117471

# Memorandum

U.S. Department
of Transportation

**National Highway
Traffic Safety
Administration**

00 DEC -4 PM 4:37

Subject: Submittal of "Perceptions of College Students Regarding Utilization of Transportation Recorders in the Highway Mode" to Docket No. NHTSA-99-5218 - 6

Date: September 29, 2000

From: Raymond P. Owings, Ph.D.
Associate Administrator for
Research and Development

Reply to
Attn. of: NRD-01

To: The Docket

THRU: Frank Seales, Jr.
Chief Counsel

Research and Development requests that the report titled "Perceptions of College Students Regarding Utilization of Transportation Recorders in the Highway Mode" be placed in the public docket.

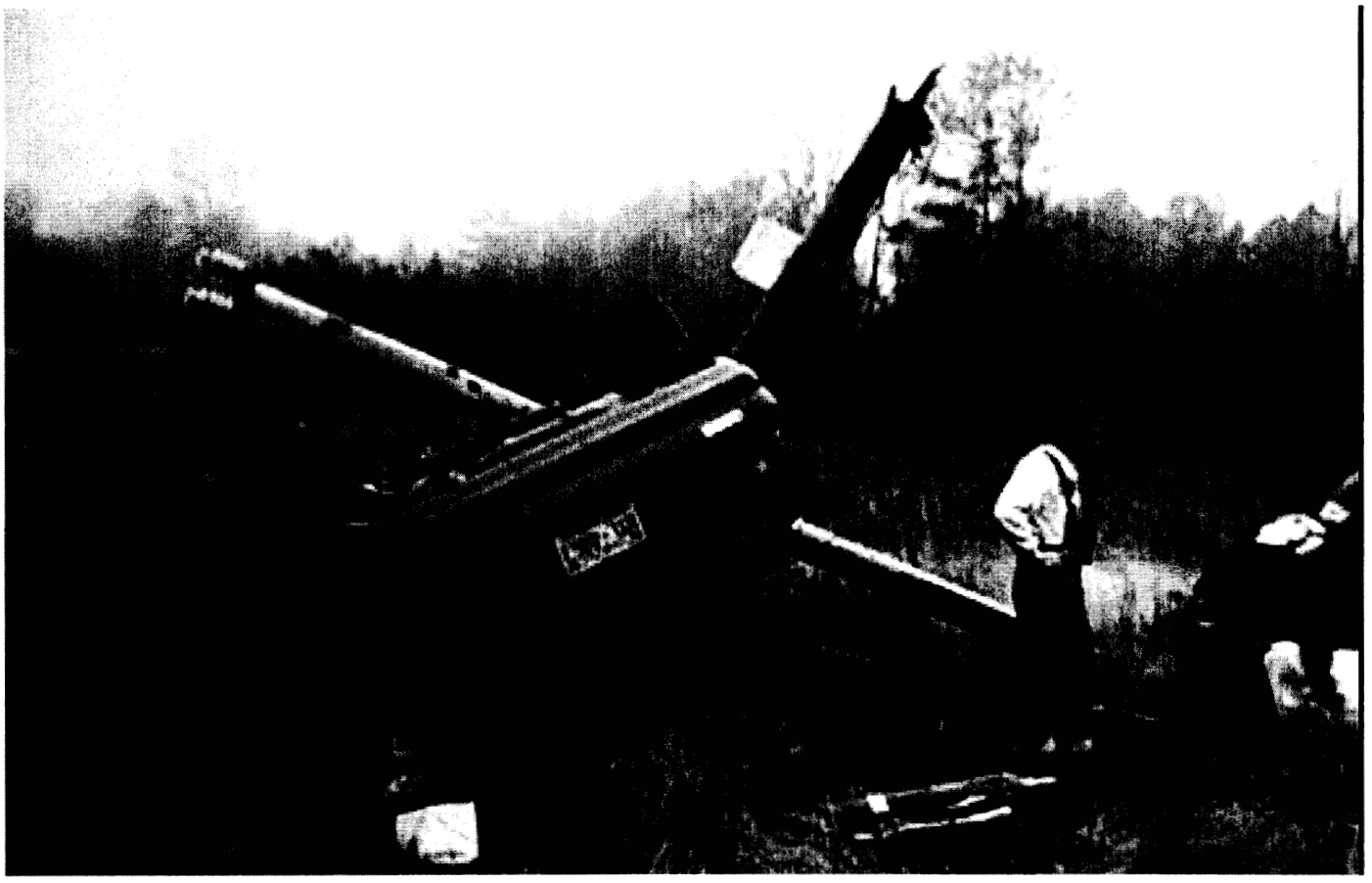
Since this report is long, NHTSA R&D is including an electronic version of this report for your use. The file name is "Perceptions of Transportation Recorders.pdf"

Attachments

#

Get it together!
SAFETY BELTS SAVE LIVES

# Perceptions of College Students Regarding Utilization of Transportation Recorders in the Highway Mode



## Research Project

Compiled by
Professor Thomas Michael Kowalick
Sandhills Community College, Pinehurst, North Carolina
June 16 to July 28, 1999
© 1999 Thomas Michael Kowalick

# Introduction

*More than 6.7 million police-reported motor vehicle crashes occurred on our highways in 1997 – one every 5 seconds.*

Photographer: Frank Staples
© 1999 Click, Incorporated, Transportation Safety Technologies

There is a great need and strong desire within the general public for awareness and understanding of emerging technologies in the highway mode of transportation. This is especially true regarding transportation recorders. Debate concerning the promise of potential safety benefits versus the fear of privacy invasion requires public involvement. Although public interest is very high in this area, public discussion is very low. Experts knowledgeable in the technological aspects are often lacking in the societal implications and ramifications, which limits public participation. Thus, this research project aims to address key issues of safety and privacy at a time in transportation history when recorders are being considered for the highway mode. Perhaps the findings can provide insights and stimulate additional interest and discussion.

# How the Research Project was Conducted

*Midnight to 3 AM on Saturdays and Sundays proved to be the deadliest 3-hour periods.*

Photographer: Frank Staples
© 1999 Click, Incorporated, Transportation Safety Technologies

- 71 students were asked to read a press release from MSNBC, dated June 1, 1999, regarding General Motors' "black box" technology. [See Attachment # 1]

- After reading this article and following a session of asking clarifying questions, these students completed a non-personal identifier template designed to establish basic information. [See Attachment # 2]

- The students were then asked to express their perceptions (not opinions) regarding this technology. They were asked to list both the positive and negative aspects, and not to draw any quick conclusions. The goal was to be objective, regardless of personal opinions. [See Attachment # 3]

- To assure objectivity, two student advocates were asked to consolidate the group responses in compendium format. Thus, one very proactive student advocate in favor of emerging technologies – Eddie C. Rochester – was asked to coordinate all of the positive input, and one very proactive student advocate in opposition to emerging technologies – Kyle D. Long – was asked to coordinate all of the negative input.

- A professor knowledgeable of transportation recorder history and current initiatives provided an objective overview without inherent bias to assure that both sides of the issue were addressed and adequately represented.

- Data was considered and compiled by a proactive student statistician – Annie K. Bridges – without inherent bias. The files and formats were completed with Adobe Acrobat™ to assure universal document exchange.

- Additionally, a local amateur photographer was identified. Mr. Frank Staples contributed the photographs to this research project.

- The data was tabulated and verified for objectivity and credibility by three faculty reviewers:

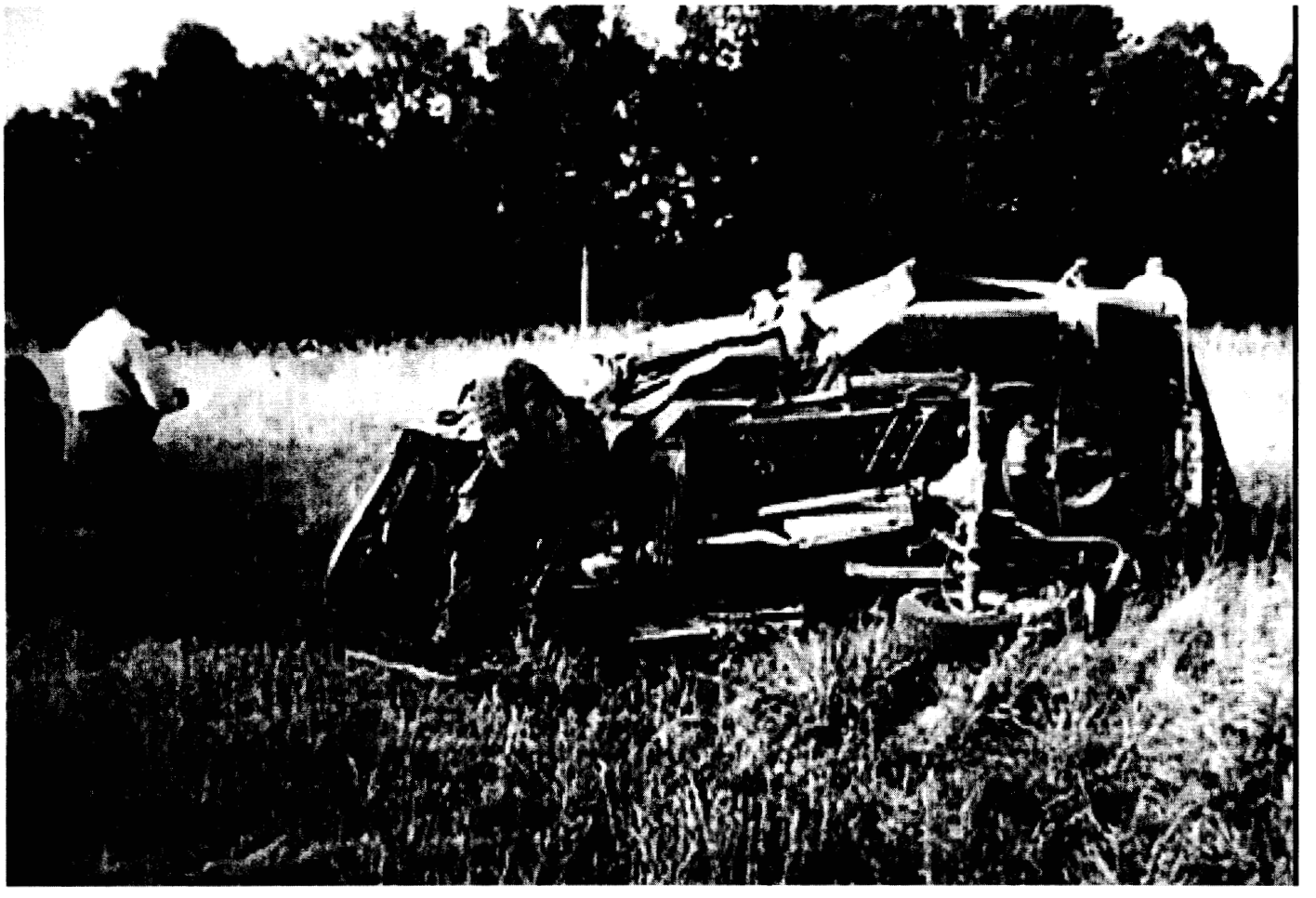
  Dr. Robert Baker, Chair, Department of Social Sciences and Humanities,

  and Professor of Sociology

  Dr. Patricia Toney, Professor of Psychology

  Dr. Wayne Adams, Professor of Sociology

# Table of Contents



*A total of 41,967 people lost their lives in motor vehicle crashes in 1997. Another 3.4 million people were injured.*

Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

Title Page i
Introduction ii
How the Research Project was Conducted iv
Table of Contents vii
Non-Personal Identifiers 1
Sex 2
Age 3
Years Driving 4
Crashes as Occupant 5
Crashes as Driver 6
% Seat Belt Usage 7
Seat Belt Used in Crash? 8
Injured by Belt? 9
Injured by Airbag? 10
Trust Airbag? 11
County of Registration 12
Data for Non-Personal Identifiers 13
Positive Perceptions 15
Category Overview 16
Crash Data Determination 17
Safety Improvements 18
Insurance 19
Health Care 20
Airbags 21
Highway Infrastructure 22
Car Prices 23
Awareness 24
Negative Perceptions 25
Category Overview 26
Privacy Implications 27
Misuse of Data 28
Consumer Knowledge and Autonomy 29
Financial Issues 30
Constitutional and Legislative Issues 31
General Functioning of Device 32
Compendium of Positive Perceptions 33
Compendium of Negative Perceptions 35
Positive Comments 37
Negative Comments 42
Addenda 47
*Automobile Collision Data* (OTA) 48
*Electronic Record Systems and Individual Privacy* (OTA) 300
*Accessibility and Integrity of Networked Information Collections* (OTA) 452
*1997 Traffic Crashes, Injuries, and Fatalities* (NHTSA) 565
*ITS Fair Information and Privacy Principles* (ITS America) 569
*Drivers Privacy Protection Act* (18 U.S.C.) 573
*Federal Advisory Committee Act* (5 U.S.C.) 577
*Freedom of Information Act* (5 U.S.C.) 588
*Government in the Sunshine Act* (5 U.S.C.) 598
*Privacy Act of 1974* (5 U.S.C.) 604
Attachments 622
MSNBC Article 623
Non-Personal Identifier Template 626
Positive / Negative Perceptions Template 627
Resources 628
Photo Credits 633
Conclusion 635

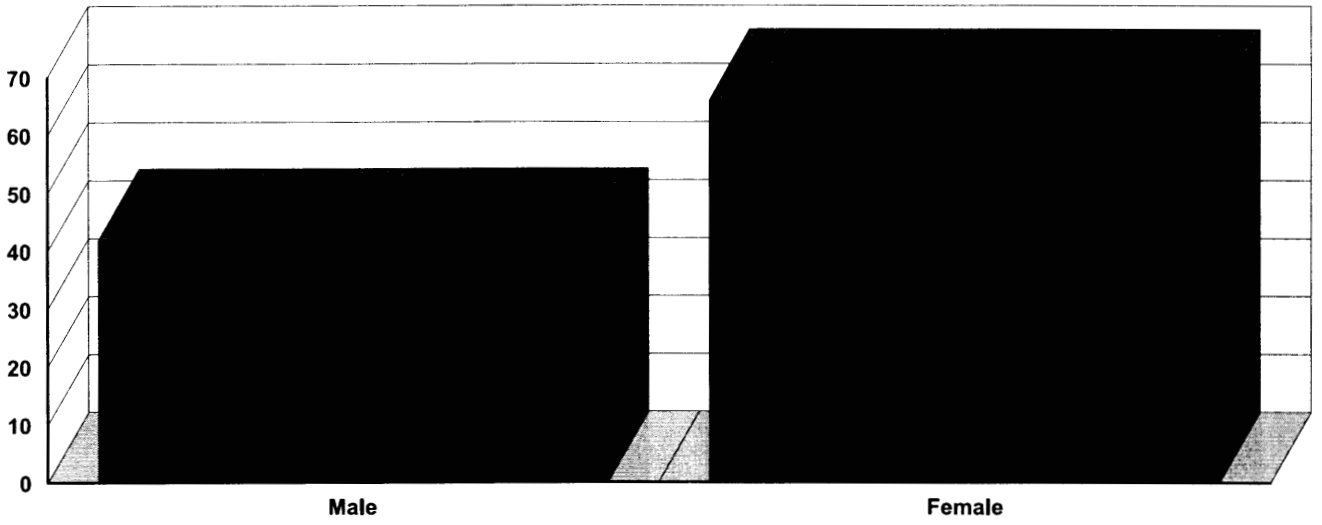
# Non-Personal Identifiers

*On average, a person was injured in these crashes every 9 seconds, and someone was killed every 13 minutes.*

Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

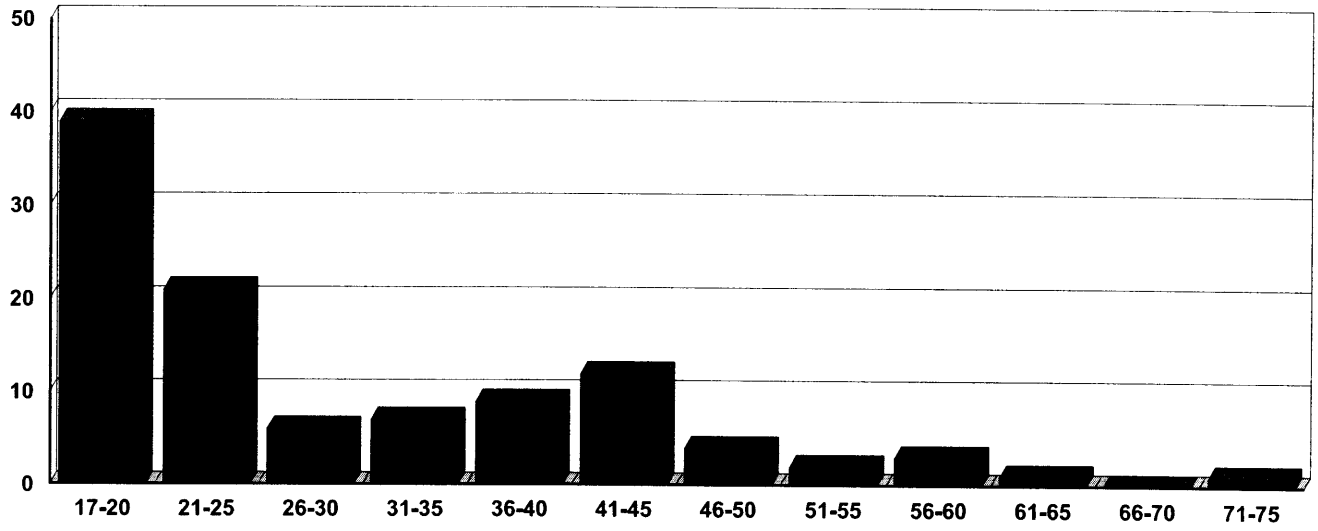
# Non-Personal Identifiers: Sex



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

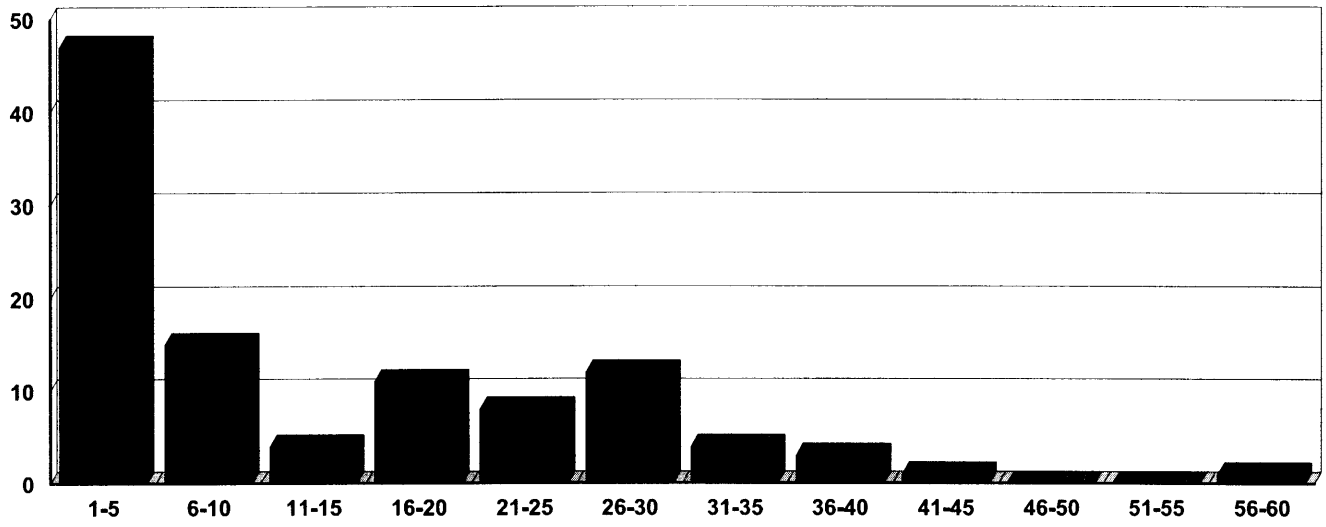
# Non-Personal Identifiers: Age



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

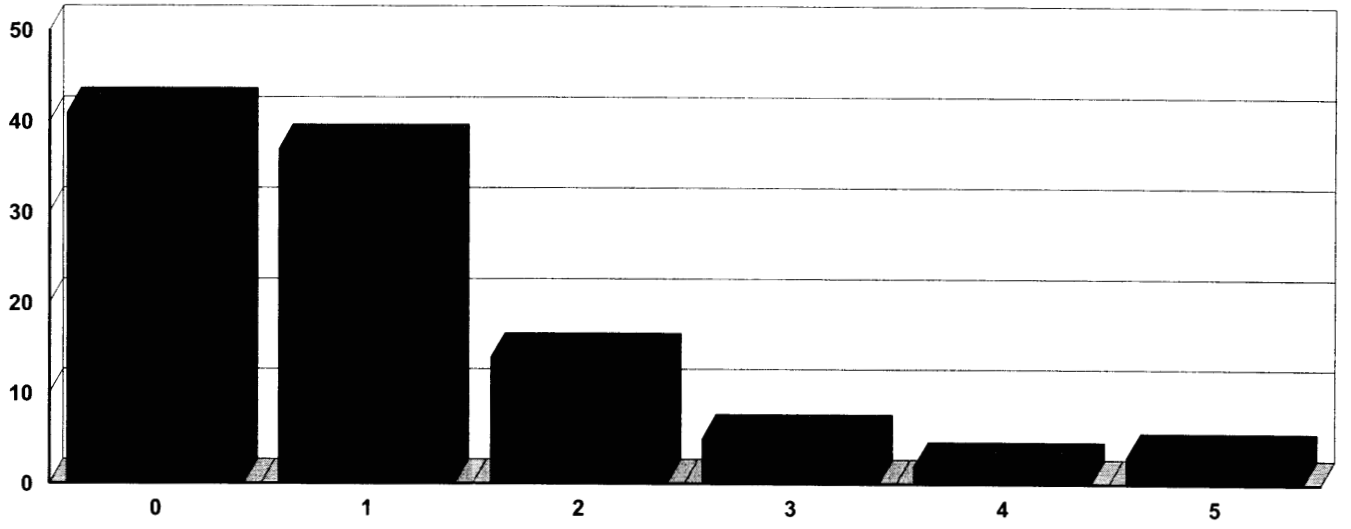
# Non-Personal Identifiers: Years Driving

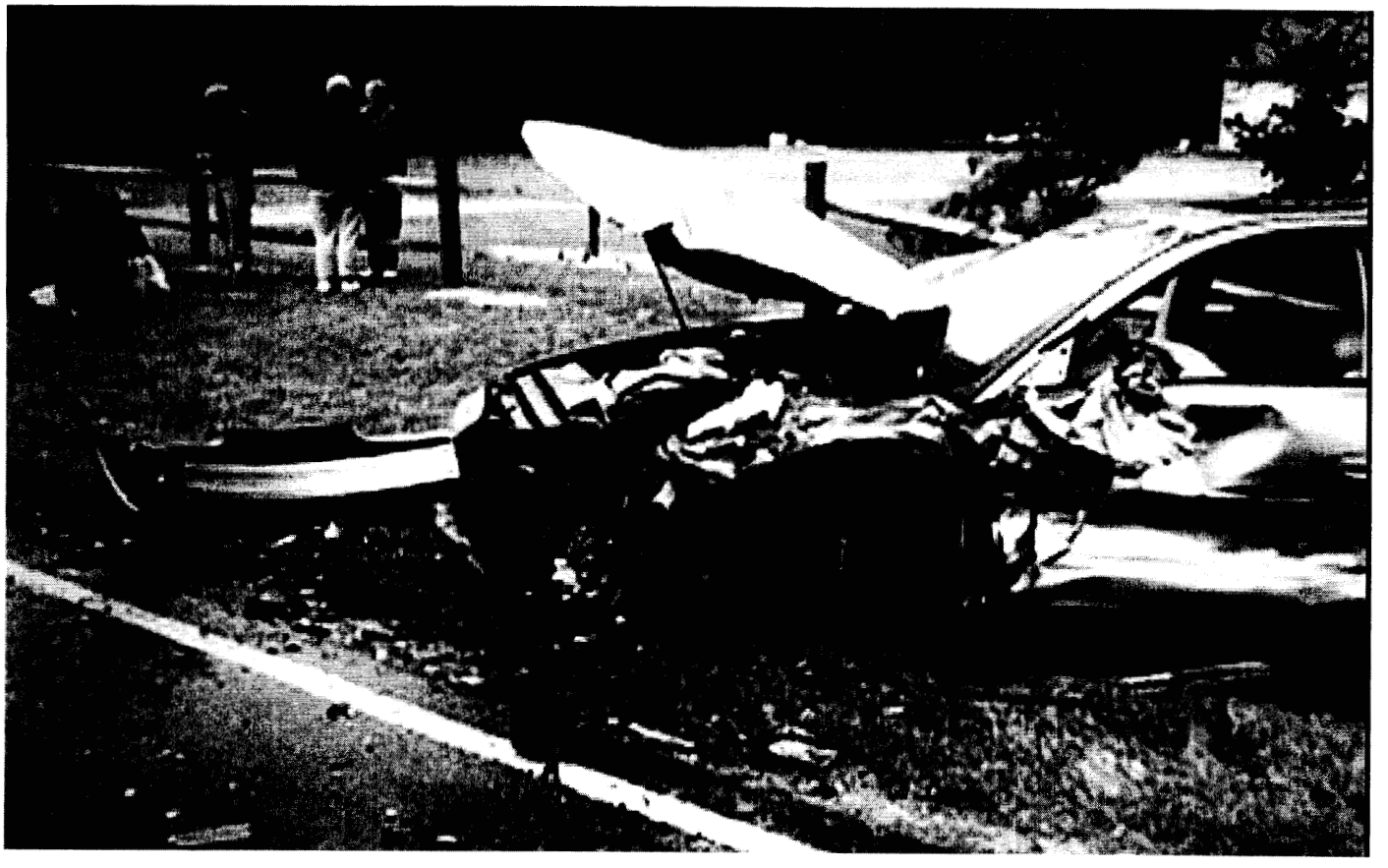


Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

# Non-Personal Identifiers: Crashes as Occupant

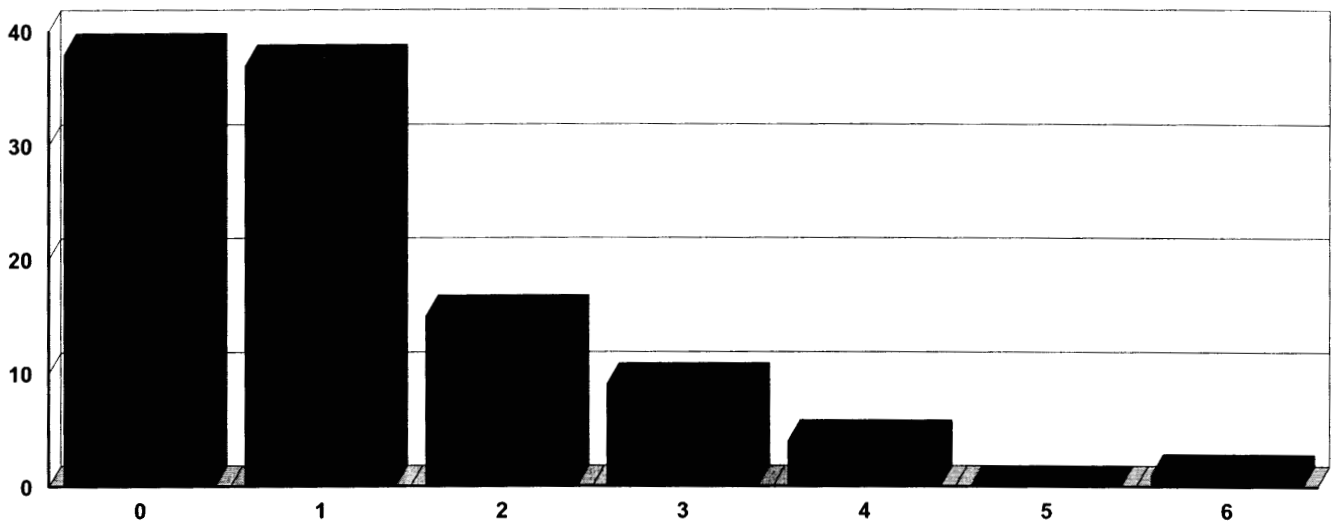


Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

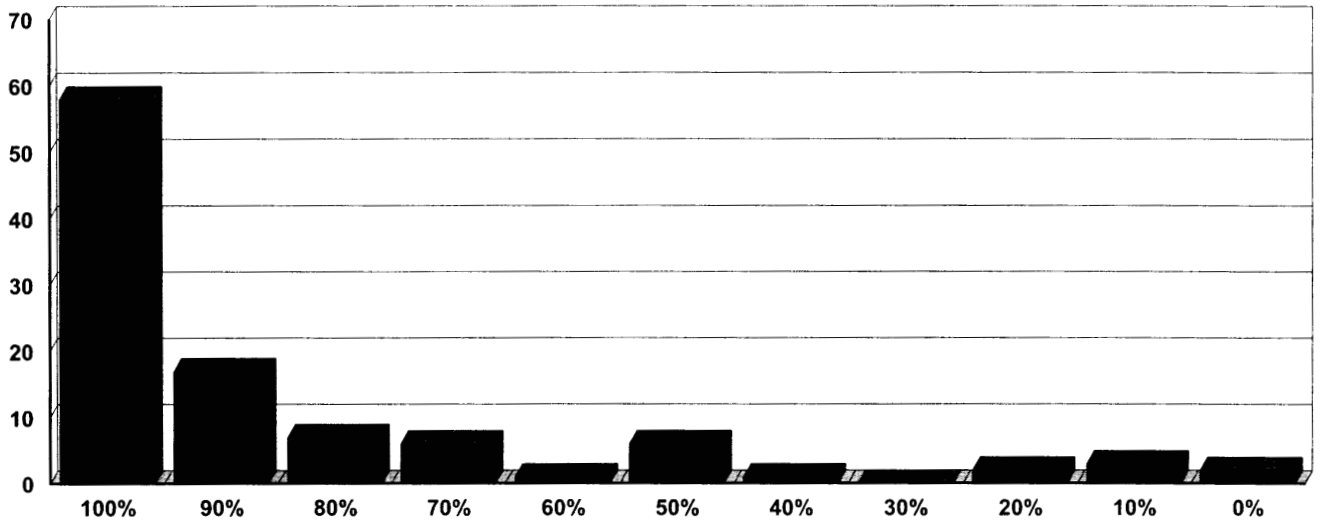
# Non-Personal Identifiers: Crashes as Driver



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

# Non-Personal Identifiers:
# % Seat Belt Usage



Photographer: Frank Staples

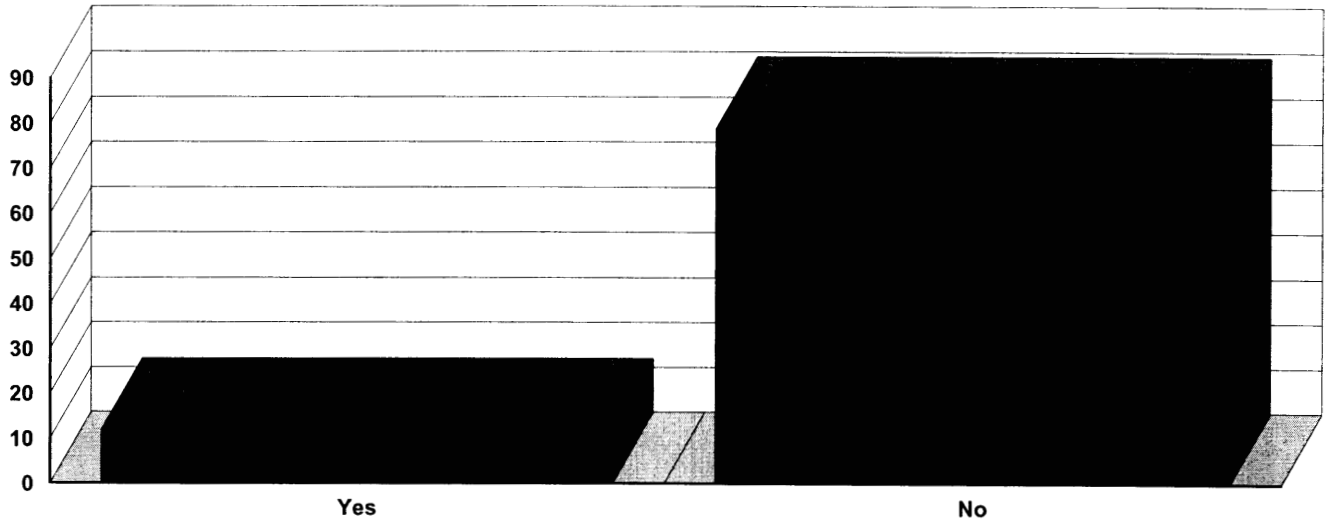
© 1999 Click, Incorporated, Transportation Safety Technologies

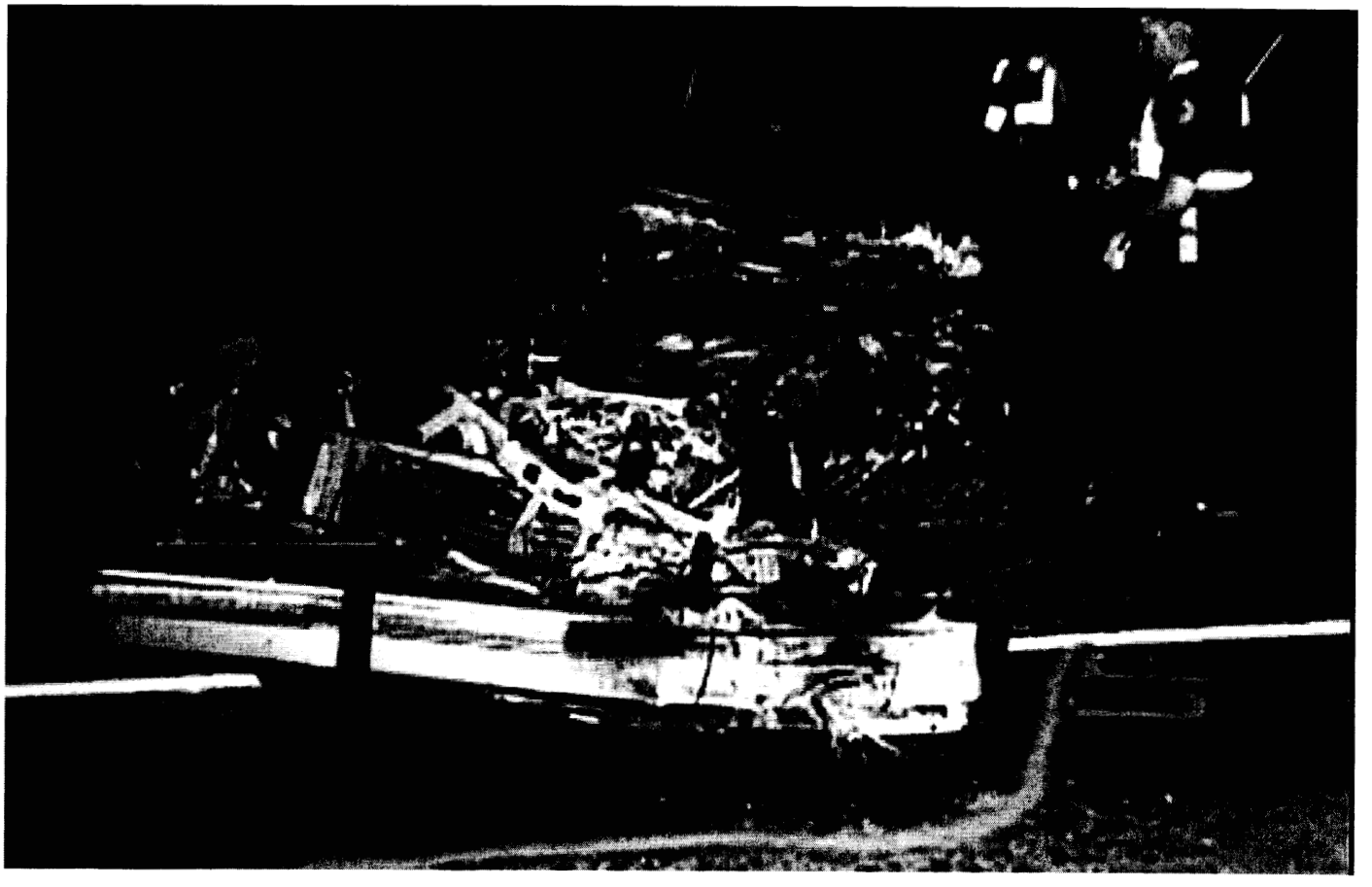
# Non-Personal Identifiers: Seat Belt Used in Crash?

Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

# Non-Personal Identifiers: Injured by Belt?

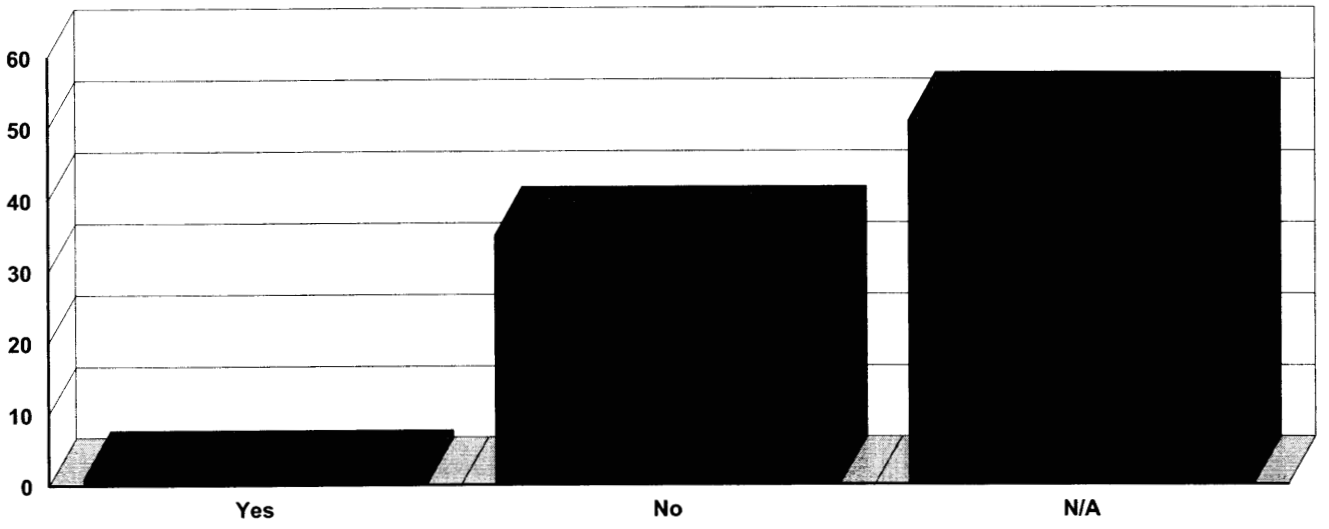


Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

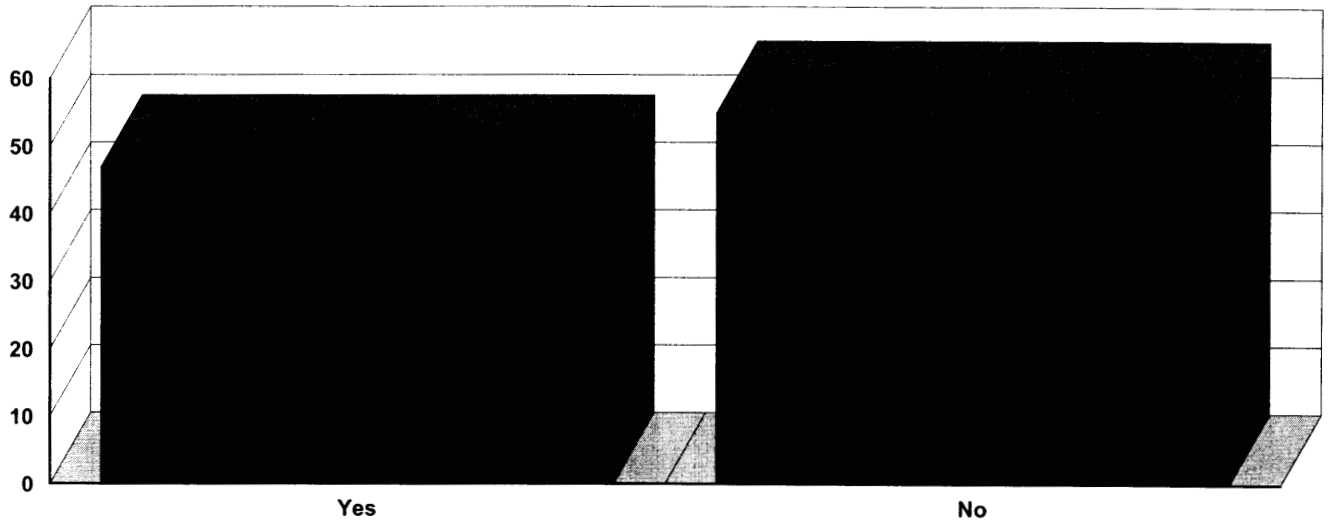
# Non-Personal Identifiers: Injured by Airbag?



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

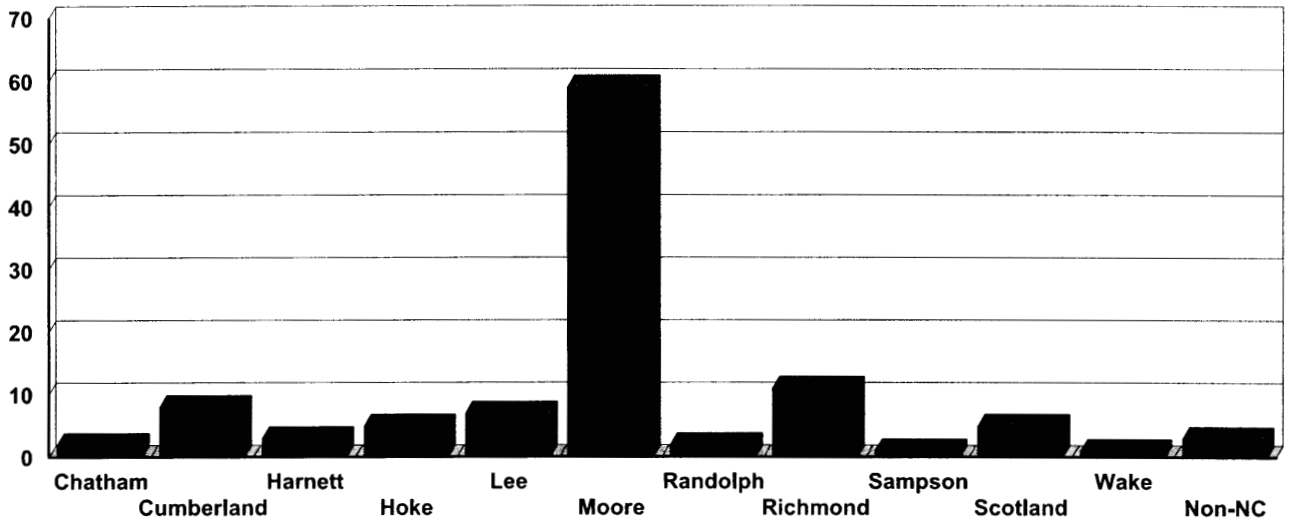
# Non-Personal Identifiers: Trust Airbag?



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

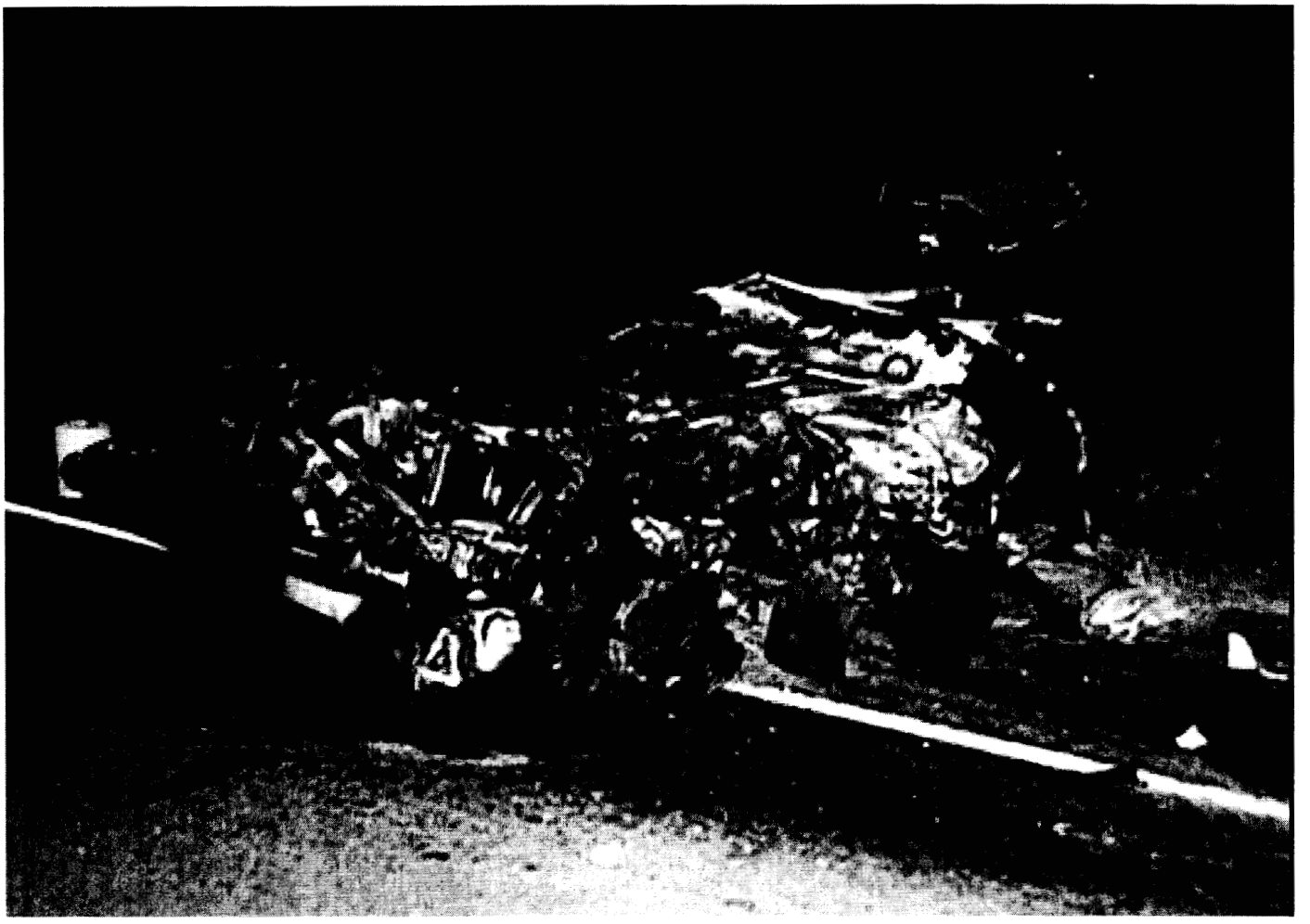
# Non-Personal Identifiers: County of Registration



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

# Data for Non-Personal Identifiers



*39% of fatal crashes involved alcohol. For fatal crashes occurring from midnight to 3 AM, 75% involved alcohol.*

Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

| Sex | # | Age | # | Years Driving | # | Crashes as Occupt. | # | Crashes as Driver | # | % Seat Belt Usage | # |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Male | 42 | 17-20 | 39 | 1-5 | 47 | 0 | 41 | 0 | 38 | 100% | 58 |
| Female | 66 | 21-25 | 21 | 6-10 | 15 | 1 | 37 | 1 | 37 | 90% | 17 |
| | | 26-30 | 6 | 11-15 | 4 | 2 | 14 | 2 | 15 | 80% | 7 |
| | | 31-35 | 7 | 16-20 | 11 | 3 | 5 | 3 | 9 | 70% | 6 |
| | | 36-40 | 9 | 21-25 | 8 | 4 | 2 | 4 | 4 | 60% | 1 |
| | | 41-45 | 12 | 26-30 | 12 | 5 | 3 | 5 | 0 | 50% | 6 |
| | | 46-50 | 4 | 31-35 | 4 | | | 6 | 1 | 40% | 1 |
| | | 51-55 | 2 | 36-40 | 3 | | | | | 30% | 0 |
| | | 56-60 | 3 | 41-45 | 1 | | | | | 20% | 2 |
| | | 61-65 | 1 | 46-50 | 0 | | | | | 10% | 3 |
| | | 66-70 | 0 | 51-55 | 0 | | | | | 0% | 2 |
| | | 71-75 | 1 | 56-60 | 1 | | | | | | |

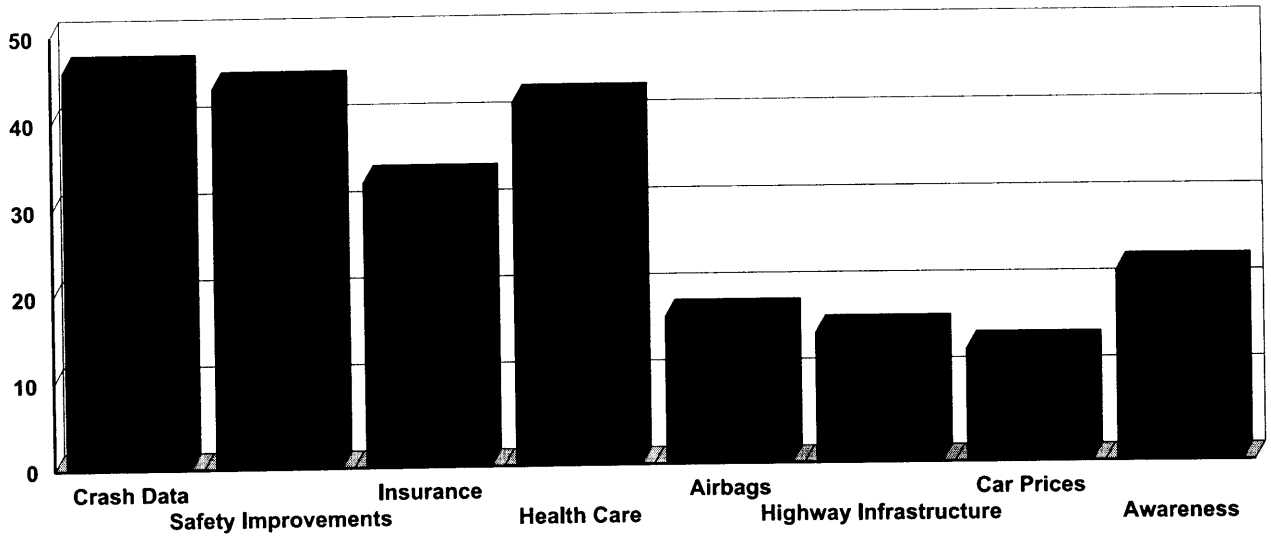
| Seat Belt Used in Crash? | # | Injured by Belt? | # | Injured by Airbag? | # | Trust Airbag? | # | County of Registr. | # |
|---|---|---|---|---|---|---|---|---|---|
| Yes | 70 | Yes | 12 | Yes | 1 | Yes | 47 | Chat-ham | 2 |
| No | 19 | No | 79 | No | 35 | No | 55 | Cumber-land | 8 |
| | | | | N/A | 51 | | | Harnett | 3 |
| | | | | | | | | Hoke | 5 |
| | | | | | | | | Lee | 7 |
| | | | | | | | | Moore | 59 |
| | | | | | | | | Rand-olph | 2 |
| | | | | | | | | Rich-mond | 11 |
| | | | | | | | | Samp-son | 1 |
| | | | | | | | | Scotland | 5 |
| | | | | | | | | Wake | 1 |
| | | | | | | | | Outside North Carolina | 3 |

# Positive Perceptions

*Fatal crashes dropped slightly (0.6%) from 1996 to 1997.*

Photographer: Frank Staples
© 1999 Click, Incorporated, Transportation Safety Technologies

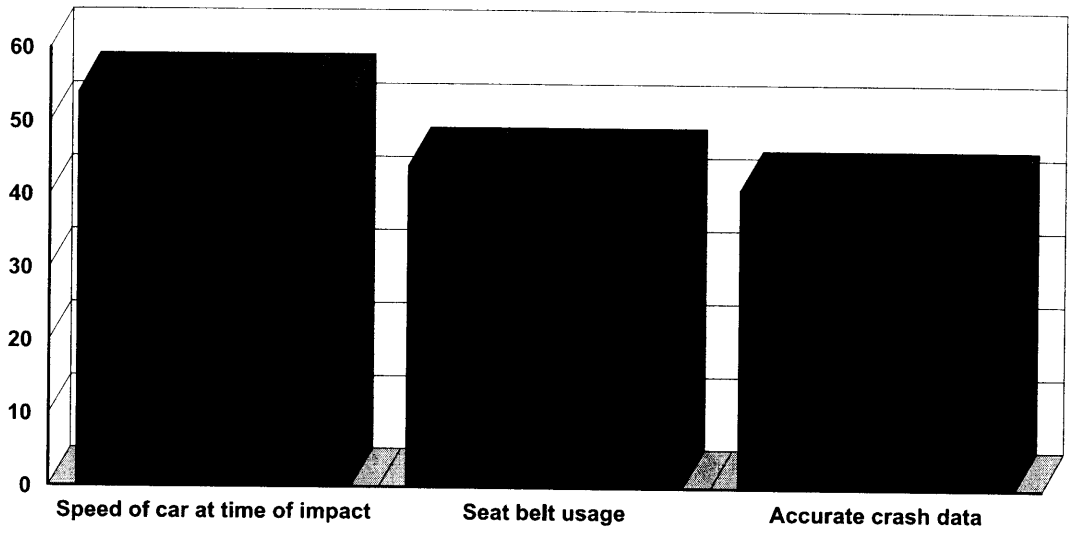
# Positive Perceptions: Category Overview



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

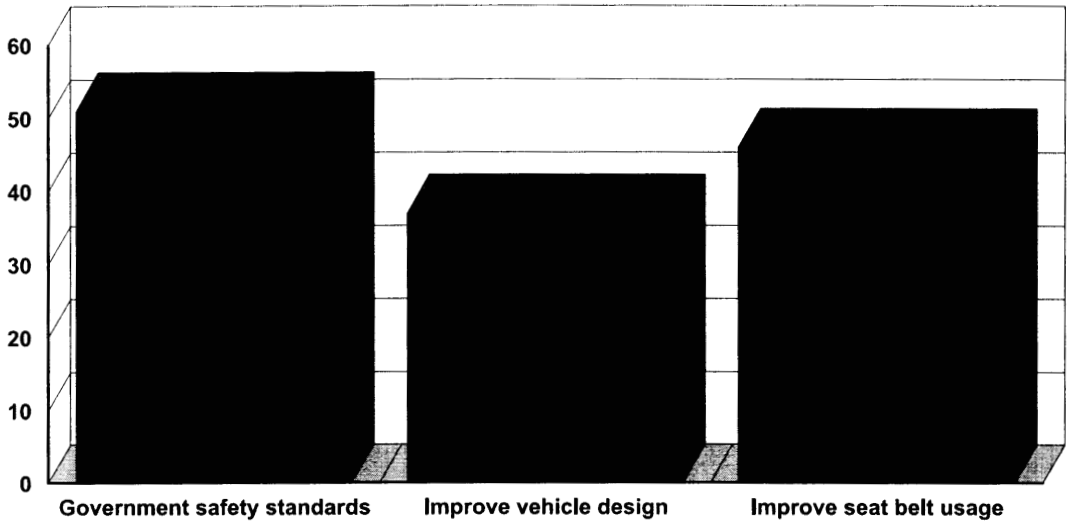
# Positive Perceptions: Crash Data Determination



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

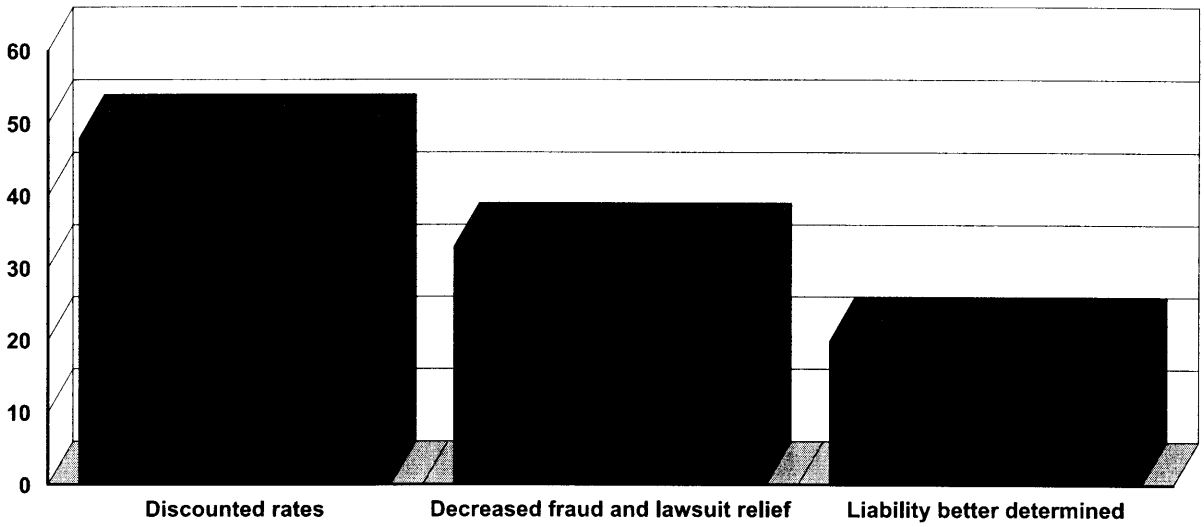
# Positive Perceptions: Safety Improvements

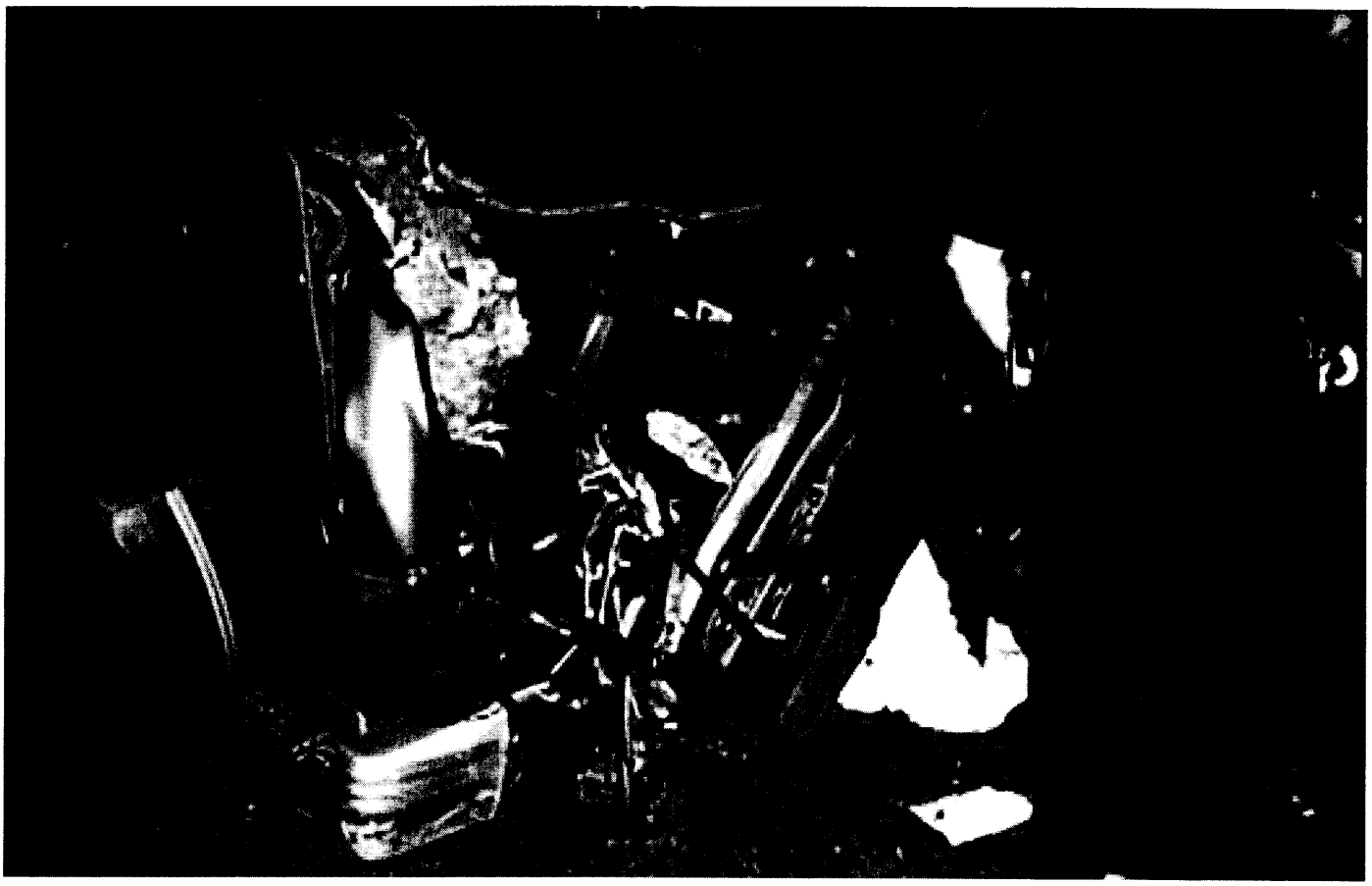


Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

# Positive Perceptions: Insurance

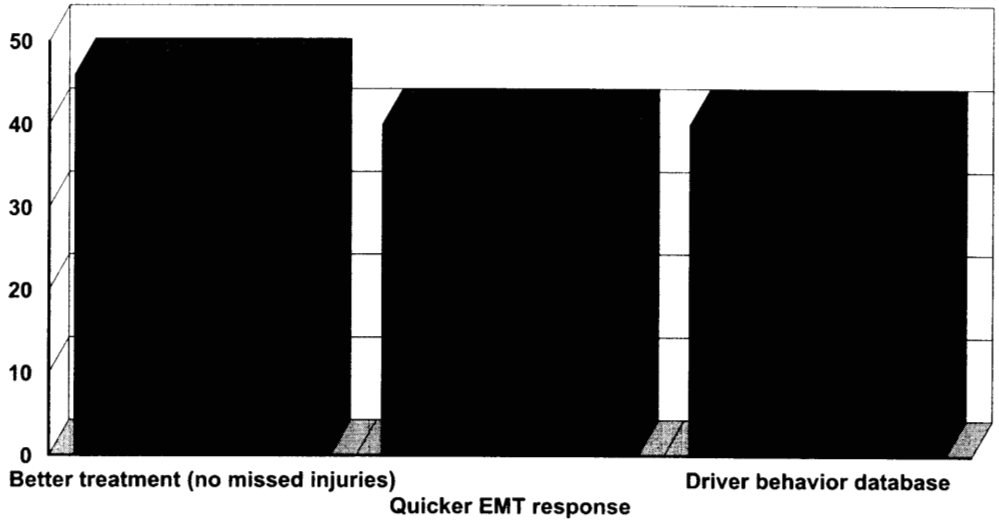


Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

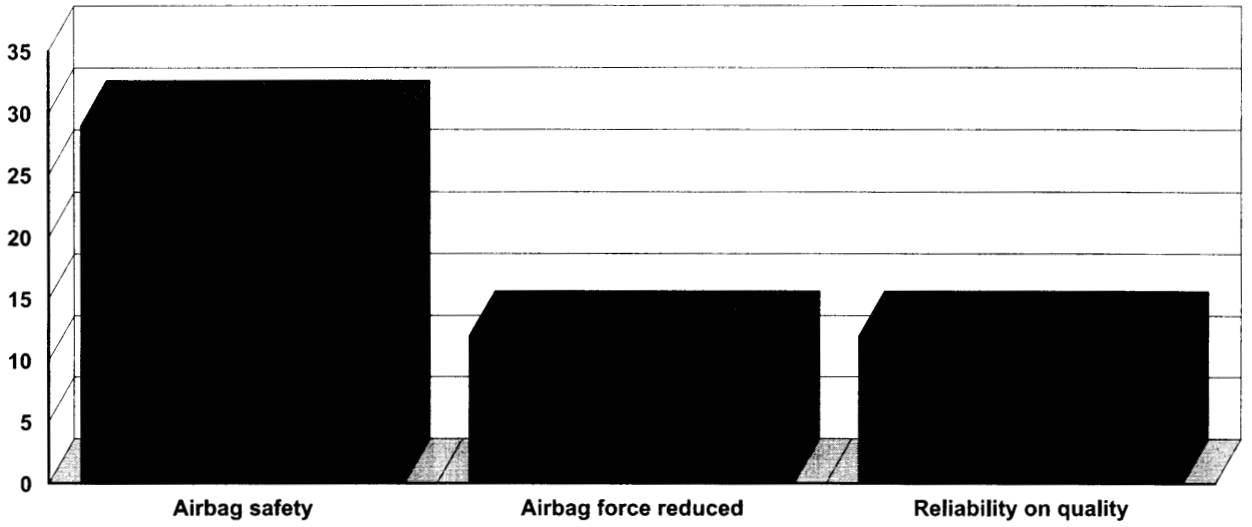
# Positive Perceptions: Health Care



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

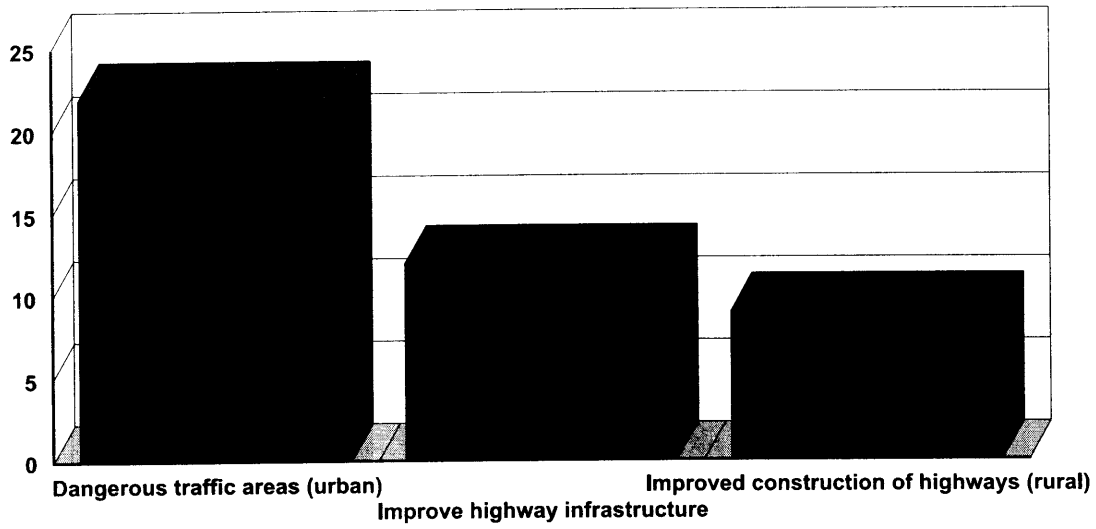
# Positive Perceptions: Airbags

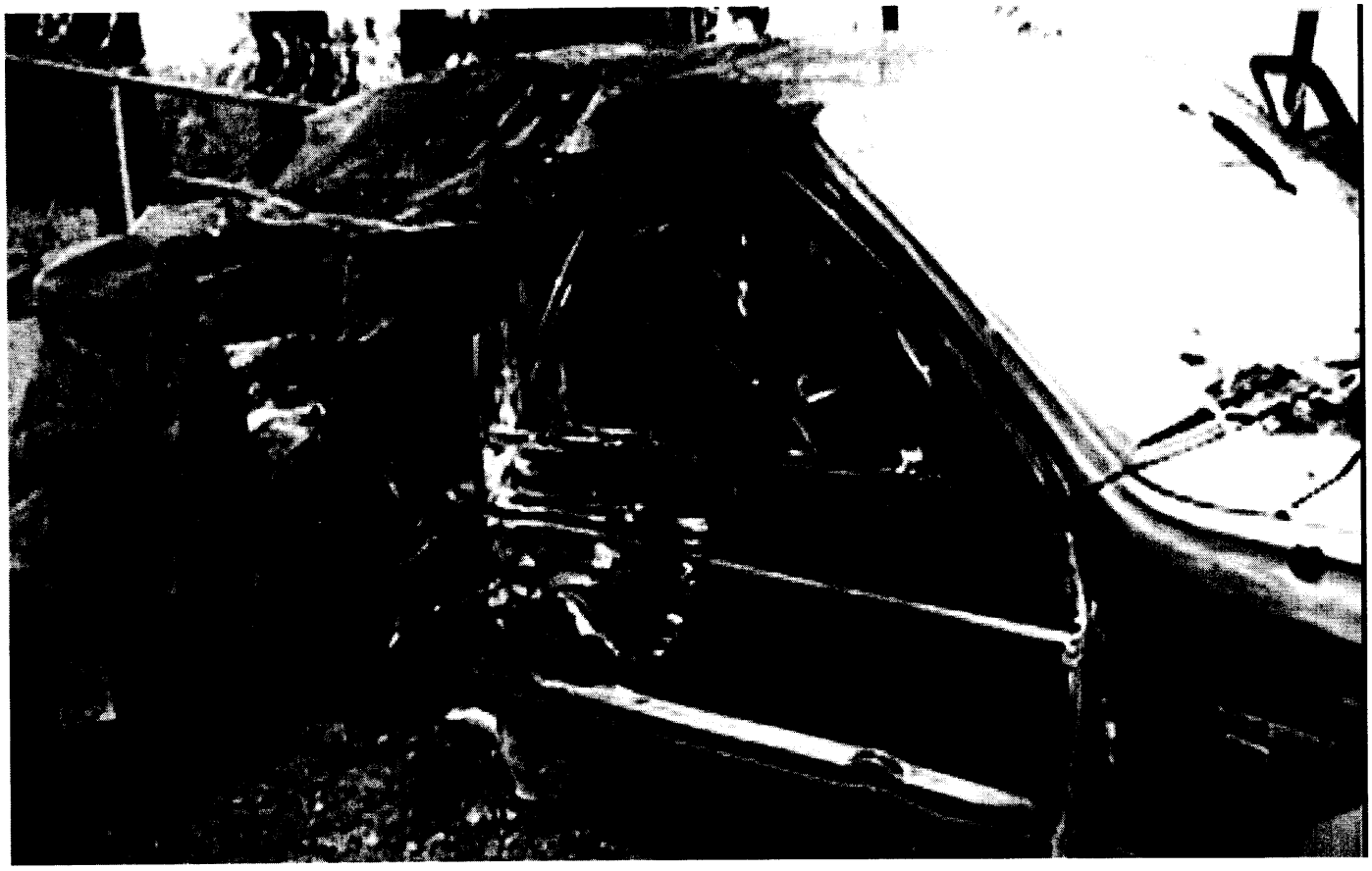


Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

# Positive Perceptions: Highway Infrastructure

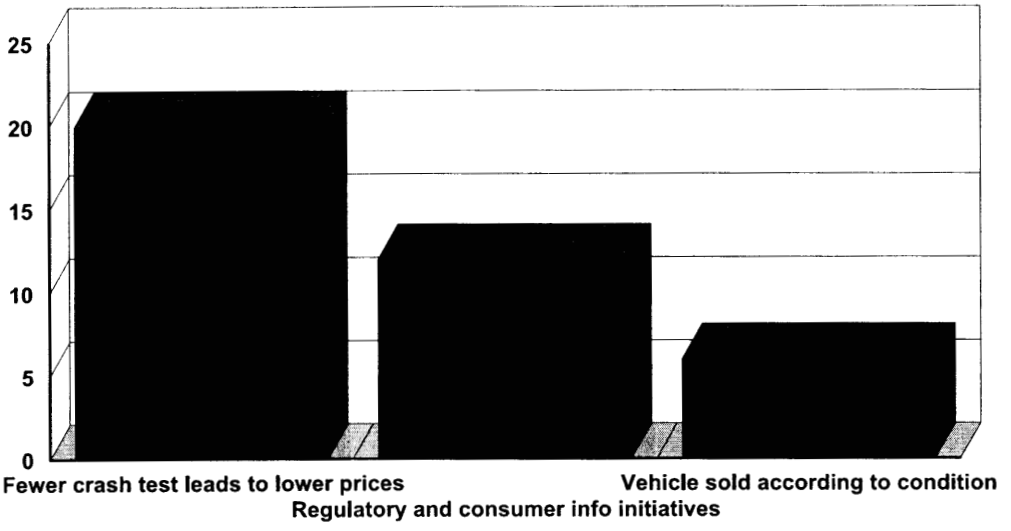


Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

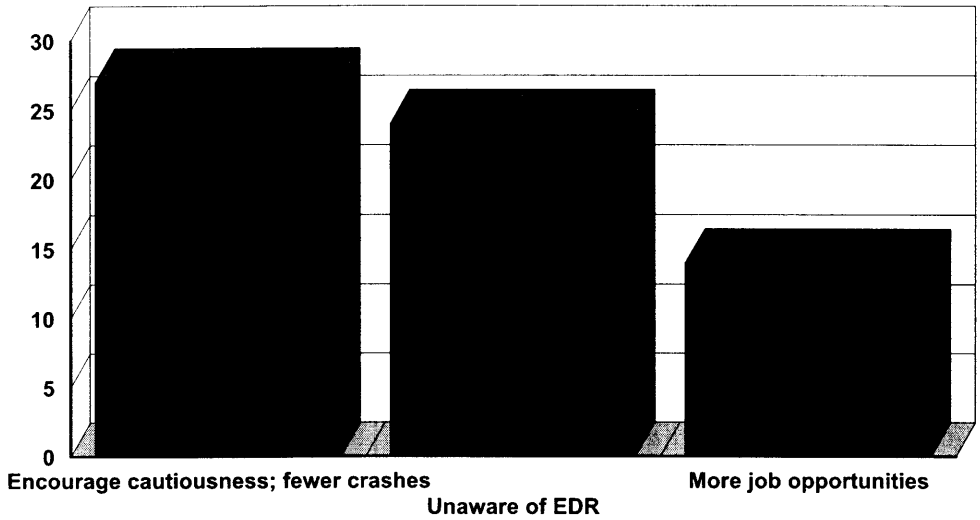
# Positive Perceptions: Car Prices

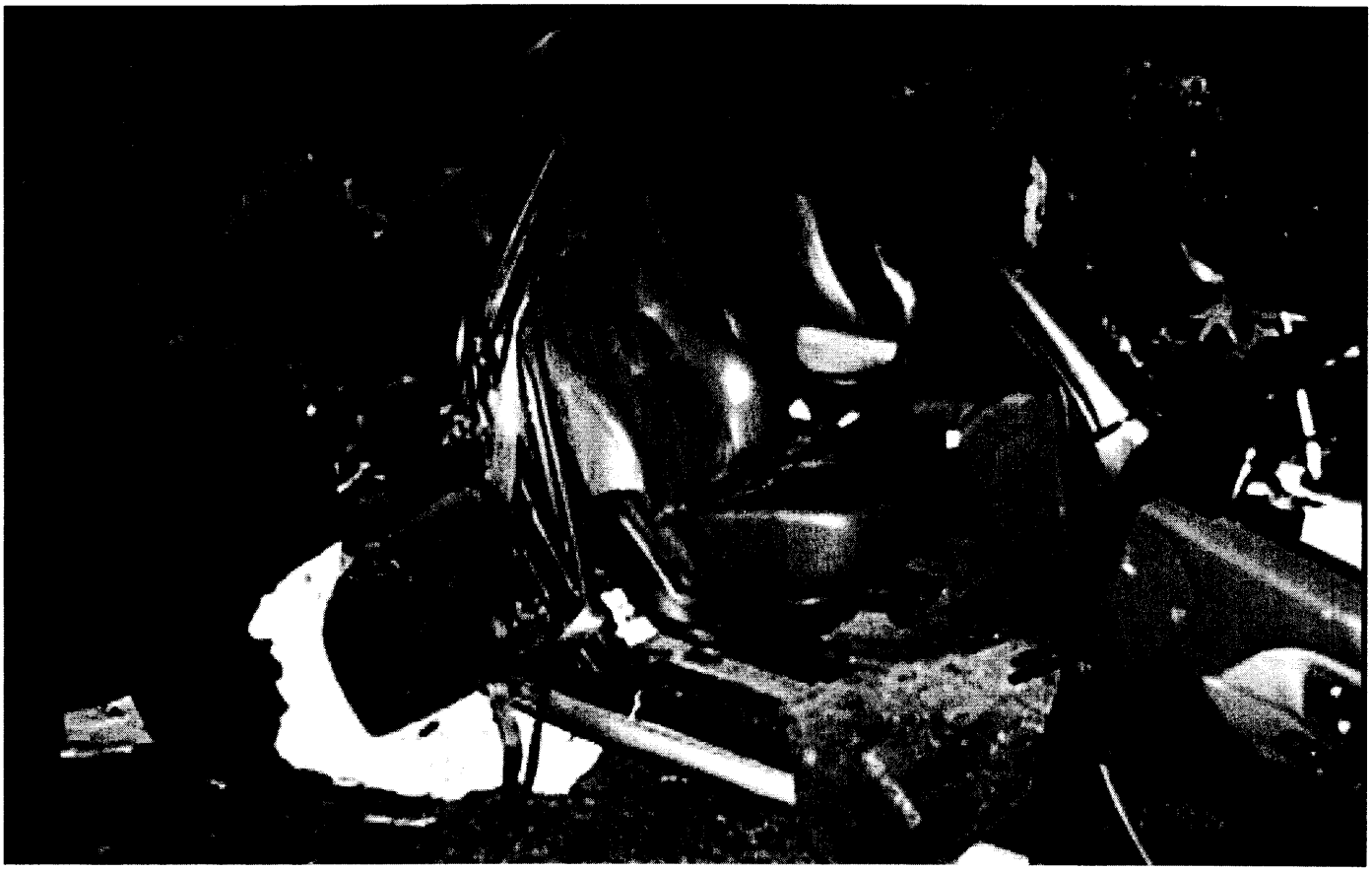


Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

# Positive Perceptions: Awareness



Photographer: Frank Staples

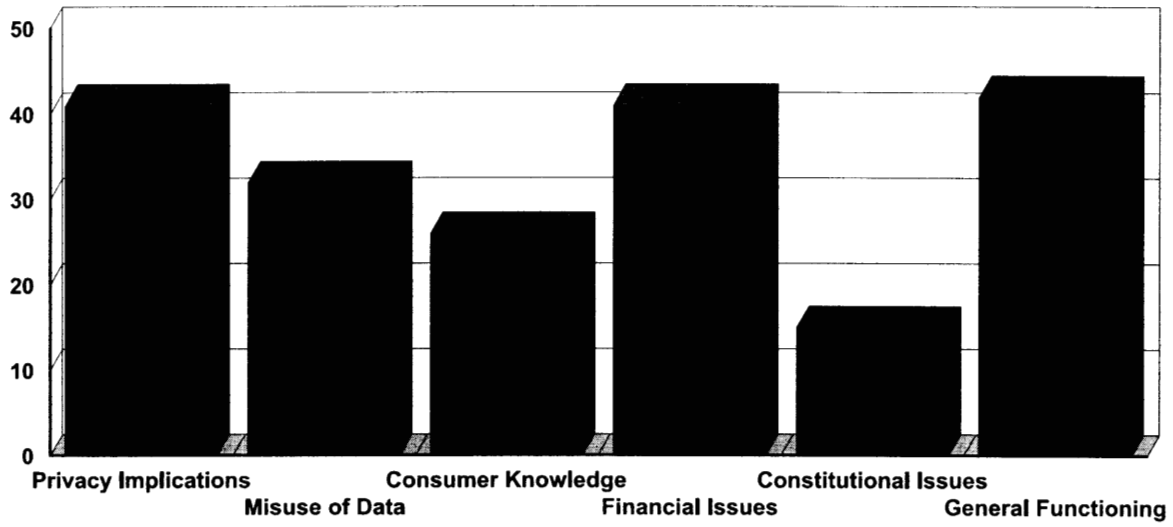
© 1999 Click, Incorporated, Transportation Safety Technologies

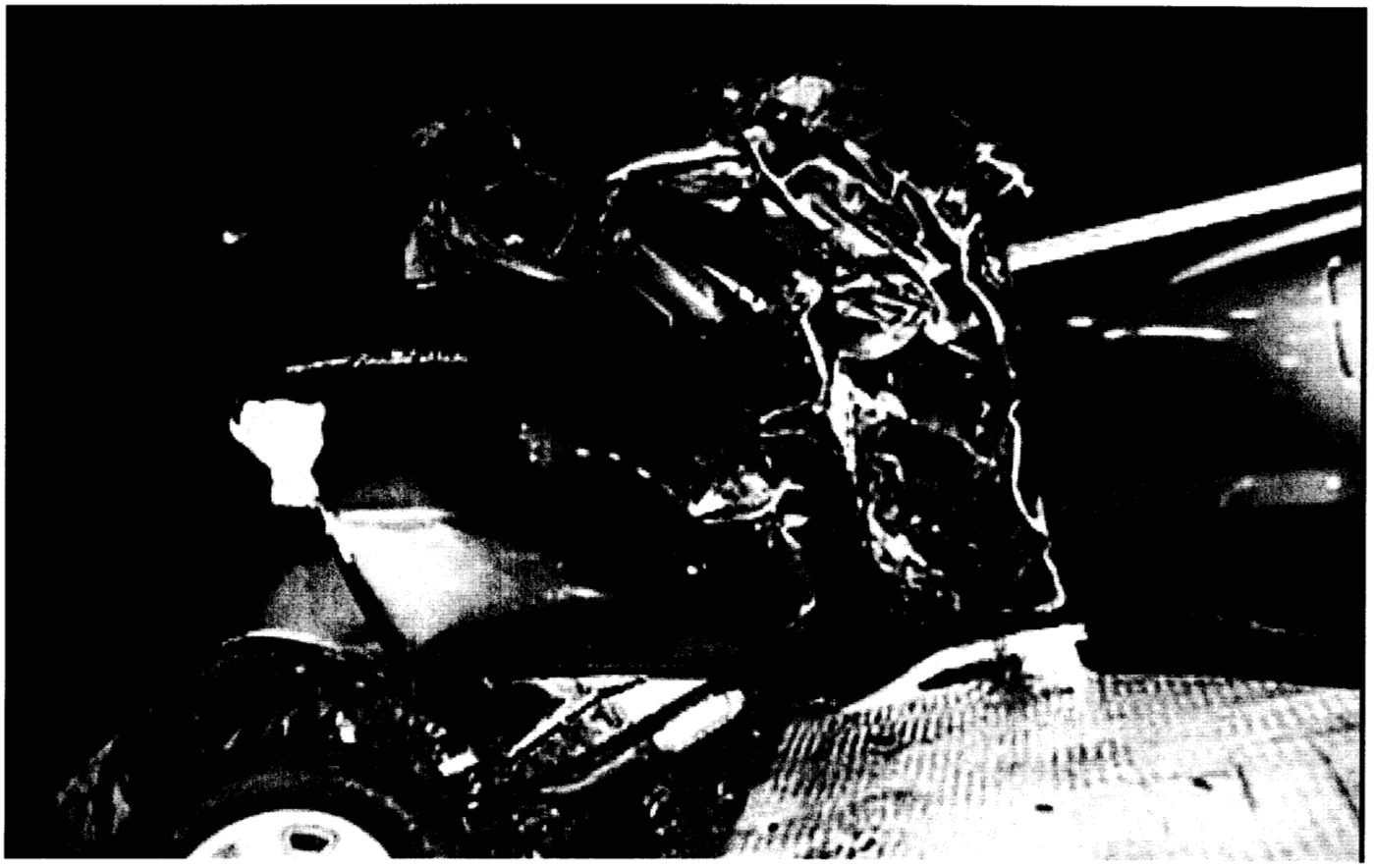
# Negative Perceptions

*Fires occurred in 0.1% of the vehicles in all traffic crashes in 1997. For fatal crashes, however, fires occurred in nearly 3% of the vehicles involved.*

Photographer: Frank Staples
© 1999 Click, Incorporated, Transportation Safety Technologies

# Negative Perceptions: Category Overview

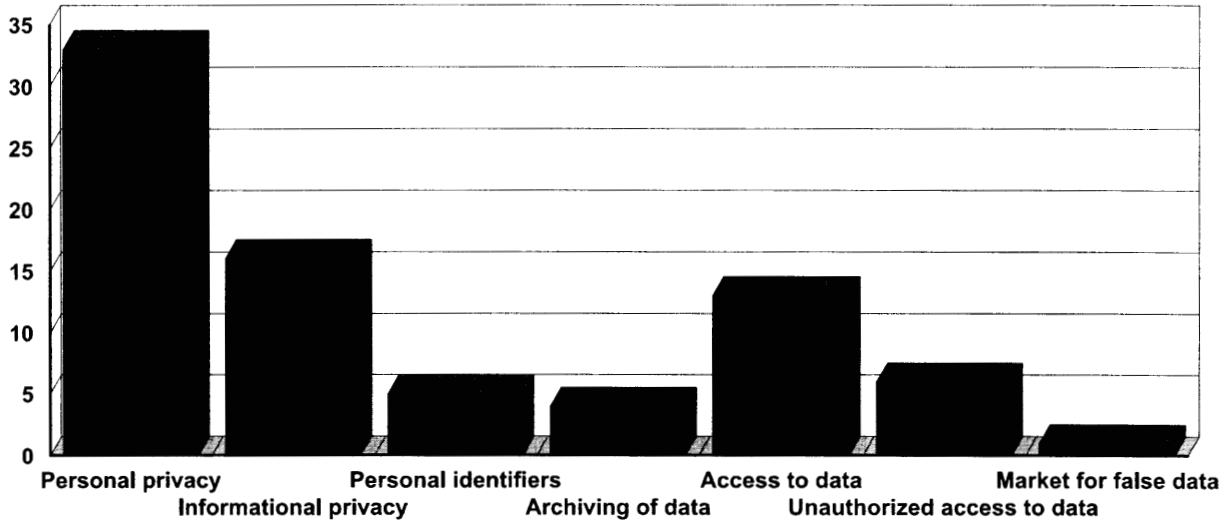


Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

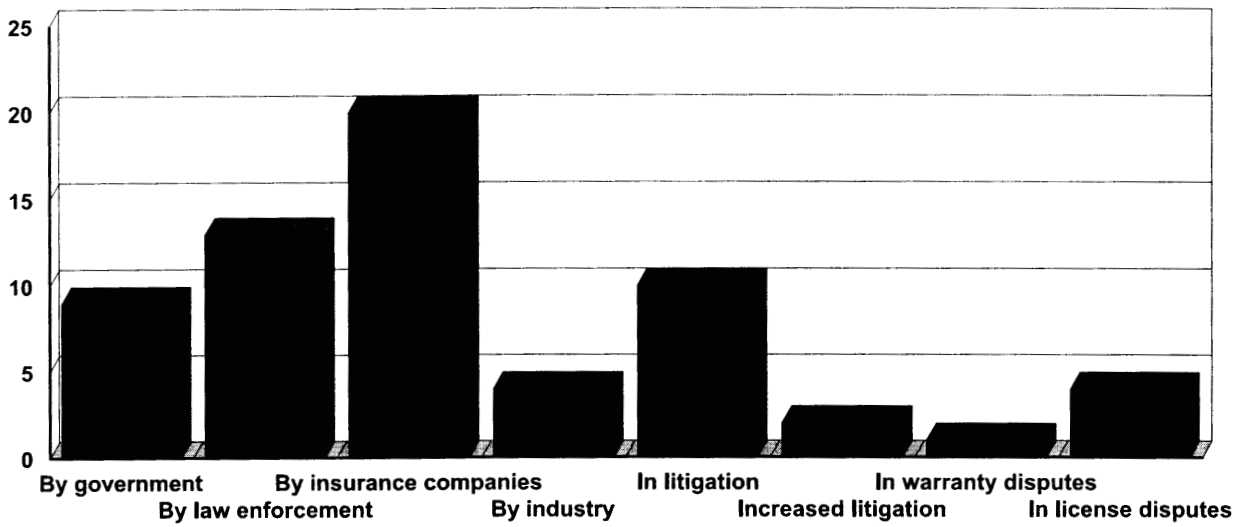
# Negative Perceptions: Privacy Implications



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

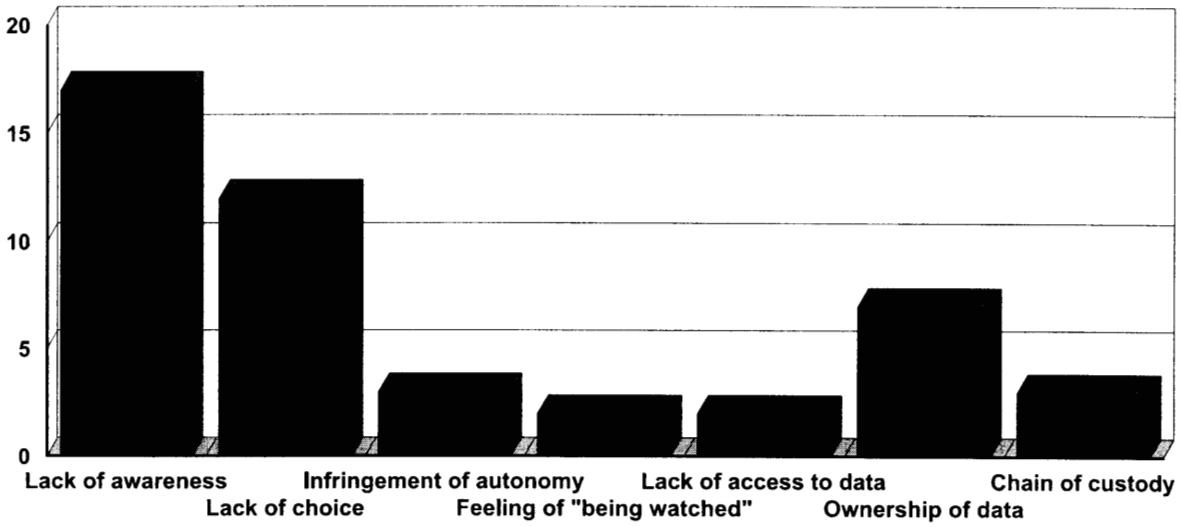
# Negative Perceptions: Misuse of Data



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

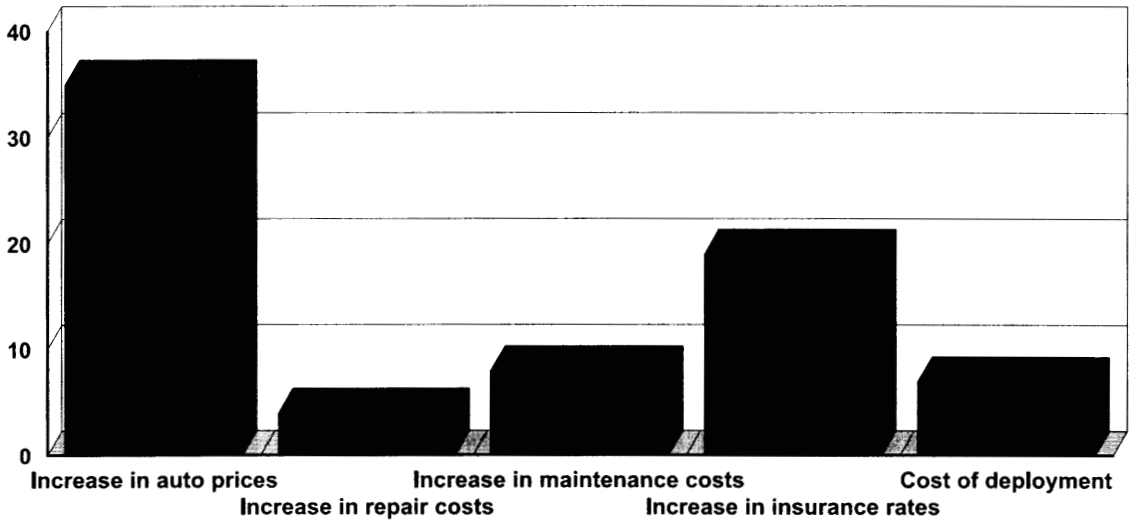
# Negative Perceptions: Consumer Knowledge and Autonomy



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

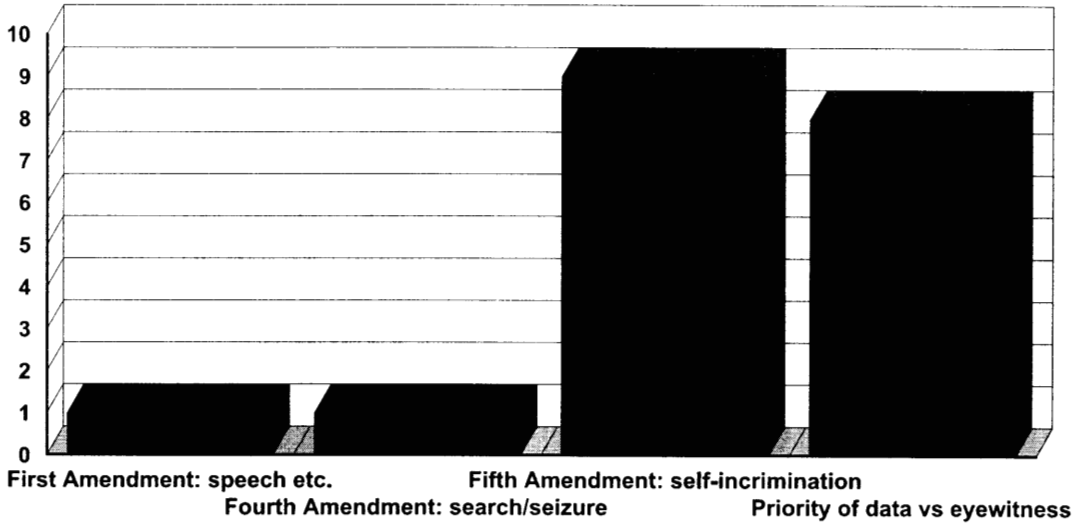
# Negative Perceptions: Financial Issues



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

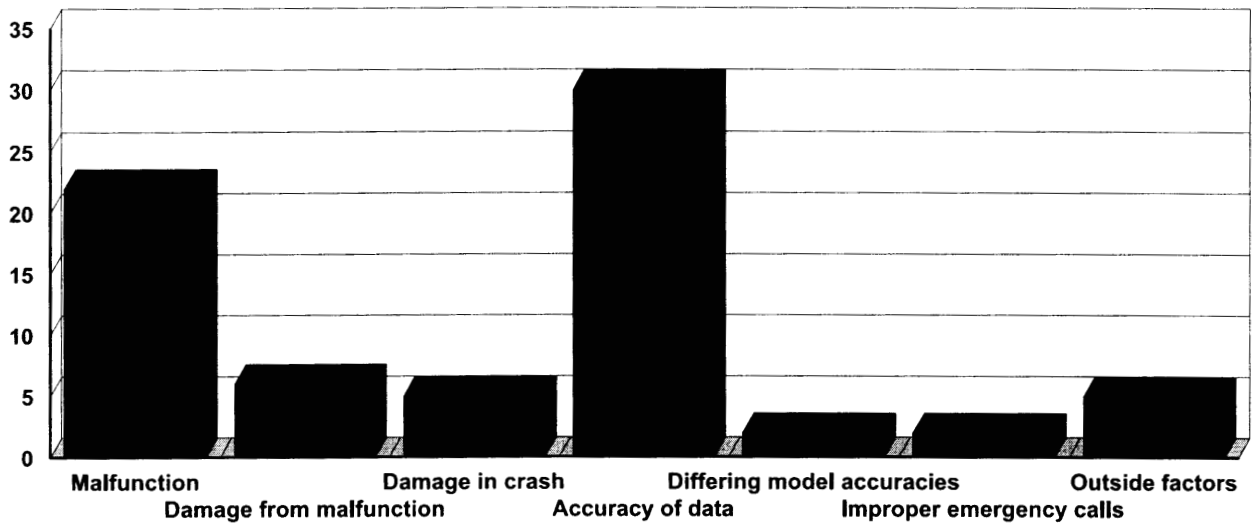
# Negative Perceptions: Constitutional and Legislative Issues

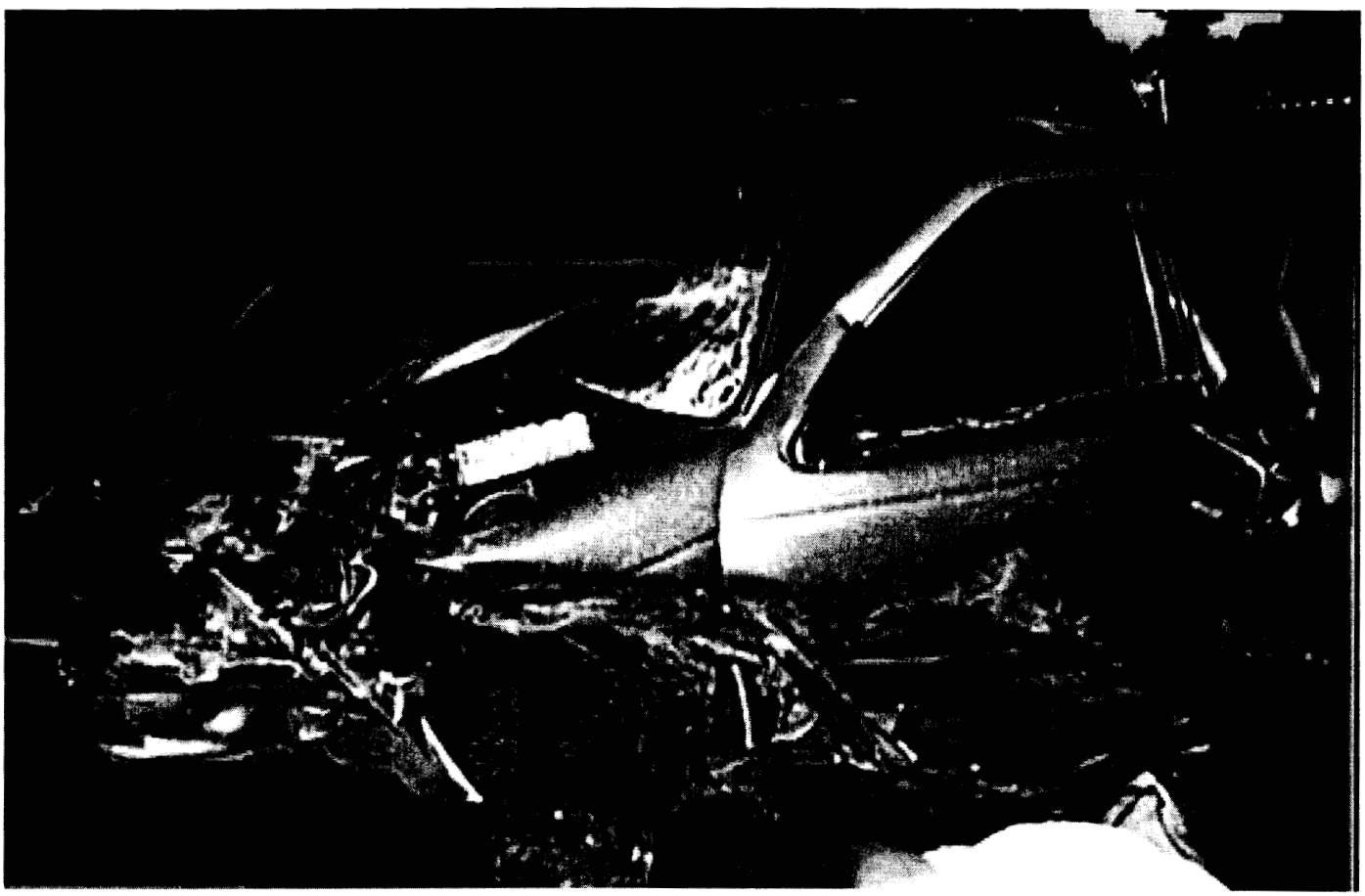


Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

# Negative Perceptions: General Functioning of Device



Photographer: Frank Staples

© 1999 Click, Incorporated, Transportation Safety Technologies

# Compendium of Positive Perceptions

*95% of the 12 million vehicles involved in motor vehicle crashes in 1997 were passenger cars or light trucks.*

Photographer: Frank Staples
© 1999 Click, Incorporated, Transportation Safety Technologies

| Concern | Number |
|---|---|
| **I. Crash Data Determination** | **46** |
| A. Speed of car at time of impact | 54 |
| B. Seat belt usage | 44 |
| C. Accurate crash data | 41 |
| **II. Safety Improvements** | **44** |
| A. Government safety standards | 51 |
| B. Improve vehicle design | 37 |
| C. Improve seat belt usage | 46 |
| **III. Insurance** | **33** |
| A. Discounted rates | 48 |
| B. Decreased fraud and lawsuit relief | 33 |
| C. Liability better determined | 20 |
| **IV. Health Care** | **42** |
| A. Better treatment (no missed injuries) | 46 |
| B. Quicker EMT response | 40 |
| C. Driver behavior database | 40 |
| **V. Airbags** | **17** |
| A. Airbag safety | 29 |
| B. Airbag force reduced | 12 |
| C. Reliability on quality | 12 |
| **VI. Highway Infrastructure** | **15** |
| A. Dangerous traffic areas (urban) | 22 |
| B. Improve highway infrastructure | 12 |
| C. Improved construction of highways (rural) | 9 |
| **VII. Car Prices** | **13** |
| A. Fewer crash tests leads to lower prices | 20 |
| B. Regulatory and consumer info initiatives | 12 |
| C. Vehicle sold according to conditions | 6 |
| **VIII. Awareness** | **22** |
| A. Encourage cautiousness; fewer crashes | 27 |
| B. Unaware of EDR | 24 |
| C. More job opportunities | 14 |

# Compendium of Negative Perceptions

*Persons 16 to 20 years old had the highest fatality and injury rates per 100,000 population.*

Photographer: Frank Staples
© 1999 Click, Incorporated, Transportation Safety Technologies

| Concern | Number |
|---|---|
| **I. Privacy Implications** | **41** |
| A. Infringement of personal privacy / surveillance | 33 |
| B. Infringement of informational privacy | 16 |
| C. Inclusion of personal identifiers | 5 |
| D. Archiving of recorded data | 4 |
| E. Access to, sale, or release of data to third parties | 13 |
| F. Unauthorized data intrusion and tampering | 6 |
| G. Black market for falsified data | 1 |
| **II. Misuse of Data** | **32** |
| A. By government | 9 |
| B. By law enforcement | 13 |
| C. By insurance companies | 20 |
| D. By automotive / technology industry | 4 |
| E. In litigation | 10 |
| F. In an increased quantity of litigation | 2 |
| G. In warranty disputes | 1 |
| H. To suspend or revoke drivers' licenses | 4 |
| **III. Consumer Knowledge and Autonomy** | **26** |
| A. Lack of public awareness | 17 |
| B. Lack of consumer choice | 12 |
| C. Infringement of personal autonomy | 3 |
| D. Could cause reluctance to drive while under surveillance | 2 |
| E. Lack of consumer access to recorded data | 2 |
| F. Ownership of device and data | 7 |
| G. Chain of custody for device and data | 3 |
| **IV. Financial Issues** | **41** |
| A. Increase in automotive prices | 35 |
| B. Increase in post-crash repair costs | 4 |
| C. Increase in general maintenance costs | 8 |
| D. Increase in insurance rates | 19 |
| E. Cost of deployment as burden upon taxpayer | 7 |
| **V. Constitutional and Legislative Issues** | **15** |
| A. First Amendment abridgement: speech and movement | 1 |
| B. Fourth Amendment abridgement: search and seizure | 1 |
| C. Fifth Amendment abridgement: self-incrimination | 9 |
| D. Legal priority of recorded vs. eyewitness evidence | 8 |
| **VI. General Functioning of Device** | **42** |
| A. Possibility of malfunction | 22 |
| B. Damage to other systems in a malfunction | 6 |
| C. Damage to device in crash | 5 |
| D. Accuracy or corruption of data | 30 |
| E. Differing degrees of accuracy for different device models | 2 |
| F. Emergency calls for minor crashes could divert resources | 2 |
| G. Device cannot see outside factors at work | 5 |

# Positive Comments

*49 states, plus the District of Columbia and Puerto Rico, have safety belt use laws.*

Photographer: Frank Staples
© 1999 Click, Incorporated, Transportation Safety Technologies

Objective use of the data may improve vehicle systems.

Objective use of the data may improve highway systems.

The data may aid in regulatory initiatives.

The data may aid in alleged defect investigations.

The data may aid in litigation cases.

The data may help in initiatives to improve driver behavior.

The data may aid law enforcement efforts.

May increase driver awareness.

May help determine dangerous traffic areas.

May help engineers design a safer car.

May help gather statistics: seat belt usage, etc.

May lead to decreased vehicle prices.

May lead to decreased insurance rates.

May identify conditions and situations were additional safety devices could be used.

May provide information as to why some crashes are fatal and others are not.

May provide actual crash velocity data in real time conditions.

May reduce the amount of crash testing in labs.

May become so ordinary that owners/drivers will not know/care if it is present.

May help provide quicker emergency response time to crashes.

May provide better understanding as to how a driver responds to a crash.

May provide better understanding as to how occupants in various positions respond.

May provide a better picture of overall crash behavior.

May catch people who intentionally crash cars to collect insurance.

May determine the number of occupants within a vehicle and help cut-down on insurance fraud.

May provide critical information that will determine causes of injuries and fatalities.

Will allow us to better understand automobile crashes.

Will help to determine who was at fault in an accident.

Will make the insurance company's job easier.

Will increase the safety of cars to be built in the future.

Will most likely increase seat belt usage.

Seat belts will save lives if increasingly worn with a sensor.

The speed of the automobile at the time of the crash can be determined accurately, whereas before it could not.

We will have factual information instead of estimated data on police reports.

It may scare drivers resulting in safer driving knowing that they are being recorded.

Insurance reports should be more consistent.

Crashes without eyewitnesses will now have evidence.

Insurance fraud will be less frequent because all the facts of the accident will be on record.

Drivers may maintain safer speeds.

Data can distinguish between two parties that disagree on what really happened.

In crashes, the driver who was not at fault will receive justice, instead of being cheated.

Could help detect defective parts that cause crashes.

May assist doctors in understanding injuries.

May determine if the vehicle systems were all operating at the time of a crash.

Could better determine if the driver was operating the vehicle in a reckless manner.

Could tell if the road conditions were poor.

Make people more aware of their vehicle.

May lead to different occupant restraint systems.

May lead to improved air bag safety.

May determine if children were in-position or out-of-position.

May determine if children were in the front seat.

May help locate stolen vehicles.

May provide an accurate number of daily, weekly, monthly, and yearly crashes in specific locations.

May provide a more realistic number of crashes that actually occur and are not reported.

May be tied-into the defect/recall system of identifying unsafe products.

May help to reduce road rage behavior.

May aid in eliminating habitual drunk drivers from the highways.

May aid in school bus safety.

May provide time of crash.

May provide location of crash.

May provide seat belt usage.

May determine faulty systems.

May signal emergency response.

May cause the driver to drive more cautious and considerate.

May create new industries and jobs:

People to manufacture the box,

People to install the box,

People to inspect & maintain,

People to analyze the data,

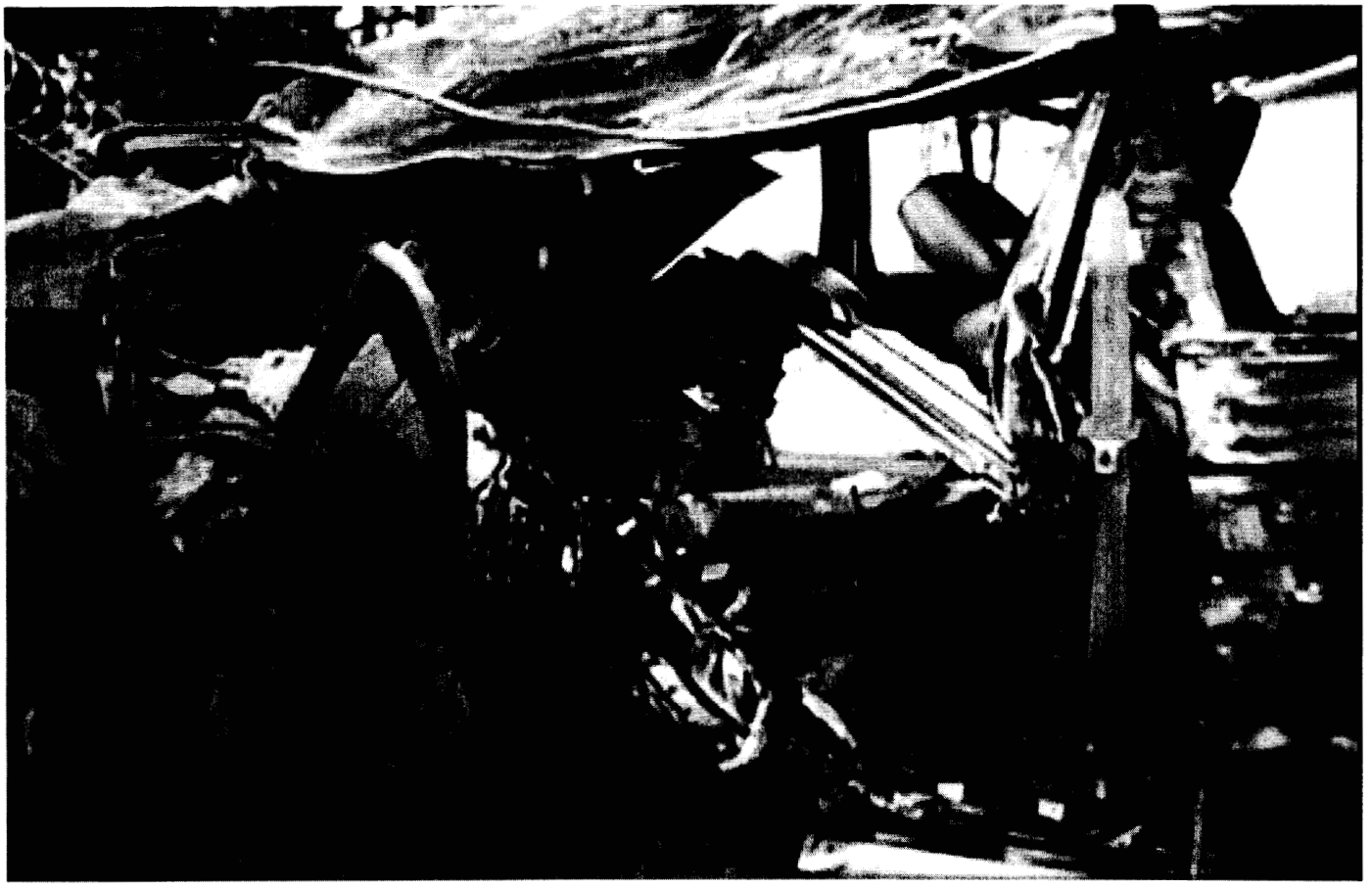
People to use the data when designing future vehicles, making safety standards, etc.

Could aid a variety of medical personnel (doctors, EMT's) in determining injuries the occupants suffered.

The data could be used in your favor and help defend your interests.

Used on a select population of at-risk drivers...say teenagers...if so, it may cut-down on irresponsible driving and save precious lives.

# Negative Comments



*The majority of persons killed or injured in traffic crashes were drivers (64%), followed by passengers (32%), pedestrians (2%), and pedalcyclists (2%).*

Photographer: Frank Staples
© 1999 Click, Incorporated, Transportation Safety Technologies

This technology would make it possible to place private vehicles under continuous surveillance.

This technology could eliminate what little informational and personal privacy remains.

Links to the Global Positioning System of satellites would make it possible to track the whereabouts of private vehicles at all times.

The cost of installing transportation recorders could increase automobile prices.

The complexity of these devices could increase repair and maintenance costs.

The data gathered could be misused by the government.

The data gathered could be misused by law enforcement.

The data gathered could be misused by insurance companies.

The data gathered could be misused in litigation, or could increase the overall quantity of insurance- and crash-related lawsuits.

The data gathered could be misused by automotive or archiving corporations, or in warranty- or drivers' license-related disputes.

The cost of deploying transportation recorders and other emerging technologies will increase the taxpayers' burden.

Personal identifiers could be compiled along with the non-personal data.

There are serious concerns regarding third-party access to the data.

There are serious concerns regarding the security of the data from unauthorized intrusion, access, corruption, or alteration.

There are serious concerns surrounding the proposed permanent archives of transportation recorder data, including:

The security of the archive from unauthorized intrusion and tampering;

The availability of the archive to third parties;

The availability of the archive to government or law enforcement agencies;

The extent and nature of the data to be archived;

The questionable need for such an archive in the first place.

The public is almost completely unaware of the existence and planned deployment of transportation recorder and related technologies.

The consumer has thus far been given no choice regarding the presence of transportation recorders in private vehicles.

There are serious concerns that no choice in this matter will ever be offered.

Transportation recorders and the data they gather could be used to infringe Constitutional rights, including:

First Amendment rights to freedom of religion, speech, and assembly could be abridged if government agencies have detailed knowledge of private vehicles' movements;

Fourth Amendment rights to freedom from unreasonable searches and seizures could be abridged if government and law enforcement have access to transportation recorder surveillance data;

Fifth Amendment rights to freedom from self-incrimination could be abridged by government or law enforcement access to transportation recorder data.

No technology is infallible. Concerns arising from this fact include:

The transportation recorder could malfunction or cease operating;

Electrical or other computer systems in the car could be damaged by such a malfunction;

The transportation recorder could be damaged or destroyed in a crash;

The transportation recorder could provide inaccurate or corrupted data.

There are serious concerns regarding the admissibility of transportation recorder data in litigation, especially in the case of malfunction, inaccurate data, or contradictory eyewitness accounts.

Different transportation recorder models could provide differing degrees of accuracy.

There are serious concerns arising from potential transportation recorder surveillance, including:

The feeling of "being watched" could cause a reluctance to use private vehicles;

The existence of such surveillance could infringe the right to personal autonomy.

The ownership of transportation recorder data, and the chain of custody for such data, is in question, especially regarding a fear that the consumer would be denied access to the data.

It is possible that a black market for falsified transportation recorder data could appear.

Emergency calls relating to minor crashes could divert resources needed for handling more serious problems.

The transportation recorder would be unable to see outside factors at work, including the actions of other automobiles and drivers, and so could not provide a complete picture of a crash.

The data gathered by transportation recorders might have little or no value to victims of automotive crashes, for instance providing no information that would assist doctors or Emergency Medical Teams in caring for the victim.

The data gathered by transportation recorders could be used as a substitute for pre-market crash testing, effectively using the victims of automotive crashes as crash-test subjects, and also leading to lowered vehicle safety as untested automobiles are released to market.

It is questionable whether transportation recorders could have any use except as a substitute for pre-market crash testing, and thus the entire premise could be flawed.

Finally, there are important questions that must be answered before any decision in this matter can be made:

Can we ensure that the personal privacy of drivers will *never* be infringed by this technology?

Can we ensure that the informational privacy of drivers will *never* be infringed by this technology?

Can we so regulate the gathering, archiving, and usage of this data that it will *never* be misused by any entity, whether government, law enforcement, private, or corporate?

Can we be utterly certain that all data gathered and archived is safe from unauthorized intrusion and tampering?

Is there any valid connection between the ability to record this data and the premise that it will in some way make driving safer?

Most importantly, are we responsible enough properly to cope with the enormous power inherent in the ability to observe, track, and archive the movements of all private vehicles and their owners?

# Addenda

*The majority of vehicles in single- and two-vehicle crashes were going straight prior to the crash.*

Photographer: Frank Staples
© 1999 Click, Incorporated, Transportation Safety Technologies

*Automobile Collision Data: An Assessment of Needs and Methods of Acquisition*

February 1975

NTIS order #PB-244867

# OFFICE OF TECHNOLOGY ASSESSMENT

## Automobile Collision Data

AN ASSESSMENT OF NEEDS
AND METHODS OF ACQUISITION

FEBRUARY 17, 1975

PREPARED AT THE REQUEST OF
THE HOUSE COMMITTEE ON APPROPRIATIONS
TRANSPORTATION SUBCOMMITTEE

PREPARED UNDER CONTRACT OTA C11 BY

ECONOMICS & SCIENCE PLANNING, INC.
1200 18TH STREET N.W.
WASHINGTON, D. C. 20036

TECHNOLOGY ASSESSMENT BOARD

OLIN E. TEAGUE, TEXAS, CHAIRMAN
CLIFFORD P. CASE, N.J., VICE CHAIR MAN

EDWARD M. KENNEDY. MASS.
ERNEST F. HOLLINGS, S.C.
HUBERT H. HUMPHREY. MINN.
RICHARD S. SCHWEKER, PA.
TED STEVENS. ALASKA
MORRIS K. UDALL, ARIZ.
GEORGE E. BROWN, JR., CALIF.
CHARLES A. MOSHER, OHIO
MARVIN L. ESCH, MICH.
MARJORIE S. HOLT, MD.

EMILIO Q. DA DADDARIO

Congress of the United States
OFFICE OF TECHNOLOGY ASSESSMENT
WASHINGTON, D.C. 20510

EMILIO Q. DADDARIO
DIRECTOR

DANIEL V. De SIMONE
DEPUTY DIRECTOR

The Honorable George H. Mahon
Chairman
Committee on Appropriations
U. S. House of Representatives"
Washington, D. C. 20515 .

Dear Mr. Chairman:

On behalf of the Board of The Office of Technology Assessment, we are pleased to forward to you the following report on Automobile Collision Data. This study was requested as an evaluation of the automotive crash recorder program proposed by the National Highway Traffic Safety Administration (NHTSA). As the assessment progressed, the implications for automobile collision data as a "whole became apparent and the report has been so titled to provide a more accurate indication of its scope.

This report is being made available to your Committee in accordance with Public Law 92-484.

Respectfully yours,

Olin E. Teague
Chairman
Technology Assessment Board

Respectfully yours,

Clifford P. Case
Vice-Chairman
Technology Assessment Board

## PREFACE

Highlights of the study findings which are especially relevant to the four questions posed by the House Appropriations Committee in its letter of request are summarized below. (The Committee letter is appended).

1. Cost and Adequacy of Current NHTSA Programs

The National Highway Traffic Safety Administration has spent a total of $15.8 million during the last three years gathering and analyzing automobile crash data. The data collected by NHTSA is inadequate to provide a basis for effective safety standard setting or measurement of the benefits of the standards in force. The inadequacies of the system are: too few reports are gathered too slowly; the file is biased toward severe injury accidents; reports do not include adequate quantitative measures of causal severity; and, the information recorded in accident reports is not that which is essential to answering the specific questions of rulemakers, accident researchers and car designers.

2. Use of Existing Crash Recorders

There are 1800 installed (disk-type) crash recorders. These provide a 3-axis acceleration time history over the actual impact interval. This information would probably be adequate to determine crash severity had a severity index been explicitly defined. After the index is defined, these same recorders might be used as part of a specialized crash severity research program.

Currently these recorders provide a limited independent measure of crash severity in air-bag equipped cars. They are also giving NHTSA practical experience in the retrieval, readout and analysis of crash records, the reliability of recorders themselves, and the reactions of fleet owners to crash recorder installations.

3* Improving the Data Base

NHTSA has not provided a sampling plan to support requested appropriations for crash data acquisition programs in the last three years. In order to rectify the inadequacies of the existing data base and the current crash data acquisition system, a comprehensive sampling plan must be developed.

The rate of acquisition of collision reports should be increased to 500,000 to 1,000,000 per year at an estimated cost of $3-10 million annually. Causal severity should be measured and reported. This could be done by using disk recorders at a cost per report of about $133. Alternately, vehicle deformation could be measured and analyzed to determine severity at a cost of about $20 per report. However, if a cheap crash severity measuring device could be developed, it would eliminate the tedious measurement and analysis of vehicle deformation.

The consequences of not getting data are, first, sustaining a continuing societal 10SS of at least $22 billion per year in automobile death, injury and damage without developing adequate tools to correct the problem; and second, imposition of $7 billion to $14 billion in consumer costs for meeting existing, proposed, and planned future motor vehicle safety standards whose benefits will continue to be uncertain.

Current NHTSA programs (multidisciplinary accident investigation, air cushion restraint system evaluation, fatal accident reporting, pedestrian-cyclist accident survey) should be continued. They are necessary to answer specific safety questions. . . . _

4. Further Considerations

If sophisticated tape crash recorders were used, there may be secondary benefits to driver training programs. For example driver errors may be more readily determined and the effectiveness of driver training may be better measured.

If crash recorders are installed, there is the possibility that their readings could be used in liability cases. This matter should be examined more fully in the legislative process.

MAJORITY MC MBERS
GEORGE H. MAHON, TEX.,
CHAIRMAN

JAMES L. WHITTEN, MISS.
JOHN J. ROONEY, N.Y.
ROBERT L. F. SIKES, FLA.
OTTO E. PASSMAN, LA.
JOE L. EVINS, TENN.
EDWARD P. BOLAND, MASS.
WILLIAM H. NATCHER, KY.
DANIEL J. FLOOD, PA.
TOM STEED, OKLA.
GEORGE E. SHIPLEY, ILL.
JOHN M. SLACK, W. VA.
JOHN J. FLYNT, JR., GA.
NEAL SMITH, IOWA
ROBERT N. GIAIMO, CONN.
JULIA BUTLER HANSEN, WASH.
JOSEPH P. ADDABBO, N.Y.
JOHN J. MC FALL, CALIF.
EDWARD J. PATTEN, N.J.
CLARENCE D. LONG, MD.
SIDNEY R. YATES, ILL.
DAN DASEY, TEX.
FRANK E. EVANS, COLO.
DAVID R. OBEY, WIS.
EDWARD R. ROYBAL, CALIF.
LOUIS STOKES, OHIO
J. EDWARD ROUSH, IND.
K. GUNN MC KAY, UTAH
TOM BEVILL, ALA.
EDITH GREEN, OREG.
ROBERT O. TIERNAN, R.I.
BILL CHAPPELL, JR., FLA.
BILL D. BURLISON, MO.

CLIFFORD A. ALLEN, REP. PA., MICH.
WILLIAM C. MINSHALL, OHIO
ROBERT H. MICHEL, ILL.
SILVIO O. CONTE, MASS.
GLENN R. DAVIS, WIS.
HOWARD W. ROBISON, N.Y.
GARNER E. SHRIVER, KANS.
JOSEPH M. MC DADE, PA.
MARK ANDREWS, N. DAK.
LOUIS C. WYMAN, N.H.
ROBERT L. TAYLOR, CALIF.
WENDELL WYATT, OREG.
JACK EDWARDS, ALA.
WILLIAM J. SCHERLE, IOWA
ROBERT C. MC EWEN, N.Y.
JOHN T. MYERS, IND.
J. KENNETH ROBINSON, VA.
CLARENCE E. MILLER, OHIO
CARL D. PURCELL, N.C.
VICTOR V. VEYSEY, CALIF.
LAWRENCE COUGHLIN, PA.
C. W. BILL YOUNG, FLA.

CLERK NO STAFF V 011, DIRECTOR
KENETH P. MAINLAND

TELEPHONE:
CAPITOL 4-3121
EXT. 52771
OR
225-2771

# Congress of the United States
## House of Representatives
## Committee on Appropriations
## Washington, D.C. 20515

November 19, 1974

Honorable Edward M. Kennedy
Chairman
Technology Assessment Board
Washington, D. C. 20510

Dear Mr. Chairman:

On behalf of Congressman John J. McFall, Chairman of the Transportation Subcommitee, and Congressman Silvio O. Conte, the Subcommittee's Ranking Minority Member, I am transmitting the attached request for a technology assessment with regard to automobile crash recorders.

with kindest personal regards.

Sincerely,

George H. Mahon

Chairman

GEORGE H. MAHON, TEXAS
CHAIRMAN

JAMIE L. WHITTEN, MISS.
JOHN J. ROONEY, N.Y.
ROBERT L. F. SIKES, FLA.
OTTO E. PASSMAN, LA.
JOE L. EVINS, TENN.
EDWARD P. BOLAND, MASS.
WILLIAM H. NATCHER, KY.
DANTE B. FASCELL, FLA.
TOM STEED, OKLA.
GERALD R. FORD, MICH.
JOHN M. SLACK, W. VA.
JOHN J. FLYNT, JR., GA.
NEAL SMITH, IOWA
ROBERT N. GIAIMO, CONN.
JULIA BUTLER HANSEN, WASH.
JOSEPH P. ADDABBO, N.Y.
JOHN J. McFALL, CALIF.
EDWARD J. PATTEN, N.J.
CLARENCE D. LONG, MD.
SIDNEY R. YATES, ILL.
BOB CASEY, TEX.
FRANK C. EVANS, COLO.
DAVID H. OBEY, WIS.
EDWARD R. ROYBAL, CALIF.
LOUIS STOKES, OHIO
J. EDWARD ROUSH, IND.
W. GUNN McKAY, UTAH
TOM BEVILL, ALA.
EDITH GREEN, OREG.
ROBERT O. TIERNAN, R.I.
BILL CHAPPELL, JR., FLA.
BILL D. BURLISON, MO.

# Congress of the United States
## House of Representatives
## Committee on Appropriations
## Washington, D.C. 20515

November 19, 1974

MINORITY MEMBERS
ELFORD A. CEDERBERG, MICH.
WILLIAM E. MINSHALL, OHIO
ROBERT H. MICHEL, ILL.
SILVIO O. CONTE, MASS.
GLENN R. DAVIS, WIS.
HOWARD W. ROBISON, N.Y.
GARNER E. SHRIVER, KANS.
JOSEPH M. McDADE, PA.
MARK ANDREWS, N. DAK.
LOUIS C. WYMAN, N.H.
BURT L. TALCOTT, CALIF.
WENDELL WYATT, OREG.
JACK EDWARDS, ALA.
WILLIAM J. SCHERLE, IOWA
ROBERT C. McEWEN, N.Y.
JOHN T. MYERS, IND.
J. KENNETH ROBINSON, VA.
CLARENCE E. MILLER, OHIO
EARL B. RUTH, N.C.
VICTOR V. VEYSEY, CALIF.
LAWRENCE COUGHLIN, PA.
C. W. BILL YOUNG, FLA.

CLERK AND ST. DIRECTOR
KEITH F. MAINLAND

TELEPHONE:
CAPITOL 4-3121
EXT. 52771
OR
225-2771

Honorable George H. Mahon
Chairman
Committee on Appropriations
U.S. House of Representatives
Washington, D. C.

Dear Mr. Chairman:

The Conference Report to H.R. 15405 (Department of Transportation and Related Agencies Appropriations Bill, 1975) states that: "The conference agreement contains no funds for the crash recorder program. The Committee intends to request an evaluation of this program by the Office of Technology Assessment. "

The purpose of this program, as proposed by the National Highway Traffic Safety Administration (NHTSA) , is to assemble detailed data on actual collisions so as to develop realistic automobile design standards. NHTSA proposed the installation of 100, 000 crash recorders in vehicles used in ordinary driving. Total cost of the 5 year program including installation of the recorders and monitoring and analysis of the data was estimated at $14.5 million in 1973. An alternate approach has also been proposed by NHTSA. This entails the controlled crashing of unoccupied vehicles along with computer s emulations o f automobile crashes. The cost of this program has been estimated as approximately the same as the crash recorder program.

Although the committees of both Houses have heard extensive testimony on this program over the past three years, substantial question and differences still exist on the necessity for gathering additional information through the installation and monitoring of the requested crash recorders.

Since this issue remains unresolved, the Conference Committee on H.R. 15405 decided to call upon the Office of Technology Assessment for assistance.

We therefore request that the Technology Assessment Board consider approving an assessment that would address the following issues:

1. How much has NHTSA spent in each of the past three years to gather accident data? Is that data sufficient, or is further data on the characteristics of automobile collisions necessary for effective NHTSA standards-setting? If the existing data base is inadequate; in what ways is it inadequate?

2. An evaluation of the type of data being produced by existing crash recorders and an explanation of how this data is being used by NHTSA should be conducted.

3. If the data base is inadequated, how might an adequate data base be obtained and what are the consequences associated with obtaining the data in different ways (including the possibility of not obtaining the necessary data)? The cost effectiveness of the crash recorder and the crash impact approaches proposed by NHTSA should be examined.

4. Secondary consequences of implementing these or other program should be identified and evaluated. Examples of these secondary consequences include legal questions associated with the existence of actual physical data from an accident and the potential value (to driver training program) of a knowledge base concerning how drivers actually respond in accident situations. For each type of approach investigated, the implementation costs to the Federal Government, industry and consumers should be identified.

We appreciate your assistance in transmitting this request to the Chairman of the Technology Assessment Board.

Sincerely,

(signed)

John J. McFall
Chairman, Subcommittee on
Transportation Appropriations

(signed)

Silvio O. Conte
Ranking Minority Member
Subcommittee on Transportation
Appropriations

## Contents

In this volume:


| | | Page |
|---|---|---|
| 1. | Introduction and Summary | 1 |
| 2. | The Need for Mere and Better Crash Data | 9 |
| 3. | Characteristics of An Adequate Data Collection Program | 26 |
| 4. | Alternatives For An Adequate Data Acquisition Program | 49 |
| 5. | Federal Responsibility and Expenditures For Collision Data Gathering | 64 |
| 6. | Legal Aspects of Crash Recorders | 68 |
| | References | 70 |


Bound Separately:


Appendices

## 1. INTRODUCTION AND SUMMARY

At the request of the House Appropriations Committee, the Office of Technology Assessment, through contract OTA-C11, engaged Economics & Science Planning, Inc. (ESP) to undertake a study of the need for and means to assemble detailed data on actual automobile collisions so as to develop realistic automobile design standards. The study examined the desirability, utility, design and cost of crash recorders and of the alternate approaches to gathering collision data, including computer crash simulation, controlled laboratory crashes and their correlation with observed vehicle deformations, and methods to improve the accuracy of accident investigation reporting and to increase the utility of national crash data files. Specific data collection programs previously proposed to Congress by the National Highway Traffic Safety Administration were studied and evaluated. This report contains the results of this effort.

We have concluded that the current national accident data base is inadequate to resolve the uncertainties in NHTSA's current and proposed motor vehicle safety programs. One of the major deficiencies is data relating collision forces and actual fatalities and injuries. The need has been clearly expressed by Professor B. J. Campbell (University of North Carolina):

> ". . . when one is forced to use nonhuman subjects [in laboratory crashes] then one is left in the situation of knowing a great deal about the physics of the crash but knowing little of the actual injuries that might have occurred in such a crash. On the other hand, in real world automobile crashes one can learn about the actual outcome in terms of survival and injuries, but the input variables mentioned before are unknown.

"The need to link these two systems is apparent. Engineers who design protective systems need to know about stopping distances, forces, decelerations, etc. But knowing these things is of too little help unless one has a way to relate them to real world injuries."

FINDINGS

1. The existing national data base is inadequate

-- only four of 40 existing standards have been shown to be beneficial based on statistical evidence.

-- the nationwide effectiveness of lap belts in mitigating fatalities is still unknown after five years; statistical evidence is available from only one state.

-- there is an immediate need for more and better crash data

- to support rulemaking and to estimate the benefits of proposed safety standards
- to determine the effectiveness of existing safety standards
- to determine causes of accident, injury and fatality to aid crashworthy vehicle design
- to identify new safety problems as they develop
- for predicting the impact of trends in motor vehicle design on accident incidence and outcome

-- Larger crash data collection expenditures than the $5 million to $6 million now programmed annually appear to be justified:

- Motor Vehicle accidents cost society $22 billion to $44 billion annually.

- Present safety standards cost consumers $2.5 billion annually

- proposed and possible safety standards could cost an additional $4 to $12 billion annually.

- Present and planned safety standards add weight to automobiles which increases gasoline consumption.

2. <u>A Comprehensive Accident Data Program</u>

-- must be designed with great care to assure that

- it is representative and avoids inadvertant biases

- it will answer the outstanding critical safety questions

- it is adequate in rate and quantity

- it provides uniformity in reporting and format

-- should be reviewed and approved by a broadly based body of experts before it is implemented.

-- elements for a comprehensive program could include:

- 500,000 to 1,000,000 crash reports per year for a mass data file at a cost of $3 to $10 million per year.

- the measurement and reporting of crash severity either by vehicle deformation measurement or a cheap and widely installed crash severity recorder, at a cost of $10 to $20 million per year.

- some measurement of crash dynamics using some mix of simulated accident reconstruction (SMAC) and collision history (disk or tape) crash recorders at a cost of $2 million to $4 million

- supplementary surveys to answer specific questions and the existing special programs now costing $5 to $6 million per year

- a cheap crash severity recorder at a development cost of about $500,000

- field trials of planned safety improvements whose costs are high and whose benefits are uncertain (as an example, the cost of a field trial of passive restraints would be $30 - $60 million)

3. <u>The Federal Government, not States, manufacturers or insurance companies, should support the central data collision activities.</u>

-- It is a national problem.

-- The Motor Vehicle Safety Standards are promulgated by the Federal Government.

-- The data has to be obtained in an unbiased and uniform manner throughout the nation.

-- The Federal Government has the resources and ready access to the sources of information.

4. Crash recorders provide data that may be admissible in a court of law.

5. Program alternatives include the following:

- Doing nothing to improve the current crash data acquisition system. If this course is followed, $22 to $44 billion in societal losses will continue to be incurred each year without developing adequate tools to analyze and correct the problem; $7-14 billion or more in consumer costs will be imposed yearly by current, proposed and advanced motor vehicle safety rule making whose benefits, in most cases, will continue to be uncertain.

- Upgrading current data collection programs without adding a mass data acquisition system. This course will neither provide statistically convincing measures of the reduced incidence of death or injury resulting from incorporation of safety features nor will it give a timely response to questions regarding the impact of vehicle design changes.

- Providing a mass accident data acquisition program at a cost of $3 to $10 million yearly. This course will begin to permit timely statistical determination of safety system benefits and identification of automotive safety problems. However, crash severity measures will be inadequate and it will be difficult to associate injury with crash severity.

- Upgrading mass accident data acquisition program to provide accurate severity reporting at a cost of $10 to $20 million annually. This action would finally provide timely determination of safety benefits with ascertainable accident severity incidence and associated injury and fatality exposure bridging the gap between laboratory and field experience.

- Use of acceleration time-history (disk) recorders. A small (10,000 to 20,000 recorders; $2-4 million) program will permit: generating baseline statistical information such as severity distribution of all collisions; the calibration of vehicle deformation estimates as a severity measure; and calibration of computer simulated crash reconstruction (SMAC). A program as large as large as 100,000 disk recorders -- $10 million -- would overdo it from the standpoint of research and be inadequate from the standpoint of mass data gathering.

  Development of a cheap and proliferable causal severity measurement device at an estimated development cost of $500,000 and a production cost of approximately $2 per unit will provide a device capable of widespread installation that permits ready read out of crash severity magnitude and direction by an untrained investigator. The need for careful deformation measurement and transformation of these measurements to equivalent barrier speed would be eliminated.

  providing a federally sponsored field trial of uncertain and/or expensive safety aids. This program will permit the evaluation of safety aids, where normal market forces do not operate, prior to their being mandated on a national scale. (In the case of passive restraints, the one time cost would be $30 - $60 million. )

This study was accomplished by an extensive literature survey; by independent analysis by members of the ESP staff; by analysis of specific assigned topics undertaken by knowledgeable members of the automobile accident research community; and through an Automobile Collision Data Workshop, convened January 16 and 17, 1975, at which the requirements for, and various approaches to, better collision data gathering were presented and discussed in depth by experts in all aspects of the problem. Individuals who participated in the Workshop were the following:

| | |
|---|---|
| Lynn Bradford | National Highway Traffic Safety Administration |
| Paul Browinski | AVCO Systems Division |
| B. J. Campbell | Highway Safety Research Center University of North Carolina |
| Charles Conlon, Jr. | AVCO Systems Division |
| J. Robert Cromack | Southwest Research Institute |
| John Edwards | Ford Motor Company |
| M. D. Eldridge | National Highway Traffic Safety Administration |
| Vincent J. Esposito | National Highway Traffic Safety Administration |
| William Fitzgerald | AVCO Systems Division |
| John Garrett | Calspan Corporation |
| Howard P. Gates, Jr. | Economics & Science Planning, Inc. |
| Lawrence A. Goldmuntz | Economics & Science Planning, Inc. |
| Walton Graham | Economics & Science Planning, Inc. |
| James Hofferberth | National Highway Traffic Safety Administration |
| John F. Hubbard, Jr. | Center for Auto Safety |
| Paul R. Josephson | Center for Auto Safety |
| Charles Kahane | National Highway Traffic Safety Administration |
| Edwin A. Kidd | Calspan Corporation |
| Phil Klasky | Teledyne Geotech |
| Gene G. Mannella | National Highway Traffic Safety Administration |
| Don Mela | National Highway Traffic Safety Administration |

| | |
|---|---|
| Charles A. Moffatt | National Highway Traffic Safety Administration |
| David Morganstein | Center for Auto Safety |
| James O'Day | Highway Safety Research Institute University of Michigan |
| Brian O'Neill | Insurance Institute for Highway Safety |
| L. M. Patrick | Wayne State University |
| Steven J. Peirce | National Highway Traffic Safety Administration |
| Louis W. Roberts | Transportation Systems Center, Department of Transportation |
| A. J. Slechter | Ford Motor Company |
| John Versace | Ford Motor Company |
| Richard Wilson | General Motors Safety Research and Development Laboratory |

We wish to acknowledge our gratitude to these individuals not only for their participation in the Workshop, but for their continuing assistance during the study effort and preparation of this report.

## 2. THE NEED FOR MORE AND BETTER CRASH DATA

The following paragraphs will discuss the general objectives of crash data collection, identify some specific data needs that are not now satisfied, and point out serious inadequacies in the current data file and acquisition systems. It will be shown that these needs and limitations lead to a requirement for mass acquisition of crash data, supplemented by special surveys and large scale real-life experiments.

### a. THE OBJECTIVES OF COLLISION DATA COLLECTION

The cost to society of automobile death and injury is conservatively estimated[2/] at $17 billion annually. The vehicle damage adds at least another $5 billion yearly[3/]. The total, $22 billion per year, corresponds to an average of $2200 in losses per each U.S. automobile during its lifetime.

The specialists in auto safety have, as their concerted objective, the reduction of this enormous waste. A body of collision data is needed that will provide a substantial part of the means to determine the causes of accidents, of injuries, and of damage.

Professor Lawrence Patrick of Wayne State University expressed the consensus view of the Workshop participants as follows:

"PREMISE

1. The only valid way to establish safety needs for automobiles is through examination of field data.

2. The only valid way to evaluate the effectiveness of safety measures is through analysis of their effect on accident data.

CONCLUSION

Accident data are essential."

The National Highway Traffic Safety Administration is responsible, under the National Traffic and Motor Vehicle Safety Act of 1966,* for the promulgation of Federal Motor Vehicle Safety Standards to which vehicles manufactured for sale or use in the United States must conform. Under the Motor Vehicle Information and Cost Savings Act (1972)** the Secretary of Transportation is also responsible for setting standards for damage-limiting bumpers and for evaluating automobile damageability and crash-worthiness.

Safety standards put into effect to date cost the consumer about $2.5 billion annually[4/] and standards proposed will cost another $4 billion or more each year[2/, 4-/]. In addition, standards suggested in Advance Notice of Proposed Rulemaking would cost $4 billion per year in first costs plus another $4 billion in added fuel costs when fully implemented. While the more than 40 existing standards , which were based on intuition, judgment and limited experience, are believed to yield in the aggregate a societal benefit greater than their consumer cost,[2/] only four of them (seat belts, energy absorbing steering column, HPR glass and head restraints) have been shown by any authority to be beneficial based on convincing statistical evidence. The problem is that the body of data is inadequate.

Thus an initial objective of crash data collection and analysis from the standpoint of the Government rulemaker, is that of evaluating the efficacies of the existing standards to determine which should be kept on the books and which should be eliminated.

---

* Public Law 89-563.

** Public Law 92-513.

A second objective from the standpoint of rulemaking is that of providing the necessary statistical support to estimates of benefits of a projected safety or damage-limiting standard. In the next section there will be discussed a projected rule that is controversial because of inadequate supporting data.

A third objective is the early identification of problem areas in automobile damage and injury so as to permit designing effective motor vehicle and highway safety programs.

The foregoing objectives from the standpoint of rulemaking have their parallel from the standpoint of the automobile manufacturers. C. Thomas Terry of General Motors has summarized [8/] the objectives of gathering accident data in the field:

a. Evaluation of production safety systems.
b. Prediction of performance of proposed safety systems.
c. Identification of problem areas and evaluation of proposed solutions on a cost/benefit basis.
d. Estimation of human tolerance to impact.

Automobile manufacturers are , of course, vitally concerned with the relative merits of specific alternative designs as well as with the validation of Safety Standards to which they are required by law to conform.

A number of universities and institutes, both profit and non-profit, have been for years involved in research in accident causation, injury causation and designs of vehicles and roads that will reduce accidents and injuries. They need accident data to discover causes of accidents and injuries; armed with this information they can accomplish and test in their laboratories design modifications and provide valuable advice to NHTSA and automobile manufacturers.

Finally, there is a need for national planners to predict the impact of new trends in automobile designs. Fuel and resource conservation programs, encouraged if not mandated by the Federal Government, will lead to lighter, lower power-to-weight ratio automobiles. Data on collision frequencies and outcome are needed as a function of these parameters to inform Federal officials.

b. UNSATISFIED NEEDS FOR CRASH DATA

The body of specialists concerned with automobile collisions -- the rulemakers, safety researchers, accident statisticians, car designers, insurers, and public interest people -- overwhelmingly agrees that there is a grave and compelling need for more and better crash data. The need is expressed by Dr. Edwin A. Kidd of CALSPAN Corporation[1/] in the following way:

> "It is essential that NHTSA have a data bank for surveillance and effectiveness studies related to the impact of standards on accident, injury and fatality frequencies. The relatively small output of the special federal teams and/or the higher quantity, but low content State data banks are inadequate for the purpose. In addition to information on the general accident environment, vehicle damage and occupant injuries, details of the impact environment -- velocity at impact, change in velocity during impact and possibly, vehicle deceleration -- are required for a sample of 100,000 to 500,000 automobiles annually."

Professor B. J. Campbell, Highway Research Center, University of North Carolina[10/], states:

"In acquiring automobile accident data several approaches are used in the U.S. : First, are intensively investigated accident crashes of which several thousand have been collected. The advantage of this approach is that the cases are extremely detailed with photographs and good injury data. The most important disadvantage is that by virtue of the changing sampling criteria and the small sample size, the ability to generalize these few cases to the population is restricted heavily.

I believe too much reliance has been made on this type of data for guiding NHTSA decisions. It leads one to situations in which too much is made of a small number of cases."

The critical need for better collision data to support rulemaking can be illustrated by the passive protection provisions of Motor Vehicle Safety Standard 208. Estimates of the cost to consumers of meeting passive protection requirements range from [2/], ~/ $220 to $400 per car, or a gross cost of $1.5 billion to $3 billion per year more than belt restraints now cost. There is also significant uncertainty in the incremental benefits that may be realized from passive protection. Estimates range from 3,000 to 8,900 more deaths prevented, and from 130,000 to 492,000 more injuries prevented.

One crucial lack of data leading to uncertainty can be pinpointed: the number of lives saved and injuries prevented by a restraint system in frontal collisions is estimated by NHTSA from a graph showing the percentage of injuries and deaths as a function of "equivalent barrier test speed."* This graph is shown in Exhibit A (Figure 4). The "equivalent barrier test speed" is that speed which would produce as much car damage, when the car is driven into a rigid barrier, as the car suffered in an actual collision.

The fatality curve of Figure 4 is based on judgment estimates of barrier equivalent speed of 51 fatal frontal collisions by General Motors and a small (unstated) number by Ford Motor Company; in Figure 3 of Exhibit A the NHTSA curve is replotted for comparison with the companies' judgment data.

In making an estimate of the fraction of lives saved by a restraint system, NHTSA attributes to the system a barrier equivalent speed below which it is effective and above which it is not effective (a conceptual convenience). On the basis of laboratory crashes with dummy and cadaver occupants, lap belts are taken as effective to 25 mph, lap-shoulder harnesses to 30 mph, and air-bag passive restraints to 35 mph.[6/] The intersections of these speed lines with the fatality curve of Exhibit A, Figure 4, then yield NHTSA's estimate of fraction of lives saved in frontal collisions. For example, the intrinsic effectiveness of the lap-shoulder harness in preventing fatalities in frontal collisions is thus deduced[6/] to be 37%, and for all collisions (of which frontals constitute 50%), is estimated at 31%. Yet extensive field experience in Sweden shows lap-shoulder harnesses have an overall fatality prevention effectiveness of 90%. The lap belt alone is estimated by NHTSA to have intrinsic fatality prevention effectiveness of 20% in frontal collisions, with 22% for all collisions. Yet extensive field experience from North Carolina indicates an overall fatality prevention effectiveness with lap belts of 75%.

---

* Technically, these curves are cumulative distribution functions for barrier equivalent speed for fatal collisions and injury collisions.

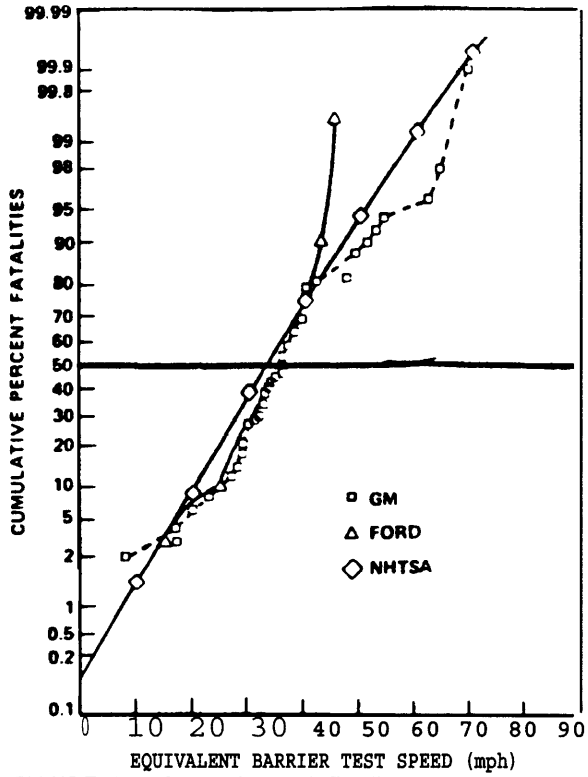



*FIGURE 3 - Comparison of Fatality Distribution Data (Frontal Collisions)*

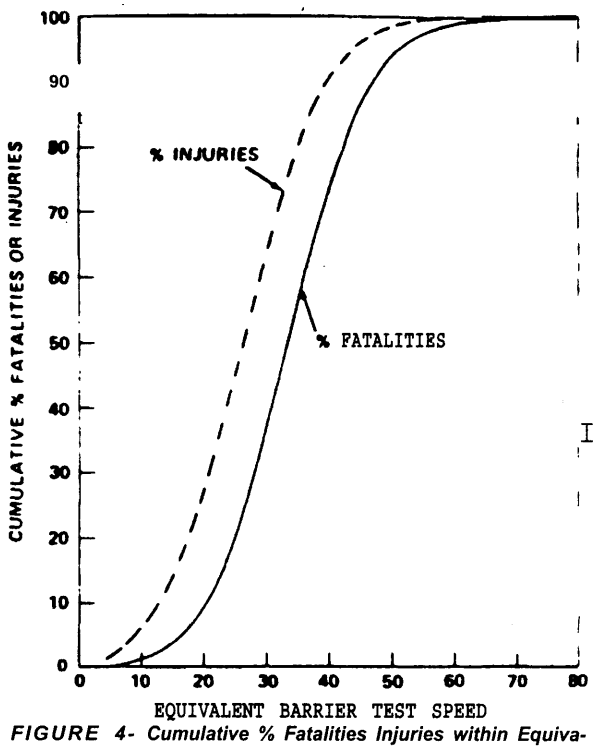



*FIGURE 4- Cumulative % Fatalities Injuries within Equivalent Test Speed Range*

EXHIBIT A

Curves from NHTSA's "Passive Protection at 50 Miles Per hour"

These discrepancies can be explained in three principal ways, any of which may be correct: 1) The Swedish and North Carolina experience is not representative of the population of U.S. car collisions; 2) The barrier equivalent speeds up to which restraint systems are effective are underestimated by NHTSA; or 3) The barrier equivalent speeds at which fatalities occur were over-estimated in the original material of Ford and General Motors.

All of these questions can be resolved by more and better data.

The uncertainty about these curves as a basis for rulemaking is confirmed by National Highway Traffic Safety Administrator James Gregory in Congressional testimony:

> " . . .we have gone out on an advanced notice of proposed rulemaking at the same time that we went out with the passive restraint notice to say that we are moving in the direction of a standard for occupant crash protection at the level of 45 to 50 miles per hour. We figure when we get there we will have pretty much attained what is cost effective and technologically feasible in today's world.
>
> "We feel, by the way, that this would still be worthwhile doing. Yet, as we move toward that, <u>without quantitative data, without persuasive data</u>, even in the public interest, <u>without being able to substantiate a standard</u> we feel is reasonable and in the public interest, the challenge would be sufficient to provide that type of occupant protection. . .
>
> "...The reason I have to be rather vague about this is that most curves that have been derived by experts and <u>from data that have been collected get very fuzzy</u> when you get much above 40 miles an hour as far as what percentage of the fatalities occur at these particular speeds.*

---

* Excerpts from Dr. Gregory's testimony before the Transportation Subcommittee of the Committee on Appropriates, House of Representatives, 93rd Congress 2nd Session 1974, Part 3, pp. 41 - 43 [emphasis ours].

> ...To establish crashworthiness, we need to know what to do to an automobile and what we need to do to the occupants from the standpoint of restraint protection under a given crash condition. <u>These precise data we now lack</u>. . .
>
> "At the present time we cannot make a judgment with accuracy and that makes us guess. And those guesses could cost, unnecessarily as far as the consumer is concerned, untold millions of dollars for protection that we may actually not need. . ."**

The doubts the Administrator expresses about the curves at speeds of 40 mph and above, we believe, as indicated earlier, also should apply to speeds lower than 40 mph.

The kinds of information needed to mitigate much of the uncertainty about the prospective incremental benefits of passive restraints are, first, a file of representative collision data from which it is possible to derive the incidence figures for injury and fatality of belted occupants, in order to establish as a baseline the capabilities for the current belt restraints; second, results of a large-scale field experiment to establish the relative capabilities of passive restraints; and third, representative files of fatal and injury collisions (involving unrestrained and restrained occupants) for which causal severity magnitudes such as BEV have been quantitatively established. With this information the lifesaving and injury prevention potential of restraint systems and the speeds to which the systems are effective can be established.

---

** Excerpts from Dr. Gregory's testimony before the Senate Committee on Appropriations (Hearings on FY 1974 supplemental appropriations, HR 11576) 93rd Congress, first session, part 2, pp. 1509-1510. [emphasis ours]

Fundamental to the statistics of accidents are the cum ulative probability distribution functions of severity for all accidents, for injury accidents, and for fatal accidents. These, though badly needed, are not now being obtained from large quantities of real-life accident data. In order to establish them, measurement and reporting of causal severity is required.

C. LIMITATIONS OF THE CURRENT DATA SYSTEM

In a later section we address the question of collision data requirements. The basic needs can be summarized as follows:

(1) The data should be representative of the population of u. s. automobile crashes.

(2) The data should be gathered in sufficient quantity to be useful, at a sufficient rate to be timely.

(3) The data should be in adequate detail and precision to permit its analysis to determine causes of accidents, injury and death (and the functional relationships between these causal factors and the probabilities of accidents, injury and death) ; and to permit answering questions that may arise relative to traffic safety and motor vehicle safety standard efficacy.

The inability of the current files to meet each of these needs is expressed by several investigators.

O'Day of the Highway Safety Research Institute, says:[9/]

"A random sample is the best way of insuring represen-nativeness. Unfortunately, no random sample of United States crashes exists."

Kidd 15/ comments:

> "For too long, those concerned with accident studies of the effects of safety standards already in force have had to make do with either too small samples of reasonably good data or relatively large samples of data whose content is inadequate for the purpose. In the first category is the data bank (and "bank" is too grandiose a term) that has resulted from the individual federal teams of multidisciplinary, professional investigators. These teams can serve useful purposes in special studies, in discovery of problems that would otherwise go undetected and, particularly, in the area of accident causation. By their very nature, they cannot provide a sufficiently large data sample relevant to the implementation of standards aimed at injury and fatality reduction without excessive expenditure of funds."

MDAI -- Multidisciplinary Accident Investigation[14/] -- is conducted by about 20 teams scattered throughout the country and sponsored by the National Highway Traffic Safety Administration and the Motor Vehicle Manufacturers Association. These teams have been performing clinical in-depth studies (both on-scene and off-scene) of selected accidents in the United States, primarily on new cars, since 1969. The accidents selected for data collection have been strongly influenced by the specific interests of the individual teams. Although the information gathered is accurate and detailed, only about 6,000 cases have been investigated and 2,500 of these have entered the computerized file in the five years since the program started. The MDAI favors accidents in which there was injury or severe damage or in which there were large disparities between the degree of damage and the degree of injury; as a consequence, there is significant bias in the file. B. J. Campbell states,[10/] "I believe too much reliance has been made on this type of data for guiding NHTSA decision. It leads one to

situations in which too much is made of a small number of cases." According to Marie Eldridge of NHTSA, "As a system for producing statistical information needed for supporting our safety standards, the on-scene in-depth investigations cannot be regarded as cost effective. The average cost per case is about $2,000. The cost decreases to about $800 per in-depth case if the on-scene investigation requirement is eliminated." Moreover, as indicated by O'Day, "The present collection of MDAI cases is a sample of an undefined and relatively undefinable population, thus limiting severely the capability to draw inferences to the national accident picture."

A program that has long been established but only recently has become operational is "FARS" -- the Fatal Accident Reporting System.[16/] This system involves NHTSA collection of state data on all fatal accidents, with recording into a uniform format that will permit central storage, retrieval, sorting and analysis. Police data plus later medical reports are included. Reports are made on each occupant, each vehicle and each accident, so that about 200,000 reports are expected to enter the file yearly. Since the file will cover all and only fatal accidents, it will be representative, but only of fatal accidents. Without supplementary information from a sample.of all accidents whose intrinsic severity distribution is the same as that for the fatals, inferences cannot be drawn as to, for example, whether sobriety or use of belt restraints affects the incidence of fatalities in crashes.

A much more representative collision data sample, structured to meet limited objectives, is being collected by NHTSA.[14/] From five selected regions of the country "Level II" data is being obtained on new cars in tow-away involvements for the purpose of evaluating active and passive restraint systems. Information is assembled from the police report, a doctor's report, photographs, a brief vehicle investigation, and driver interviews. Data is collected on all

occupants, whether injured or not, but information gathered is limited to that needed for the statistical analysis of restraint system effectiveness. The design of the sampling process was accomplished centrally, by NHTSA, so that the process will be free of the biasing influence of the investigators (a serious problem in MDAI investigations) . The cost is about $100 per crash. The sampling plan has been designed in such a way that NHTSA expects to be able to make national estimates based on post-stratification.

NHTSA has under development a system for sampling pedestrian and bicyclist accidents in several hundred localities. This is a "bilevel" investigation effort in which there is a supplementary investigation carried out by police (with the added costs borne by NHTSA or others) to establish the nature and location of the accidents and factors affecting visibility. It will answer questions at the level of detail needed to determine gross behavior and counter-measures.

The States, of course, collect accident reports in great number. The reporting thresholds vary from State to State. Within a State, sampling may not be representative or uniform. For example, a city with a high crime rate may devote little effort to investigating and reporting traffic accidents, while even the slightest crash may be reported in smaller towns. Efforts by the NHTSA to use collision data files directly from the States have proved unsuccessful primarily because of the nonuniformity of reports and the consequent inability to properly combine, analyze and process the information. A second problem related to the sheer volume of records that was derived from the States.

On review of the information required on HS Form 214 used in the Fatal Accident Reporting System (FARS) we observe that certain information critically required by both rulemakers and injury researchers is not supplied by the reporters. Specifically, provision of vehicle crush measurements that could be converted to Equivalent Barrier Impact Speed (EBS) using the method of K. L. Campbell[20/] would make possible construction of the cumulative distribution function of EBS in fatality accidents, a function needed by the rulemakers in analysis and prediction of the effectiveness of restraint systems. Provision of information on the vehicle interior points of impact, occupant's height and weight and more detail on the precise nature of injuries suffered by injured and killed occupants would provide vital injury cause information.

It is clear from the foregoing that there is no existing national crash data collection program that is designed to meet national needs. As indicated earlier, NHTSA has contracted with the Highway Safety Research Institute of the University of Michigan to design a national accident data sampling system based on a probability sample. NHTSA hopes that through control of the selection of accidents that a sample can be acquired whose characteristics can be generalized to the national crash population.

d. MASS ACCIDENT DATA ACQUISITION

In summary, to meet data needs and to overcome the limitations of the current national data files and collection systems, a mass accident data acquisition system is needed. In addition, measurement and reporting of accident causal severity is important to the classification and analysis of accidents and

often can be important to drawing credible inferences as to the projected benefits of proposed safety standards. The following chapter will discuss the problems of design of the data acquisition system and of measurement of causal severity in more detail.

The need for more and better data does not mean the current data collection programs should be abandoned. However, each of these programs should be reviewed as to its specific objectives and upgraded as necessary to meet them. For example, MDAI team investigations should conform to a sampling plan rather than being entered into to satisfy the personal interests of the investigators. An effort should be made to get causal severity information and information on injury mechanisms into FARS reports.

An extremely important characteristic of the Fatal Accident Reporting System that might be overlooked as "just a detail" is that it provides uniformity in the reporting from all states, using computerized forms. This uniformity makes it possible to combine, sort and analyze data. Extension of this uniformity to general accident reporting systems used by states would enormously simplify the central collection and analysis of mass accident data, and should be encouraged through a system of incentives.

Even with a very good mass accident data acquisition system in being and operating, it will not be possible to answer certain questions that were unanticipated at the time the system was designed. Supplementary data acquisition systems will be needed to answer such questions; the restraint system

collection system and the pedestrian cyclist system now operating are examples of systems designed and needed to answer specific questions at this time.

Mass accident data acquisition may not, by itself, answer questions with regard to the benefit of a projected safety standard. When the costs of such a standard are large, or the benefits uncertain, it may be necessary to undertake a large scale experimental program to provide the needed answers.

Section 3, following, is necessarily quite technical. However, much of the discussion is summarized in the introduction to Section 4. Readers more interested in the various alternatives for remedying deficiencies in the existing data may wish to proceed directly to Section 4.

## 3. CHARACTERISTICS OF AN ADEQUATE DATA COLLECTION PROGRAM

In Section 2 the general needs of an adequate accident data collection program have been identified and the inadequacies of the present system have been presented. In this section, three characteristics of a satisfactory data collection program are discussed: the quantities and rate of data acquisition, the importance of an unbiased sampling plan and the measurement of causal crush severity.

### a. QUANTITIES AND RATES OF DATA COLLECTION

It is reasonable to require the data collection system to provide timely evaluation of the effects of automobile design changes, whether voluntary or made in compliance with official safety standards. This suggests that the national data collection system should be designed to gather vital information within a single year.

As Kidd points out,[15/] ~Definition of the total number of accident cases required annually for an adequate national data bank can be made if (1) the questions to be asked of the system can be identified both for the present and future; (2) the accuracy with which the particular data elements can be measured is known or can be appropriately examined; and (3) the statistical analysis techniques to be employed can be agreed upon."

But rate depends also on the speed with which results must be realized. Rapid feedback from the field is essential to the evaluation of the effectiveness of changes, so as either to reinforce the decision made by the designer or rulemaker or to dissuade him from an erroneous decision.

In the case of general accident statistics, the population of crashes does not represent the statistically stable ideal (stationary time series) because of continually changing mixes of car sizes and weights, changing rules under which cars are operated (for example, the Federal 55 mph speed limit) , changes in the quality and extent of highways, variation from season to season and year to year in total miles driven, and modifications to vehicle designs, both voluntary and in compliance with safety standards.

The allowable lag in production of statistics, based on the foregoing considerations, appears to be about one year. This, in turn, suggests that a sufficient body of data should be gathered within one year to detect differences in injury incidence as a result of actions on the part of the government or the carmakers.

In the following paragraphs we will estimate what this may mean in terms of the number of reports required per year and, if causal severity were to be obtained through the use of crash recorders, the number of crash recorder installations that would be needed. Some less important data might be acquired over longer periods, lessening the amount of data required annually.

We have previously indicated that one objective of collision data gathering is the construction of cumulative distribution functions for severity for all accidents, all injury accidents, and all fatal accidents. The first of these is needed to provide reference or baseline statistical information from which other important statistics may be derived; the second and third are needed to validate the rationale used in rulemaking. A statistical technique* permits prediction of the number of

---

* The Kolmogoroff-Smirnov test; see, for example, "Non-parametric Statistical Inference." J.D. Gibbons, McGraw Hill 1971.

observations in a random sample that would be required to construct these distribution functions with a confidence of xpercent that the function derived from the sample will be within Y percent of the true distribution. Table 1 tabulates the number of samples required for several levels of confidence and accuracy.

Table 1

Number of Observations Required
To Construct Cumulative Distribution Functions

| Deviation From "Truth" | Confidence Level | | |
|---|---|---|---|
| | 80% | 90% | 95% |
| 1% | 11,449 | 14,884 | 18,496 |
| 2% | 2,862 | 3,721 | 4,624 |
| 3% | 1,272 | 1,653 | 2,055 |
| 4% | 716 | 913 | 1,156 |
| 5% | 458 | 595 | 740 |
| 8% | 179 | 233 | 289 |
| 10% | 115 | 150 | 185 |

The table indicates the number of reports that would be required to construct distribution functions of severity if severity could be measured for each year.

The tabulated numbers represent also the number of reports needed in a segregated category to construct a severity distribution function for that category. Taking a typically acceptable statistical level of 95% confidence, 5% accuracy, 740 fatality reports would be required to construct a severity distribution function for fatalities; 740 injury reports would be required to construct severity distribution function for injury cases. Suppose it were desired to examine the distribution function for car weights in injury cases, independent of all other factors; again, 740 reports would be required in which weight was stated.

The need for a large number of annual reports arises when a particular set of events to be examined has low probability of occurrence in the sample. Suppose, for example, one wishes to determine the distribution of car weight in rollover injury accidents for two categories of occupants: belted and unbelted, 740 reports in each of the two categories would be required. Injury accidents constitute 33% of reportable accidents, and the probability that an injury accident was a rollover [3/] is about 8%. Perhaps 25% of those injured wore belts. Thus 0.67% of reportable accidents were rollover-injury-belted, and to find a sample of 740, an aggregate of 111,000 reports in the 'reportable accident" category would be required. (This same set of reports would provide more than enough unbelted-rollover-injury events.) If only injury accidents were reported, a sample of 37,000 reports would suffice. If the same analysis were to be done for fatal rollover accidents drawn from a mass accident file, the file would have to number 3,500,000 to find 740 fatal-rollover-belted events. The reason for the much larger data file in this case is that there are far fewer fatalities than injuries.

---

* $0.25 \times 0.08 \times 0.333 = 0.0067$.

Analysis of infrequent events requires many input reports. But the fact that events are infrequent does not make them unimportant. The best example of this is traffic fatalities, which, though infrequent, cost society almost as much as automobile injuries and damage combined.

Suppose that a new restraint system modification were implemented, and one wished to confirm, to a confidence level of 95%, that it reduced the incidence of occupant fatalities in the population of all accidents by 10% over the old restraint system.* Assuming the old system had a (perfectly known) fatality rate (when used) of 0.06%. We are seeking to verify that the new restraint system gives a fatality rate of 0.054% or less. The use rate on the new restraint system is expected to be 50%. An upper bound on the number of accident reports required to determine the fatality incidence to the desired accuracy is found to be 768,000. If this were to be accomplished in the first year Of the new installation, reports would be needed on about 30% of all accident involvements of new U.S. automobiles. Clearly, reports on fatal accidents alone would not be useful, as fatality incidence could not be determined.

The foregoing calculation makes use of an expression for the number of samples n required to determine with accuracy $\sigma$ a proportion p in the population from which the sample is drawn, namely:

$$n = \frac{p \quad (1-p)}{\sigma}$$

Clearly, if the same question were restricted to side impact accidents a sample of 768,000 side impact accidents would be needed, but since side impacts constitute 1/6 of all accidents and were drawn from a sample of all accidents, that sample would have to number 4.6 million.

---

* A practical example of the kind of question NHTSA and safety researchers seek answers to.

One can now see, from the examples given, the extent to which numbers of reports required depend on the questions asked. Efficient sampling to minimize the number of samples requires a basic set of questions to provide baseline statistics with supplementary surveys to obtain the answers to specific questions.

Based on the previous examples of questions that might be asked of an accident file, we believe that 500,000 to 1,000,000 cases per year, collected in accordance with a carefully designed sampling plan, is needed by NHTSA and others.

We determine now the number of crash recorders that would be needed to determine accident severity distributions if recorders were the chosen technique to measure accident severity. The number of recorders required depends on the probability occurrence of the type of collision. About 7.5% of all cars are involved in reportable accidents, 2.5% in injury accidents, and 0.04% in occupant-death accidents each year.

Table 2 indicates the number of recorders required to get the needed data each year to construct severity distribution function curves to 5% accuracy (5% corresponds to approximately 2 mph in estimate of barrier equivalent impact speed) . The figures in the column headings are the probabilities that a recorder equipped car will be involved in an accident of the type indicated; 100% recovery of recorder data is assumed. 30% of involvements are considered to be of "reportable" severity: that is, that the damage to the vehicle is of sufficient extent, or that there is an injury, either of which would require reporting the accident to police.

Table 2

Number of Recorders Required to Secure in One Year Data Needed to Construct Severity Distribution Functions to 5% Accuracy

| Confidence Level | Accidents Above a "Reportable" Severity Level P = 0.075 | Injury Accidents of All Types P = 0.025 | Fatal-to-Occupant Accidents of All Types P = 0.0004 |
|---|---|---|---|
| 80% | 6107 | 18,320 | 1,145,000 |
| 90% | 7933 | 23,800 | 1,487,500 |
| 95% | 9867 | 29,600 | 1,850,000 |

If it were further required to construct these distributio functions for smaller classes of accidents (frontal, side, rear, rollover) the number of recorders required, for 90% confidence and an accuracy of 5%, would be as shown in Table 3. (Based on accident type probabilities given in references 3 and 6.)

Table 3

Number of Recorders Required to Secure in One Year Data Needed to Construct Severity Distribution Functions With 90% Confidence of 5% Accuracy

| | Accidents Above a "Reportable" Severity Level | Injury Accidents | Fatal Accidents |
|---|---|---|---|
| Frontal | 16,190 | 64,324 | 2,917,000 |
| Side | 46,665 | 58,048 | 5,313,000 |
| Rear | 27,355 | 170,000 | 29,750,000 |
| Rollover | 198,000 | 297,500 | 9 ,297,000 |

As the cell size becomes smaller -- that is, as the data is subdivided into more and more classes of interest -- the number of reports needed in each cell for the construction of the particular distribution function of severity remains the same; but the number of recorders required to assure that required number of reports in each cell increases rapidly. Clearly, either a very large number of recorders would need to be installed in the U.S. automobile fleet, perhaps one in each car, or alternate methods of obtaining a measure of severity, such as measuring structural deformation of the automobile, should be used.

If a very cheap (say, $2) crash recorder does not become available, then it is clear that crash recorders become impractical because of costs as a means of measuring severity for mass accident data files, which are needed to evaluate events of low probability yet events of great importance.

b. THE NEED FOR DEFINITION, MEASUREMENT AND REPORTING OF CAUSAL CRASH SEVERITY

Throughout earlier sections of this report, reference has been made to accident severity. It is important to note that what is meant is intrinsic or causal severity, as opposed to the severity of the outcome of crash, such as the degree of injury or damage. As indicated earlier, selection of a sample based on outcome inherently biases the sample and masks the effects of design changes. What is needed, instead, is a bank of data that will permit determining, for a given causal severity or range of causal severities, the outcome as a function of other factors -- car weight, occupant age, passenger compartment design, etc.

For example, in establishing bumper standards, it would be useful to know, first, the probability distribution for causal crash severity and second, the relationship between costs to repair car damage and the severity of the collision in the absence of damage limiting bumpers. From this information could then be predicted the gross benefits of new bumpers that prevented damage in accidents up to a specified severity level.

In determining the efficacy of an existing motor vehicle safety standard for occupant protection, it is important to be able to establish how the probability of injury (or degree of injury) is affected by meeting the standard. This implies a need to develop a file of crash reports whose inclusion is based on causal severity level (as opposed to outcome) , so that the incidence of injuries can be compared for cars that meet the standard and those that do not. Stratification of the data by causal severity levels would make it possible to draw inferences about benefit of the standard as a function of severity. Without the severity measure, the levels of exposure of uninjured occupants cannot be determined, and the basis for finding and comparing injury incidence is lacking.

It has been pointed out in an earlier section that there are doubts about the validity of the NHTSA curves of the cumulative distribution functions of barrier equivalent impact speed (BEV or EBS) for injury accidents and fatality accidents. Validating these curves from real-life accident data would require measurement and reporting of the causal severity of fatal and injury accidents.

The measurement and reporting of causal severity in crashes provides a relatively unbiased method of screening crashes for investigation and introduction into a file. Once the severity distribution function for all crashes is established with sufficient

accuracy, reports can be identified by severity level, and only the number of reports needed in each stratum can be selected for admission to the file. Knowledge of the severity distribution functions both for the population and for the file permits analysis of the constrained file and extending inferences to the universe of crashes. At the same time, the size of the file can be reduced by preventing the entry of 'the voluminous reports of low severity crashes whose frequency is high.

B. J. Campbell[10/] feels that a crucial need in the field of crash injury is the means to forge a meaningful link between laboratory test crash data and events as they occur in the field:

> "In the staged crashes in the laboratory, telemetric procedures are used for recording data and one can justify in considerable detail the physical system in which the crash occurs -- the 'g' -forces, the rate of onset, delta 'v' etc. But when one is forced to use nonhuman subjects then one is left in the situation of knowing a great deal about the physics of the crash but knowing little of the actual injuries that might have occured in such a crash. On the other hand, in real world automobile crashes one can learn about the actual outcomes in terms of survival and injuries, but the input variables mentioned before are unknown.

> "The need to link these two systems is apparent. Engineers who design protective systems need to know about stopping distances, forces, decelerations, etc. But knowing these things is of too little help unless one has a way to relate them to real world injuries."

Clearly, a measure of real-world crash severity would help provide such a link.

The question remains as to what constitutes a proper causal severity measure, or "Vehicle Crash Severity Index (VCSI)"[19,22] This question is independent, of course, of what parameters are being or can be measured, such as vehicle deformation, acceleration time history, speed at impact, etc.

The severity measure that has been used in tests, some crash reports from the field, and in motor vehicle safety standards is Barrier Equivalent Impact Velocity (BEV or EBS). It is of interest to examine whether this is a reasonable measure of causal severity, both as regards occupant injury and vehicle damage.

What injures unrestrained and loosely restrained occupants is the so-called "second collision" of the occupant with the interior of the automobile, such as the windshield, dashboard, B-pillar, etc., or with the restraining belts or air bag. The speed with which an occupant impacts an interior element has fair correlation with the injuries he suffers. The speed of impact is determined by the average car acceleration component in the direction from the object to the occupant and the distance between the two:

$$V = \sqrt{2ad}$$

The commonly used head injury criterion is:

$$HIC = \left[\frac{\Delta V}{\Delta t}\right]^{2.5} \frac{\Delta t}{g^{2.5}}$$

Where $\Delta t$ is the time duration and $\Delta V$ is the head speed change during the hardest bump. If the final head speed is zero and there is only one bump, this becomes

$$HIC = V^{2.5} / (\Delta t)^{1.5} g^{2.5}$$

or, in terms of car average acceleration during the crash, is:

$$HIC = 2.38\ a^{1.25}\ d^{1.25} / (\Delta t)^{1.5}\ g^{2.5}$$

Thus, we observe that the criterion for head injury severity increases with car acceleration during the crash interval, but at a slightly greater rate.

If the occupant is tightly restrained, he is subjected to the same acceleration as the occupant compartment of the automobile. The forces he experiences are in proportion to this acceleration and the weight of his own body. It has been determined by investigators[23/] that human tolerance limits can be best expressed in terms of the acceleration to which a person is subjected during the crash interval. It is important to note that rapid variations of acceleration with time are not felt by the unrestrained occupant in crashes in which his motion has a forward component relative to the car, as he is in "free flight" until he impacts the interior. The fully restrained occupant feels these changes (called "jerk") but there is no evidence to indicate that they inflict more than minor punishment; the damage to the restrained occupant appears to result from the average level of acceleration he is subjected to during the crash.

Thus we observe that the two most important measures of injury tolerance can be related directly to vehicle acceleration during the crash. The next question is whether and how barrier impact velocity is related to this acceleration.

Running a car into a barrier causes deformation of the car ("crush") . It has been found in the laboratory that there is a linear relationship observed between impact speed and residual crush. The average acceleration during the crash[3/] is:

$$a = -\frac{V_o k}{2}$$

where $V_o$ is the barrier impact speed and k is a measure of the "stiffness" of the car. Thus we observe that the car acceleration is directly proportional to the barrier impact speed, but also t. the stiffness, which is higher in small cars than it is in full size vehicles.

We conclude, therefore, that barrier impact speed is a reasonable indicator of injury-related causal severity provided that car stiffness is taken into account.

K. L. Campbell[20/] has evolved a sophisticated approach to relating vehicle damage to collision severity. In this approach the dynamic force-deflection characteristics are used to estimate the energy absorbed in plastic deformation of the vehicle. A linear force-deflection characteristic is the simplest (but not necessarily the most accurate) model leading to the observed linear relationship between impact speed and crush distance, and is used by Campbell. The energy can then be expressed as an equivalent barrier speed (EBS or BEV). The approach has been partly validated for frontal impacts in angle and offset barrier tests: The BEV estimates based on vehicle damage differed from the true impact speeds in the angle barrier case, over impact speeds ranging from 18 to 31 mph, by an average of -0.35 mph, with a standard deviation of 2.85 mph; and in the offset barrier case, over a narrow range of impact speeds around 30 mph, by an average of -0.01 mph, with a standard deviation of 1.64 mph. The input

information items required to make the estimate were the crush coefficients as determined from pure frontal barrier tests for each of the various automobiles, together with the actual detailed crush measurements in the test impacts. K. L. Campbell believes that the technique can be extended to side and rear impacts; such an extension would, of course, require determination of side and rear crush coefficients. The crush coefficients, as defined by K. L. Campbell, are the slope and intercept of the curve of impact speed as a function of crush distance. The slope is identical to the reciprocal of the "stiffness" constant we used in the previous paragraphs.

A. B. Volvo employed a series of eleven full-scale frontal barrier, car-to-car and car-to-pole impact tests[24/] to obtain data on crush characteristics of the Volvo model 140 automobile. This information was used in conjunction with detailed measurements of deformation incurred in real-life impacts to estimate barrier equivalent speeds for 128 collisions.

In uncomplicated collisions, we believe that similarity between real-life collision-caused vehicle deformation and that produced in a laboratory staged crash having the same point and direction of impact, implies correspondence between the forces and rates of application. Thus measurements of vehicle deformation can be analyzed, compared with the outcome of staged crashes, and used to estimate barrier equivalent impact speed. However, it is not possible to say that equivalence of deformation always implies equivalent dynamic forces.

Average acceleration during the crash interval appears to be a reasonable measure of causal crash severity. There are several methods by which it can be measured:

(1) By a crash recorder that records acceleration time history (later to be time-averaged over the crash interval to get a severity measure) absent a cheap crash recoder, that directly averages accelerations over the crash interval. The limitation of this approach relates to the large number of recorders required for mass accident files designed to illuminate rare events and the substantial expense associated therefore with this technique. For special measurements such as severity distribution functions, the number of recorders required becomes much smaller, and then this technique of severity measurement becomes appropriate.

(2) By measurement of vehicle deformation (the vehicle is its own crash recorder) and conversion to barrier equivalent speed or average acceleration. The limitation of this approach relates to the limited availability of calibrated deformation information derived from laboratory crashes. Another limitation for mass accident files is the limited ability of police, at the scene of an accident, to judge deformation either using the calibrated crash deformation information, or some other technique, in a consistent reliable manner.

(3) By computer reconstruction of the collision[15/] (SMAC) in an iterative simulation process that is driven to match the reconstructed accident to real-life observations of skid marks, vehicle positions, etc. Momentum changes, in conjunction with known vehicle stiffness characteristics, can be used to estimate crash accelerations. The limitation of this technique is that it requires trained investigators who can estimate the initial conditions of the crash so as to initiate the computer simulation. If the simulation does not converge to the actual disposition of vehicles after the crash, the estimated initial conditions must be revised.

It must be recognized that the crash severity index is a vector, and has magnitude and direction. Two linear accelerometers are necessary to measure its components in the horizontal plane. A third (vertical) component is measured with experimental crash recorders, but does not appear to be very useful.

A problem arises in using vehicle deformation to measure damage-related crash severity; obviously, the cause and the outcome are related. If the outcome is defined as physical deformation, the relationship is one to one. If the outcome is defined as cost to repair, the cause and the outcome are not identical. There is also a flaw in the use of acceleration during the crash interval as a measure of causal severity: if vehicle exteriors were softened, so that average collision accelerations were lowered, average severity would decrease even if the average impact speeds remained the same. So the injury mitigating effects of vehicle softening would be obscured in the collected data. Similarly, where vehicle crush is used to determine severity, if vehicles are designed using resilient materials that do not permanently deform, the average severity would decline despite unchanged average impact speed.

Thus we believe it is important that the National Highway Traffic Safety Administration undertake the job of defining causal crash severity in the most useful and realistic way.

There are several measures of severity currently in use that are quite crude and inaccurate and should be supplanted by better methods.

The deformation extent, a quantity somewhat related to severity, is often reported in Level II (greater depth than the police report) and Level III (in-depth) investigations. The deformation extent is one element of the collision deformation classification (CDC) code assigned in accordance with the Society of Automotive Engineers recommended practice SAE J224a. However, SAE recommended practice J224a warns "The extent number should not be used as a tool for determining severity or energy required to duplicate the damage. For vehicles of the same basic type, it does serve as a tool for gathering together vehicles which have similar damage characteristics. "

Some reports give the full CDC (sometimes known as "VDI") code,* which describes the direction of force, general area of deformation, specific horizontal area, specific vertical area, type of damage distribution, and extent. The Fatal Accident Reporting System reports only impact points and an abbreviated damage extent number.

Pollee reports often include estimates of traveling speed prior to impact, a very poor indication of severity because of the uncertainty of the effects of braking just prior to impact. Sometimes "impact speed" is estimated and reported; again this is a very dubious measure of severity because it is neither uniformly defined nor readily estimated. It may be, depending on the investigator, either speed relative to the ground at the instant of impact of speed relative to the struck or striking object. Ford Motor Company[21/], in an analysis of the differences between investigators' reports of impact speed and the speed

---

* See, for example, reports on crash recorder equipped cars, reference 19.

changes indicated by crash recorders, found differences as great as 40 mph and a standard deviation of 11.9 mph in 20 collisions involving crash recorder equipped cars. The average was a speed overestimate of 14.7 mph by the investigators.

MDAI teams and other in-depth investigators may report their judgment estimates of equivalent barrier speed (EBS) based on their background of understanding of the relationship between EBS and vehicle deformation in laboratory crashes.

To summarize,

(1) Average acceleration during the crash interval is a reasonable measure of the intensity component of a causal crash severity index, but has some deficiencies as such.

(2) NHTSA should, with the approval of the accident research and statistical community, settle on and begin to use an acceptable definition of crash severity index.

(3) If average acceleration during the crash interval is the appropriate measure, there are several ways of measuring or estimating it with reasonable accuracy.

(4) Several indices of severity currently in use are so erroneous, misleading, or ill-defined, as to be valueless, and should be either upgraded or discarded.

## C. THE CRITICAL IMPORTANCE OF AN UNBIASED, RELEVANT, AND ADEQUATE SAMPLING PLAN THAT IS APPROVED BY EXPERTS

In order to meet requirements for collision data collection, it is necessary to generate a plan for sampling and to implement it. The plan should call for collection of a representative sample of crash data in quantity sufficient to be useful at a rate sufficient that the data is timely, and in enough detail and with enough accuracy to permit answering outstanding essential questions.

Thus there are three separable issues:

(1) The methods of assuring that the sample is representative.

(2) The quantities and rates of data gathering.

(3) The information content, detail, and accuracy of reporting.

The problem of securing a representative sample is a difficult and subtle one. To quote Versace (Ford Motor Company)[16/] on the need for scientific sampling:

> "Not only is an increased quantity of data required but the sampling of the accident universe must be by sophisticated protocol. The last of the three reasons given above implies the need for a disciplined approach to the data, to avoid ending up with data which are biased in the factors underlying them. That requires a scientific approach to data collection, not just pouring more dollars into it and

cranking up the administrative machine to get a bigger program going but doing it in the same old way, Data gathering programs must be designed by the same people as will design the analyses that will be applied to the data. No less expertise than the Census Bureau applies, or the Gallup Poll, will suffice. Fortunately,NHTSA has been bringing in very competent people of late, people who know that a data collection scheme must be designed from the start with the method of analysis of the resulting data a key determiner of how the data should be gathered."

The importance of representativeness of the sample is hard to overstate.[8/, 9.] The sample should be representative of the entire population of automobile collisions or have an accurately known relationship to that population. If the sample is selected in some way -- that is to say, if the sample is biased -- inferences drawn from the sample may be faulty. For example, consider a sample in which only injury accidents are represented. If, say, wearing' belts reduces the risk of injury 50%, belted occupants will be underrepresented by 50% in the sample. Two incorrect inferences might be drawn by a naive observer: 1) occupants in accidents don't wear their belts; 2) most of the belted occupants in the sample were injured; obviously belts are not very effective.

Despite the importance of avoiding sample bias, much of the material in the existing national files is heavily biased and, until recently, little thought was given to rectifying this deficiency. NHTSA has contracted with the Highway Safety Research Institute of the University of Michigan to evolve a national crash data sampling plan which, presumably, will be based on sound statistical principles.

The questions to be asked of the data file determine the sampling plan: that is, the selection of regions to be sampled and, within those regions, the collisions on which information is to be collected; the quantity and rate of acquisition of case reports; and the information -- kind and reporting precision -- required in each report.

Examples of such questions are:

(1) How effective have the requirements of MVSS 206 (which specifies crash load requirements on locks, latches, and hinge systems) been in preventing occupant ejections? In preventing occupant injury? Are there significant differences in capability between makes and models of automobiles?

(2) How effective are belt restraint systems (specified by MVSS 208) in preventing injury and death? How does the effectiveness vary with accident severity? Car weight? Occupant age?

(3) At what collision severity level should the bumper system prevent damage to the automobile? Should the requirements be different for front and rear bumpers? For different car sizes and weights?

(4) How important is car visibility in preventing collisions? Are the requirements of MVSS 108 (for lighting) effective in satisfying the needs for nighttime visibility?

(5) What are the factors in passenger compartment design that are of significance in contributing to or preventing occupant injury? To what extent do the characteristics of the occupant himself influence the injury picture? What are the interactions of these factors?

As an example, the last question suggests a number of items of information required for inclusion in reported crash data. According to Lawrence Patrick of Wayne State University[10/], "complete injury data must be included in the accident data. Sex, age, weight, height, and general physical condition are all important factors . . . The type and degree of injury of each occupant including the minor bruises and abrasions and going through the severe bone and soft tissue damage are required. It is important to have complete data on the restraint systems used and the interior components of the vehicle that caused the injury." Also needed, according to Professor Patrick, are impact velocity (as a measure of severity) and direction, location of the impact, seating positions of the occupants, vehicle rigidity, and vehicle interior design.

The design of the sampling plan is critical to the utility of the bank of data that will be acquired through the sampling process. If the reported information is inadequate, crucial questions that one wishes to ask of the file will be unanswerable. If the sample fails to represent the U.S. crash universe, or contains biases, the answers to questions may be quite wrong. And if the quantities of cases on which answers are based are inadequate, the confidence one can assign to the answers is low.

Thus we believe that the National Highway Traffic Safety Administration should proceed urgently with the development of a sampling plan (hopefully, the contract with HSRI will provide the necessary result; if not, it should be augmented).

When completed, but before the plan is implemented, it should be submitted to, reviewed by and approved by a jury of nationally known experts representing the disciplines of accident and injury research, motor vehicle design, rulemaking, and statistical sampling and analysis.

## 4. ALTERNATIVES FOR AN ADEQUATE DATA ACQUISITION PROGRAM

The elements of an adequate data acquisition program have been previously described as comprising a mass data acquisition system with acceptable crash severity capability, a precision crash dynamics measurement system and special investigatory procedures such as multidisciplinary accident investigating teams (MDAI) and fatal accident reports (FAR).

Section 3 has described the quantitative requirements for mass accident data collection. It has been indicated that approximately 500,000 to 1,000,000 accident reports per year are needed to obtain early warning of motor vehicle hazards and to obtain confirmation of the effectiveness of various safety programs in a timely way to a reasonable level of significance. The exact number of annual accident reports needed depends on the level of detail of the desired results, the frequency of the event being investigated, the desired accuracy and confidence level of the information being obtained and the time by which the information is desired.

For example, if one wishes to determine the fatality rate in rollovers of belted drivers in one year to an accuracy so that the standard deviation is 30% of the mean, 130,000 accident reports would be needed. However, if one wished to determine the probability distribution function of car weight in cases where belted drivers are killed in rollovers to an accuracy of 5% with a confidence of 95%, 3,500,000 accident reports would be needed.

The kind of data needed for this mass acquisition system is generally agreed to be a causal severity index, vehicle identification number, road and visibility data, injury scale, restraint

system and usage, driver and occupant descriptions and seating positions, with many other items required, perhaps on a special survey basis, to answer specific questions.

There are a number of ways to obtain a causal severity index. If a cheap ($2) two axis crash recorder can be developed --and there are some concepts worthy of exploration--their installation on production cars is justified. This possibility is more fully discussed later in this section.

In the absence of a cheap crash recorder, vehicle deformation should be used as a causal severity index. There are at least two major approaches, one following the lead of Professor B.J. Campbell at the University of North Carolina, and the other following the approach of Professor Lawrence Patrick at Wayne State University, the Biomechanics Research Center and practiced in a recent Volvo-Wayne State University study.[24]/

The State of North Carolina uses police reports of severity reported by the TAD system.* Police training has evidently been sufficiently good to obtain useful reports[7]/ although the data base has been small and the severity reporting system quite simple. The disadvantage of this approach is summarized by Griffin:[7]/

---

* A police officer using the TAD system rates severity on a 1 to 7 scale by matching the damaged vehicle with a manual of photographs of typical accidents.

"Rural accidents tend to be more severe than urban accidents, therefore, police level data for a given state must be generalized with caution, even within that state.

"It is not simple to generalize police level data from one state to other states. States differ with respect to traffic density, number of interstate highways, and weather conditions. All of these factors interact with accident types and configurations, and thereby affect the benefits to be derived from a safety device.

"Finally, police level data are not recorded in detail. Levels of vehicle damage and occupant injury are evaluated by an officer who may be trying simultaneously to summon medical aid, direct traffic, and determine whether or not a law has been broken. Under these circumstances, the data yielded by these investigators is very good, but necessarily the collection of data should not be considered the officer's area of expertise or his major area of responsibility."

Professor Campbell[10/] feels the cost of improved police reporting could be nominal and that it would be important to extend the North Carolina system, or some improvement of it, to a number of states that might together provide 600,000 - 1,000,000 reports which would be less biased than those from rural North Carolina alone.

It is difficult to accurately determine the cost of this system, but $3-10 per report is approximately correct, or a total of $10 million for one million reports. However, there is some question of the adequacy of police data for many needs.

professor Patrick's approach to the recent Volvo experiment[24/] might be utilized to improve the reporting of causal severity by police. Staged crashes of major U.S. models, front, side and rear into poles, barriers and cars at three speeds could be used to obtain calibrated deformation data. The one-time cost of such a program is estimated* to be $3-5 million. There are a number of possible ways to use these data. Police could be trained to photograph** the damaged vehicle from a few aspects after having placed appropriate identification placards and scales on the damaged vehicle. The film could be subsequently processed at various centers to derive the severity data by analysis Of the photographs and by comparison with the calibrated deformation data. The total accident report including police and medical data, if any, could be assembled at the photographic analysis center.

Alternatively, it might be possible to train police equipped with appropriate templates to measure the collision deformation in conformance with a handbook based on the calibrated deformation data from the staged crashes. Appropriate supplies, compensation and incentive would have to be provided to local police. A cost of $10-20 per accident report might be sustained by more detailed analysis of this reporting system. Therefore this type of mass accident data system might cost a total of $25-30 million for the first year including non-recurring capital as well as operating costs.

---

* Conversations with Professor Patrick.

** Mr. John Garrett of Calspan reports some success in Western N. Y. comparing estimates of severity from police photographs with estimates of professional accident investigation teams.

In Section 2 there was also described the need for some precision reference data. This need was stressed by almost every participant in the Workshop.[1/, 10, 11, 13, 14/]. In particular sume 10,000 sophisticated recorders with an accuracy of 1-2 mph*, are needed to obtain in one year's time a representation of the probability distribution of severity of accidents (above the police reporting threshold) with severity (barrier impact speed), to an accuracy of 5% and at a confidence level of 95%. If this representation of the distribution of severity were limited to frontals only, the confidence level would be only 80% with an accuracy of 5%. Alternatively, 20,000 recorders could be used to obtain this distribution for frontal collisions to an accuracy of 5% at a confidence level of 95%. The cost of sophisticated crash recorders in these quantities is approximately $200. Therefore the total cost of this basic program is between $2 and $4 million plus the cost of data retrieval and analysis.

The cost per accident report from the sophisticated crash recorder** would be approximately $2,000 the first year, declining to $1,000 over the first two years, $500 over the first four years, $200 over the first ten years. This is the normal characteristic of the flow of benefits over a period of time from an initial capital expense.

---

* This corresponds to a 3.8 - 7.6% change in the cumulative distribution of fatalities or an annual dollar cost equivalence of approximately $250-500 million in estimating the effectiveness of occupant restraint systems.

** Described later in this section.

The SMAC system of computer-aided accident reconstruction could also be used to obtain precision reference data, and is competitive with the sophisticated crash recorder. It is our opinion that the SMAC system, while extremely clever and promising, has not completed its development cycle, and must be operated by full time professionals. These might be specially trained police. However, some means would have to be found to compensate state and city police for performing NHTSA work. If a SMAC van is to operate around the clock, a crew of eight per vehicle would be required. If as many as 100,000 accidents were to be investigated per year with 500 vans, a total crew of approximately 4000 men would be required at an annual cost of $60 million. Thus, the manpower cost seems to limit the SMAC system to obtaining relatively small numbers of reports, say 10,000 per year or lower. The SMAC system like the sophisticated crash recorder, seems most useful for special data gathering programs requiring precision severity data. If 2500-5000 accidents are to be investigated per year, perhaps 15-20 vans would be required at a total manpower cost of $1.8 - 2.4 million plus the cost of equipped vans and processing centers, or roughly $500 per case.

These costs should be compared to the current costs of MDAI investigations at $2000 per case on scene and $800 per case off scene , FAR reports at $15 per case, Level II reports at $100 per report.

Some safety devices, particularly those with uncertain performance and high cost to the consumer, could be subjected to a field test prior to general introduction. Some Federal agencies, The Food and Drug Administration, for example, do require extensive tests of products before general use. These tests, if properly designed and monitored, could yield invaluable data on the benefits from such devices.

However, a safety feature like the 5mph bumper or passive restraints can probably not be sold on a trial basis depending on market forces alone. Therefore, Federal sponsorship would be necessary to design the field trial, pay the cost of installation and monitor the results. This process would be expensive but, when viewed against huge consumer costs, may be worthwhile.

Such a test has been suggested for passive restraint systems by the National Motor Vehicle Safety Advisory Council, a body advisory to the Secretary of Transportation, by a Resolution adopted by an 11 - 5 vote on November 19, 1974.*

It is the feeling of a number of both the academic and automotive participants in the Workshop, and the authors of this report, that a field trial of 100,000 - 200,000 passive restraint systems is necessary.

The size of the field trial of passive restraints arises from the following considerations. If one assumes that the passive restraint is effective in reducing fatalities by 50%, then it would require three years of field trial of 200,000 equipped cars to determine the probability density of severity given a fatality to an accuracy of 10% with 80% confidence. On the other hand, if one wished to determine whether the fatality rate in all passive restraint equipped cars had decreased by 50% to an accuracy of 20%, 125,000 installations would be required to obtain an answer in one year. If on the other hand, one wished to determine the performance to the same accuracy in light cars as compared to heavy cars, one would have to wait two years, assuming the 125,000 car sample was split equally between heavy and light cars.

---

* See Appendix L.

For this field trial to be unbiased, these systems would have to be installed in small and large vehicles in representative parts of the country with a representative set of drivers. Since market forces cannot be depended upon to provide this, it is probably in order for the Federal mandator of the proposed regulation to support the trial. The cost of such a program could be $30 - $60 million.

In summary, an extensive mass accident data system of one million reports annually may cost

(1) $3-10 million annually using the North Carolina approach of upgrading police reporting, plus the cost of improvements in severity estimation;

(2) $10-20 million annually using the Wayne State - Volvo approach to obtaining accident severity, PIUS the costs of reporting factors other than severity, plus a one-time cost of $5 million for calibrated vehicle crash data and other capital expenditures:

(3) $10 million annually to obtain severity information alone if a cheap ($2) crash recorder could be developed and installed on 50% of all new production. One would have to add. to this cost the cost of collecting the records, analyzing the data and coalescing this information with other accident information in a mass data file.

These several approaches are potentially mutually supportive rather than competitive. There is presently no such thing as a cheap recorder, so one cannot depend on it for severity data. Should one be developed, it would be extremely useful for mass accident data. A serious effort toward this objective should be undertaken. If the Wayne State - Volvo approach to obtaining accident severity could be developed to apply to the U. S. problem, then it might be used in conjunction with the North Carolina approach as a better method of estimating severity.

A needed tool for precision research on the crash dynamics of a few thousand accidents annually may be obtained by either SMAC simulation or precision crash recorders.

(1) $2-4 million first cost for 10,000 to 20,000 sophisticated crash recorders plus the cost of the facilities and personnel needed to analyze and correlate the data produced as an annual expense.

(2) $2-2.5 million annually for personnel on vans plus the vans themselves and analytical equipment.

It would seem possible to put emphasis on one or another of these programs. In doing this NHTSA should take into account the somewhat higher first costs of the crash recorder program as compared to the somewhat higher annual operating costs of the SMAC program. Obviously this cost analysis must be viewed against the differences in the kind of data obtained from the two approaches. The SMAC vans do get trained investigators to the scene. NHTSA can best evaluate if this capability is justified in view of the multidisciplinary accident investigating teams. Since MDAI teams report on 1500-2000 cases per year from a perspective that is broader than crash dynamics, it seems advisable to maintain this capability.

The field trial of 100,000 - 200,000 passive restraint equipped cars in a representative sample would cost 30 - 60 million dollars first cost plus annual analysis expense.

Thus in addition to the current accident program of approximately $5 million covering such activities as MDAI, FAR, Level II reports, NHTSA and the Congress should consider adding a mass accident data system that might cost $5 - 20 million annually, a precision crash dynamics system (probably sophisticated crash recorders) at a first cost of $2 - 4 million, and finally a field evaluation of passive restraints costing $30 - 60 million. Table 4 summarizes the existing programs and the recommended alternatives for the additional data that we deem to be required.

The genesis of this OTA study was an issue concerning sophisticated crash recorders and their proper use in accident data retrieval.

TWO types of crash recorders have been developed under NHTSA sponsorship.

one of these, commonly known as the "tape recorder," was developed by AVCO Systems Division, Wilmington, Massachusetts. It is designed to measure and record vehicle parameters before, during and after a crash. The time history of the following quantities is recorded prior to the crash:

TABLE 4

EXISTING AND PROPOSED PROGRAMS

| DATA NEEDS / ALTERNATIVES AND COSTS | MASS ACCIDENT FILE (500,000-1,000,000 REPORTS ANNUALLY) | PRECISION CRASH DYNAMICS (2,500-5,000 REPORTS ANNUALLY) | SPECIAL STUDIES |
|---|---|---|---|
| Medical and Police Reports Using TAD | $3- $10 Per Report, North Carolina Prototype | | |
| Medical and Police Reports Using VDI or CDC | Upgraded Severity Capability as Compared to system Above | | |
| Medical and Police Reports Taking Photos to be Compared to Calibrated Crashes | $10-$20 Per Report Wayne State - Volvo Prototype, Probably the Best Severity been Demonstrated for this file | | |
| Medical and Police Report Using Cheap Crash Recorders if Available | $10 Per Report for Severity Plus $3-$10Per Report for all other information | | |
| Computer Simulation (SMAC) (15-20 Vans) | | $2 - $2.5 Million Annual Personnel Charge Plus $1.5-$2 Million First cost | |

Table 4 - continued

| DATA NEEDS / ALTERNATIVES AND COSTS | MASS ACCIDENT FILE | PRECISION CRASH DYNAMICS | SPECIAL STUDIES |
|---|---|---|---|
| Sophisticated Crash Records (10,000-20,000) | | $2- $4 Million First costs Plus Annual Analysis & Maintenance costs of $0.5 - $1 Million | |
| Multidisciplinary Accident Investigation Teams (MDAI) | | | 1500 Reports/year At $2000 Per Report on Scene, $800 Per Report Off Scene |
| Fatal Accident Reporting System | | | 55,000 Death Reports Per Year Contemplated At a Cost of $1 Million, Uncertain Severity Indications |
| Level II Restraint System Investigation | | | Analysis of Restraint System Effectiveness From Police and Medical Reports, $100 Per Case |
| Field Trial of Uncertain and/or Expensive Safety Aids | | | 100,000 - 200,000 Car Field Trial of Passive Restraints $30-$60 Million One Time Cost |

Brake pressure (200-2000 psi, accuracy ± 7%

Steering wheel motion (1260°, accuracy ± 3%)

Speed (as derived from the speedometer cable)
(o - 120 mph, speedometer accuracy)

Longitudinal and lateral vehicle acceleration
± 1 g, accuracy ± 5%

During the crash is recorded the time history of:

Longitudinal acceleration (± 50 g, accuracy ± 3%)

Lateral acceleration (± 50 g, accuracy ± 3%)

Vertical acceleration (in vehicle coordinates)
(± 50 g, accuracy ± 3%)

Prior to the crash, the recorded data are sampled at a 20 per second rate. During the crash, the recorded data are sampled at a 200 per second rate. The duration of the tape record is from 6 minutes prior to the crash to 10 seconds after the crash. A garden variety endless-loop 8-track cartridge is used as the storage element.

Recording is done in digital (PCM) format. The total system includes each of the several sensors, a crush sensor and a recorder, packaged separately.

The other recorder, commonly known as the "disk recorder," was developed by Teledyne Geotek, Garland, Texas. It is a single unit that records, only during the crash interval, the time history of lateral, longitudinal accelerations. The range of accelerations measured is ± 50 g, with an accuracy of ± 8%.

The disk recorder is much simpler and less expensive than the tape recorder, and has been purchased and installed in experimental quantities by NHTSA. 1050 have been installed in fleets throughout the country, including air bag equipped cars.

The tape recorder is intended to provide data that could give useful information on the handling, braking, speed and forces experienced by the vehicle prior to the crash. Both recorders provide a crash-acceleration time history, which yields information on the forces to which the vehicle was subjected during the crash, and which, if properly interpreted, can give magnitude and direction of crash severity.

In Fiscal Year 1975 testimony, a total cost estimate of $10 million for a crash recorder program was presented. This program would have procured 100,000 disk recorders as compared to the previous 85,000 disk recorders (at $75 per unit) and 15,000 of the more expensive tape recorders for a total cost of $15 million. The program costs include support for initial purchase and funds allocated for analysis of the data provided by the recorders.

The Transportation Systems Center of the Department of Transportation (Mr. Louis Roberts) has examined the feasibility of a somewhat cheaper, all solid state, more accurate alternative to the Teledyne Geotek disk recorder, and have concluded that such a unit could be built at a unit cost of $125 in quantities of 100,000. With this recorder, three-axis accelerations would be measured to 1%.

C. Y. Warner and Joseph Free of Brigham Young University, and Brian Wilcox and Donald Friedman of Minicars, Inc.* have proposed as a severity measuring device a very simple two-axis integrating accelerometer whose outputs are change in velocity during the crash interval. The Breed Corporation is also developing two cheap crash recorders. One will provide information indicating that the crash resulted in a velocity change of more than 30 mph. This is accomplished by a latching system. The other system provides a direct reading of crash severity. A combination of Coulomb and viscous forces acting on a mass provide a system that is insensitive below a threshold, responds to the vehicle change in velocity during the crash, and latches after the crash indicating the change in velocity experienced.

We believe that development of a cheap and simple severity measuring and recording device is highly desirable. There appear to be many feasible design alternatives to the Warner device, and they should be examined. A recorder that is designed to measure average acceleration during the crash interval, as opposed to velocity change alone, should be considered. Lynn Bradford, NHTSA crash recorder program manager, concurs that only the two horizontal components of acceleration need be sensed, and that the third axis can be omitted.

## 5. FEDERAL RESPONSIBILITY AND EXPENDITURES FOR COLLISION DATA GATHERING

The Federal Government through the Department of Transportation, has undertaken the responsibility for setting safety and damage-limiting standards for motor vehicles. The costs of standards put into effect thus far is more than $2.5 billion annually. It would appear that prudent and responsible rulemaking would imply that each such standard should be promulgated only after acquiring through data collection and large scale experiment a thorough understanding of the frequency of occurrence of the hazards to which the standard was addressed, the extent to which a design to the standard would mitigate the outcome in terms of damage or injury, and the consequent benefits as related to the estimated costs. But because of the dearth of data, rulemaking has been based instead on guesswork and judgment. Fortunately, two standards (energy absorbing steering column and belt restraints) appear on the basis of limited evidence to be highly successful. TWO others, HPR glass and head restraints, appear to be beneficial; but the others remain to be evaluated, and in the meantime, their costs continue to be borne by the public.

Motor vehicle collision loss is an enormous national problem that requires centrally coordinated solutions, both in terms of motor vehicle standards and highway designs. Implicit are both the need and the responsibility for centrally supported collection of collision data, representative of all the States, from which may be drawn inferences regarding the need for and benefit of vehicle and highway design changes. The establishing of a central collision data file further implies a need and responsibility for standardization of reporting systems and formats so that input data from

many sources can be combined. The federal government should undertake these responsibilities as the central and coordinating activity for collection of crash data.

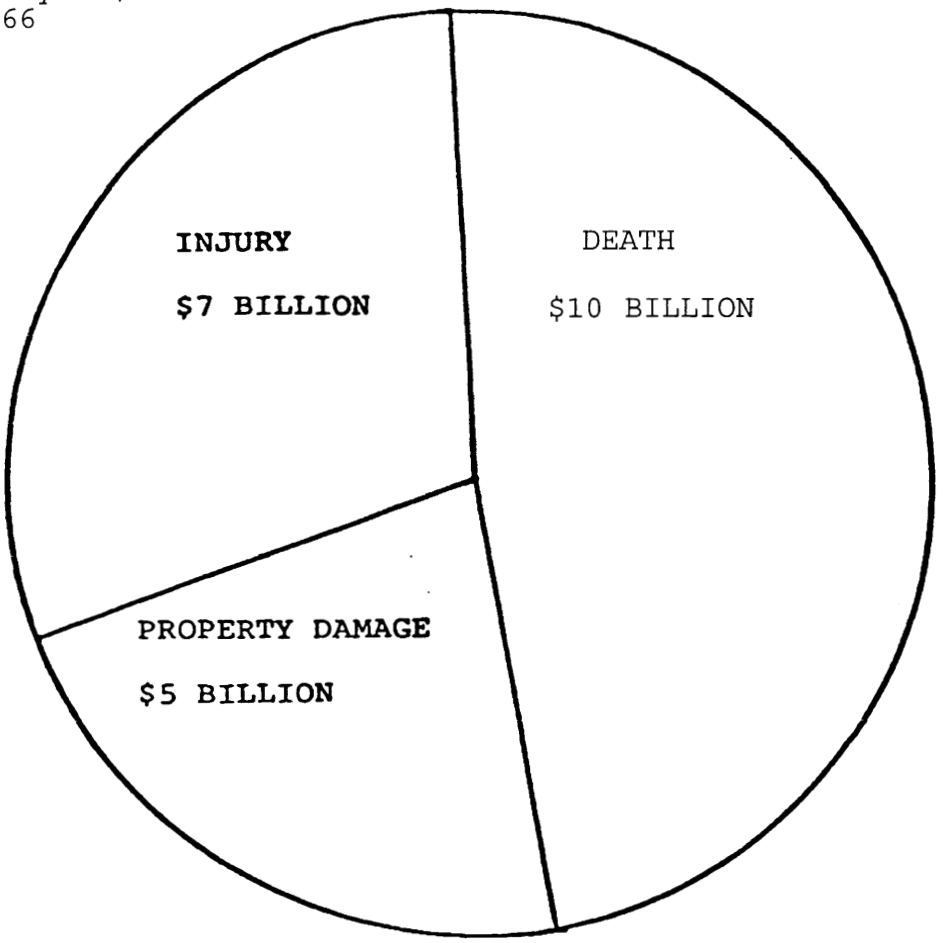
In addition to the question of responsibility, there is the question of capability. On this question, John Versace of Ford Motor Company[16] has the following comment.

> "Mass accident data acquisition, processing, analysis, and broad scale distribution requires great effort and much resource. Only the federal government has the necessary resource and easy access to the agencies which can supply information. Furthermore, it seems that it is the responsibility of the federal government to assemble data which will allow an accurate public review of the real dimensions of the crash and injury problem on our highways."

The current level of Federal expenditure for the collection and analysis of automobile collision data is $5-6 million yearly. A few examples will be presented to illustrate that the justifiable levels of expenditure may be much higher than the current amounts.

1. Each traffic fatality is a catastrophe that costs society approximately $200,000.[2] current Federal expenditures 'or collision data gathering average less than 0.06% of the cost of traffic deaths.

2. 28 million automobile accidents cost the United States $22 billion annually. Federal expenditures to collect data average less than 22¢ per accident-involved automobile, and less than 0.03 % of total losses (see Figure 1).




ANNUAL COST OF MOTOR VEHICLE ACCIDENTS

•

$6 Million

NHTSA EXPENDITURES ON CRASH DATA COLLECTION

FIGURE 1. Comparison of the cost of motor vehicle accidents with Federal expenditures to acquire and analyze crash data.

3. The cost of 5 mph no-damage bumpers front and rear has been estimated as $119 per car (first cost) plus about $100 in added lifetime fuel costs. The total consumer expenditure required to equip all cars is about $2.2 billion per year. Because of the paucity of hard statistics or the frequency distribution and cost of low-severity accidents whose damage the bumpers tend to mitigate, there is an uncertainty of at least 10% or about $200 million, in the estimate of the benefits; this uncertainty alone is more than 30 times the current Federal data collection expenditures.

4. Continuing uncertainties about the effectiveness of seat belts lead to differences in estimates of numbers of lives saved (at 50% belt usage) of at least 8000 annually representing a societal gain or loss of $1.6 billion. This uncertainty is more than 250 times the current Federal expenditures on data collection and analysis.

Thus high levels of expenditure appear justified by the magnitude of the motor vehicle collision loss program and its uncertainties. They are not necessarily required to do the job. The actual amounts needed must be determined after the development of a comprehensive plan that specifies in detail the information needed, the quantities of data and rates at which it is to be gathered, and how the plan is to be implemented.

The benefits of a data collection and analysis effort can be easily seen when it is used to resolve a choice between two approaches to solving a problem. The benefits are less obvious, just as in any research effort, when the outcome is unpredictable in terms of establishing the measures and costs of reducing damage, injury and death.

## 6. LEGAL ASPECTS OF CRASH RECORDERS

Questions that are often brought up with regard to automobile crash recorders are (1) whether crash recorder evidence is admin-sible in a court of law; (2) should it be admitted?; (3) can it be prevented from being admitted?

There is a useful parallel in the inflight recorders installed in commercial airplanes. In the event of a crash, the data in these recorders is read out and interpreted by the Federal Aviation Administration or National Transportation Safety Board staff personnel. Section 701 (e) of the Federal Aviation Act forbids the use of the NTSB report in any suit or action for damages arising out of an accident. The original policy considerations were that if such possibly legally damaging reports could be used in court, it would inhibit possible sources of information important to the cause of NTSB in promoting safety. But it is possible to get the FAA or NTSB staff member who read out the recorder to testify as to the facts and thus the "facts", data read or heard from the recorders can be received as evidence toward the proof or defense of an allegation of negligence. Neither the airlines nor the government has any privilege to exclude or restrict such evidence.*

Similarly one could expect that automobile crash recorder data could be admitted in evidence in a court of law; but there would be the usual problem of qualifying the evidence. In the absence of a stipulation of the opposing party as to the authenticity of the data and the reliability and accuracy of the recorder, the moving party would successfully have to demonstrate to the court the reliability and accuracy of the recorder and the expertise of the person who read out the data.

---

* From a private legal opinion.

On the question of whether crash recorder data should be admitted, the main point again is whether the recorder is reliable, accurate, properly read out, and provides a record of the particular event in question. The data of itself is not dispositive of liability, but merely serves as certain evidence of the event. As indicated earlier in this report, there is good correlation between the crash severity a recorder might measure and the extent of crash deformation to the vehicle in which it is installed; and it would be difficult to refuse evidence on the crash severity magnitude as interpreted from vehicle deformation. Thus if the recorder provides good evidence of the event, it seems appropriate that that evidence should be admitted.

It may be possible to restrict through legislation the admissibility of crash recorder evidence, particularly if the recorders are government-owned and the records are retrieved and interpreted by government employees. Consider, however, the objective of a very simple and widely used integrating accelerometer that is conveniently and reasily read by any police accident investigator without special training. It would appear difficult to prevent testimony by a layman -- say a tow-truck operator or an auto mechanic -- as to what he saw immediately after the accident.

In summary, we believe that (1) the data from a crash recorder would be admissible, if it meets necessary qualifications, in a court of law; (2) the data should be admitted if it is good evidence; (3) it will be difficult to prevent admitting crash recorder data, even by Federal law, if the record can be easily read by an untrained person.

## REFERENCES

1. More Sophisticated Data Collection for an Improved Accident Data System, Edwin A. Kidd, Calspan Corporation, January 27, 1975; included as Appendix A of this report.

2. Review and Critique of National Highway Traffic Safety Administration's Revised Restraint System Cost Benefit Analysis, Economics and Science Planninq, January 22, 1975.

3. Cumulative Regulatory Effects on the Cost of Automotive Transportation (RECAT) , Office of Science and Technology (White House), February 28, 1972.

4. Letter from Richard Wilson, General Motors Corporation February 4, 1975. Included as Appendix E to this report

5. Passive Protection at 50 Miles Per Hour, U. S. Department of Transportation, National Highway Traffic Safety Administration, June 1972.

6. Analysis of Effects of Proposed Changes to Passengers Car Requirements of MVSS 208, Motor Vehicle Programs. National Highway Traffic Safety Administration, Department of Transportation, August 1974.

7. Analysis of the Benefits Derived from Certain Presently Existing Motor Vehicle Safety Devices: A Review of the Literature, Lindsay I. Griffin III, University of North Carolina, December 1973.

8. National Accident Data Systems, C. Thomas Terry, General Motors Environmental Staff, Proceedings, Automotive Safety Seminar, June 20-21, 1973.

9. Statistical Inference from Multidisciplinary Accident Investigation, James O'Day, Highway Safety Research Institute, University of Michigan, August 1974, (Department of Transportation Report--DOT-801 111) .

10. Statement by B. J. Campbell, Highway Safety Research Center, University of North Carolina, Automobile Collision Data Workshop, January 17, 1975. Included as Appendix C to this report.

11. S ummary of Remarks at the Automobile Collision Data Workshop, Lawrence Patrick, Wayne State University. January 20, 1975; included as Appendix D to this report.

12. Statistical Rationale for the Number of Automobile Crash Recorders Purposed for Procurement and Installation by NHTSA; National Highway Traffic Safety Administration (Received February 5, 1975) ; included as Appendix F to this report.

13. Crash Recorders and Alternate Methods of Defining Crash Severity, James O'Day, Highway Safety Research Institute, University of Michigan (received February 8, 1975) ; included as Appendix G to this report.

14. Adequacy and Limitations of Current Data Systems, Marie D. Eldridge, National Highway Traffic Safety Administration, January 16, 1975; included as Appendix H to this report.

15. A Discussion of Data Gathering Systems, Edwin A. Kidd, Calspan Corporation, January 16-17, 1975; included as Appendix I to this report.

16. Mass Accident Data Acquisition and Why It's Needed, John Versace, Ford Motor CO., January 16, 1975; included as Appendix J to this report.

17. Reports on Traffic Accident Research, Volvo, March 1973, A.B Volvo, Car Division, Gothenburg, Sweden.

18. Position Statement on an Expanded, Low-Cost National Accident Data Collection Program. J. R. Cromack, B. J. Campbell, L. Patrick and B. O"Neill, February 7, 1975; included as Appendix K to this report.

19. Automotive Recorder Research - A Summary of Accident Data and Test Results. S. S. Teel, S. J. Peirce and N. W. Lutkefedder, Proceedings of SAE 3rd International Conference on Occupant Protection, July 10-12, 1974.

20. Energy Basis for Collision Severity, Kenneth L. Campbell, General Motors Corp. , Proceedings of SAE 3rd International Conference on Occupant Protection, July 10-12, 1974.

21. Ford Motor Co. submission to NHTSA Docket 74-15, Higher Speed Protection Requirements, September 19, 1974.

22. Automotive Recorder Research and Its Effects on Future Vehicle Safety, S. S. Teel and N. W. Lutkefedder, National Highway Traffic Safety Administration (undated) .

23. The Mechanics of Automobile Collisions, Felix Rosenthal et al, Naval Research Laboratory Memorandum Report 2417, May 1972.

24. Comparison of Three point Harness Accident and Laboratory Data, L. Patrick, N. Bohlin and A. Anderson; Wayne State University, Aug. 20, 1974.

25. Resolution of the National Motor Vehicle Safety Advisory Council, November 14, 1974.

# OFFICE OF TECHNOLOGY ASSESSMENT

## Automobile Collision Data

AN ASSESSMENT OF NEEDS AND METHODS OF ACQUISITION

FEBRUARY 17, 1975

PREPARED **AT THE REQUEST OF**

THE HOUSE COMMITTEE ON APPROPRIATIONS

VOLUME II

**PREPARED UNDER CONTRACT** OTA CII

ECONOMICS & SCIENCE PLANNING, INC.
120018 TH STREET N.W.
WASHINGTON, D. C. 20036

## APPENDICES

Page

APPENDIX A ..... 1

More sophisticated Data Collections for an Improved Accident Data System. Edwin A. Kidd, Calspan Corporation, January 27, 1975.

APPENDIX B ..... 7

Letter from John Versace, Ford Motor Co., February 6, 1975.

APPENDIX C .*****9 .OmmO*** .*mm*o** .*9**,.* .*.m.... . 31

Statement at the Automobile Collision Data Workshop, B. J. Campbell, Highway Safety Research Center, University of North Carolina, January 17, 1975.

APPENDIX D ..... 37

Summary of Remarks at the Automobile Collision Data Workshop, Lawrence Patrick, Wayne State University, January 16, 1975.

APPENDIX E ..... 53

Letter from Richard Wilson, General Motors Corp., February 4, 1975.

APPENDIX F ..... 61

Statistical Rationale for the Number of Automobile Crash Recorders Proposed for procurement and Installation by NHTSA; National Highway Traffic Safety Administration (received February 5, 1975).

APPENDIX G ..... 77

Crash Recorders and Alternate Methods of Defining Crash Severity, James O'Day, Highway Safety Research Institute, University of Michigan (received February 8, 1975).

Page

APPENDIX H ..... 87

Adequacy and Limitations of Current Data Systems, Marie D. Eldridge, National Highway Traffic Safety Administration, January 16, 1975.

APPENDIX I ..... 101

A Discussion of Data Gathering Systems, Edwin A. Kidd, Calspan Corporation, January 16-17, 1975.

APPENDIX J ..... 119

Mass Accident Data Acquisition and Why It's Needed. John Versace, Ford Motor Company, January 16, 1975.

APPENDIX K ..... 133

Position Statement on an Expanded, Low-Cost National Accident Data Collection Program, J. R. Cromack, B. J. Campbell, L. Patrick and B. O'Neill, February 7, 1975.

APPENDIX L .*.**** ..*.**** . .** ** **0= .*0*. *OG*O .. ....139

Resolution of The National Motor Vehicle Safety Advisory Council, November 14, 1974.

APPENDIX M ..... 141

The Need for Standardization in Reporting Collision Damage and Injury in Traffic Accidents, J. Robert Cromack, January 16-17, 1975.

APPENDIX N ..... ...147

Regulatory Rulemaking Based on Less Than Total Information, David Morganstein, Center for Auto Safety, and L. A. Goldmuntz, Economics & Science Planning, Inc., Received February 21, 1975.

APPENDIX O ..... 151

Automobile Collision Data Workshop: Agenda; Scheduled Presentations; Salient Residual Issues. January 16-17, 1975.

APPENDIx P ..... .....159

Letter from John Garrett, Calspan Corp., March 12, 1975, and Attachment.

APPENDIX A

# MORE SOPHISTICATED DATA COLLECTIONS FOR AN IMPROVED ACCIDENT DATA SYSTEM

Edwin A. Kidd
CALSPAN CORPORATION

January 27, 1975

# MORE SOPHISTICATED DATA COLLECTION FOR AN IMPROVED ACCIDENT DATA SYSTEM

**EDWIN A. KIDD**
CALSPAN CORPORATION

January 27, 1975

## NEED

It is essential that NHTSA have a national data bank for surveillance and effectiveness studies related to the impact of standards on accident, injury and fatality frequencies. The relatively small output of the special federal teams and/or the higher quantity, but low content state data banks are inadequate for the purpose. In addition to information on the general accident environment, vehicle damage and occupant injuries, details of the impact environmmt -- velocity at impact, change in velocity during impact and possibly, vehicle deceleration -- are required for a sample Of 100,000 to 500,000 automobiles annually.

CANDIDATE SYSTEMS

Candidate systems for achieving the required information are:

- Crash recorders, with accident, vehicle and occupant information supplied by "conventional" investigations by police and/or special teams.

- Off-scene computer reconstruction of accidents as reported by police and/or special teams.

- Computer aided investigation and reconstruction of accidents (e.g., SMAC) using appropriately equipped police and/or special teams.

Use of crash recorders alone to provide data on the impact environment for the required number of accident cases would be prohibitive in cost. For example, if accidents of tow-away severity or higher are of principal interest, then 30-40 times as many automobiles must be equipped with crash recorders as the number of accidents needed annually for analysis*. Also, and most important, the crash recorder only provides a portion of the data required; conventional investigative methods must still supply accident and vehicle descriptions, vehicle deformation, occupant injuries, restraint system use, etc.

---

* Additional cost for each accident case would be 30-40 times the cost of each crash recorder installation.

Off-scene computer reconstruction (SMAC) of the more extensively reported accidents now resulting from the special team studies has already been demonstrated. A modest amount of additional work on SMAC is required to increase the generality of the reconstruction of the various accident types. Also, application of this reconstruction aid to accidents as presently reported by police should be studied with the objective of determining the minimum information required for each accident.

The computer aided investigation and reconstruction of accidents by police offers the most promise for the attainment of the large data base required. If the use of appropriately equipped accident investigation vans is determined by individual police agencies to be beneficial for their present activities, in terms of overall efficiency, then the mechanism for providing all of the necessary accident data for NHTSA will be accomplished. Providing the police with equipment that will be cost/beneficial for their present needs will obviously provide the means for the attainment of the data required by NHTSA at the lowest cost. Also, the digital format of the accident descriptions and reconstructions that would be output from this equipment would result in minimal data processing for a fast response data bank.

RECOMMENDED APPROACH

The overall objective is the attainment of a consistent, coherent data bank that will be adequate both in size and specific content for the purpose and is practical in terms of development time and cost.

Whether police or special teams are used as the basic collection agency need not be decided immediately. What should be decided as soon as possible is the efficacy of the computer aids to reconstruction and

investigation that have already been developed. The crash recorder may play a role in further validating these aids; it cannot be seriously considered as the ultimate data collection method because of prohibitive cost. A decision to continue with special teams should include the provision of demonstrated aids for these teams. If police want these aids and their efforts can be integrated into a national data system, then it appears axiomatic that a police based system would provide the most for the least cost.

A program is outlined below for achieving the improved accident data system:

1.) Install crash recorders in special automobiles, e.g., air bag equipped vehicles, for additional validation of computer reconstruction aids.

2.) Accident Reconstruction

a.) Continue validation of SMAC via staged crashes (including crash recorders) for a broad accident spectrum.

b.) Determine accuracy achievable on police reported accidents and establish minimum data requirements as function of accuracy achievable.

3.) On-Scene Reconstructions

a.) Accident vans for special investigative teams to improve data consistency and achieve more accurate reconstructions.

b.) Field trials with police agencies to determine accident reporting and/or accident reconstruction configurations.

4.) As the result of 1, 2, and 3 (above), establish national data collection system elements.

TIMING AND APPROXIMATE COST

Aggressive pursuit of this plan would provide detailed requirements and set up of the overall system within two years. Meanwhile, the present multilevel data collection centers would continue to provide data, but with a transition toward the final system in the second year.

An acceptable, complete data system could be achieved at a cost of five to ten million dollars annually. Actual costs to NHTSA are dependent upon the usefulness and acceptability of the investigation and reconstruction aids to selected police agencies and their subsequent integration into the system.

APPENDIX B

LETTER FROM JOHN VERSACE
FORD MOTOR COMPANY

February 6, 1975

Ford Motor Company

20000 Rotunda Drive
Dearborn, Michigan 48121
Mailing Address:
P.O. Box 2053
Dearborn, Michigan 48121

February 6, 1975

Mr. Howard P. Gates, Jr.
Economics & Science Planning
1200 18th Street, N.W.
Washington, D.C. 20036

Subject: OTA Automobile Collision Data Workshop

Dear Howard:

It did take some time in a very busy schedule to meet with you and to put our thoughts down, but we appreciate the opportunity to express our understanding of, and our position on the subject of accident data. In regard to societal costs: the Ford Motor Company submission to Docket 74-15 -- Advance Notice Concerning Higher Speed Protection Requirements -- contains some estimates of the additional consumption of resources entailed in trying to meet a high speed requirement.

It is difficult to determine all the ways in which inadequate accident data would lead to unnecessary expansion of costs, but we believe this one example will provide a general picture of the possible magnitude of such expense. I don't believe we conclude that raising the crash requirements is the wrong thing to do, but rather because the cost implications are so great nothing less than a commensurately significant analysis and determination of need -- which has not been done -- should precede any decision.

It is easy to lose sight of the fact that a good intention, or want, or objective gets converted, by means of a regulation, into very specific operational requirements and specifications which the manufacturer must meet, specifications which may have little to do in the last analysis with the intentions of the regulation. However, the regulation, in its specific detail, is often defended on the basis of its motivation rather than on what the particular requirements of the regulation are

likely to actually accomplish. Specifically, in this case, if it is deemed desirable to provide better protection for those people who are in high speed crashes, then it may or may not follow that running an automobile into an immovable wall at 45 or 50 mph, and then comparing readings gotten on accelerometer in dummies against some mandated criterion level somehow validly signifies accomplishment of the societal goal which motivated the standard. The likelihood of gross erosion of relevance is probably nowhere better seen than in the accident avoidance series of standards, where little or no validation has been attempted.

A contrary argument is likely to be heard: that the need is so great we cannot wait for all the evidence to be in, that utterly adequate evidence will never be forthcoming, and thus we must act now. But such an argument seems to beg the question: for how can we know we must act now -- especially with some particular countermeasure -- if that determination depends on having adequate data? A variant on this argument is that it can do no harm and might do some good. But, without data there is no assurance that particular countermeasures will do no harm, and certainly a cost without a compensating benefit is a net harm.

I am attaching a COPY of the Ford docket submission on the higher speed protection requirements proposal, but you will probably want to give special attention to the brief summary, "societal Cost Implications of Inadequate Accident Data," which puts forth the main points made there.

In addition, I am attaching an updated copy of the remarks which I made at the Workshop. They are essentially the same as the statement I read, but there have been some additional clarifications which I felt were appropriate in view of the discussions which took place at the meeting.

Sincerely,

John Versace

John Versace
Executive Engineer
Safety Research

Attachments

## SOCIETAL COST IMPLICATIONS OF INADEQUATE ACCIDENT DATA

The demonstration of need for any safety standard must ultimately be established by accident data -- in all its forms -- if objective safety standard performance levels are to be achieved. If standard performance levels are established on a subjective basis, the possibility of very high societal cost with inadequate return for that cost is very real.

As an example of proposed performance levels which could have severe societal cost implications consider NHTSA's Advanced Notice of Proposed Rulemaking (ANPRM), Docket 74-15 Notice 1. This ANPRM proposes to increase frontal barrier crash requirements from 30 mph to 45/50 mph -- an increase in crash energy management requirements of 125 to 177% above that required today. The notice also proposes to implement the rule on September 30, 1980.

Ford Motor Company's response to this notice is attached. It presents the implications of implementing such a proposal in terms of increased car weight and car length. For example, to meet the frontal crash requirement alone, a 1974 Ford would be 500 pounds heavier and 16 inches longer; a 1974 Pinto would be 600 pounds heavier and 37 inches longer. Additional weight would be required to meet side and rear impact, roof crush, and fuel system crash requirements currently in being or presently proposed in other standards.

Weight increases of the magnitude discussed above imply completely re-designed cars -- not modifications to on-going designs. In addition to new metal structures, the added weight would require higher performance powertrain and running gear (brakes, suspensions> steering systems, etc.. . . . ) which in turn would tend to weigh more. Ford Motor Company markets 16 domestically manufactured car lines built from eight separate body shell platforms. To completely redesign these platforms would involve staggering engineering and investment costs. Annual increased car purchase costs to consumers -- assuming such a gigantic task could be done at all -- would be on the order of billions of dollars annually.

Such a major weight increase in cars would have a two-fold effect on the consumption of energy. The fuel economy of vehicles would deteriorate and secondly, additional energy would be used to manufacture the added weight.

Fuel economy may be expected to decrease from the current average of 13.6 miles per gallon by about 10%. This represents an increase in fuel usage of 25 million barrels each year. Should this weight increase be applied to the entire vehicle population, the annual fuel economy penalty would be nearly 200 million barrels. In ten years gasoline purchase costs would be on the order of $5 billion more per year than 1975.

Adding this weight to 10 million new cars each year would increase manufactured material requirements by about 3 million tons annually. The gross effect of the vehicle weight increases would be to increase the demand for finished steel, steel castings and rubber for the auto industry by about 20%. The energy consumption for manufacturing this added material weight in 10 million new cars each year would approximate 130 trillion B.T.U's.

If all the cars on the road were at the higher weight levels, the total annual cost increase to consumers would be the sum of the annual cost of the decreased fuel economy (projected at $5 billion), PLUS the higher costs and energy associated with manufacturing the heavier vehicles (projected to be billions of dollars annually). This sustained annual societal cost impact could take place because of a regulation whose need has not been definitely or definitively established.

Ford Motor Company

The American Road
Dearborn, Michigan 48121

September 19, 1974

National Highway Traffic
Safety Administration
Docket Section -- Room 5108
400 Seventh Street, S. W.
Washington, D. C. 20590

Gentlemen:

Re: Advance Notice Concerning Higher Speed Protection Requirements (Docket 74-15: Notice 1)

Enclosed are Ford Motor Company's comments on the Administration's advance notice of proposed rulemaking to increase the frontal barrier crash requirements of Federal Motor Vehicle Safety Standard No. 208, Occupant Crash protection, to 45 or 50 mph effective September, 1980. Ford has also participated in the preparation of comments being submitted by the Motor Vehicle Manufacturers Association and respectfully requests that those comments be incorporated herein by reference.

The comments address the several areas of interest cited by the Administrator in the subject advanced notice of proposed rulemaking. It is appropriate, however, to highlight certain salient points on which the comments expand.

There is the apparent assumption that a 'manifold increase in lifesaving capability of occupant crash protection systems" can be demonstrated merely by increasing the velocity at which a test vehicle impacts a fixed barrier and having the recorded test results satisfy essentially arbitrary criteria.

As the Administration well knows, there are many unsettled questions and unresolved issues with regard to Standard 208 including the correlation of test device responses to those of humans, the subjectivity of test procedure, the questionable appropriateness of the criteria, etc. Barrier crash tests are not representative of actual traffic accidents. Meeting some requirement using a test device having a superficial resemblance to a 50th percentile male adult positioned in a normal seated position is no guarantee that human occupants will survive in actual collisions of apparent equivalent severity.

Despite the uncertainty associated with Standard No. 208, in an effort to aid the Administration in defining the potential effects of adopting requirements such as those in this proposal, Ford has conducted a theoretical study related only to front end impacts using a Simplified model and idealized assumptions as to restraint systems, structure behavior, etc. That study, as explained in the attached comments, convinces us that the results of the Administration's proposal would be to increase the weight of a vehicle with a Pinto size passenger compartment by about 600 pounds and that of a Ford size vehicle by between 500 and 900 pounds for a 50 mph barrier impact speed. Length increases of as much as 37 inches for the Pinto and 16 inches for the Ford would be required. Specific modifications would be dependent upon restraint systems parameters that are yet undeveloped.

It is obvious that vehicle weight increases of this magnitude will have a pronounced effect on vehicle cost. The engineering and investment costs necessary for major redesigns of all existing cars in a short time period of a few years might best be described as staggering. Based on our analysis to date, Ford would not be able to meet the proposed effective date of September, 1980.

These weight and length increase estimates are based on a simplified, idealized analytical study and we consider them the minimum changes required, if only the requirement for front end impact speed was increased. It is significant that these results are not greatly dissimilar to those that could be derived from an analysis of the vehicle designed and built under the Experimental Safety Vehicle programs. It is also significant to note that none of the full sized Experimental Safety Vehicles were successful in meeting the requirements during a 50 mile per hour barrier crash despite, in some cases, the somewhat exotic designs employed.

Ford believes that the increased speed requirement with its attendant cost and weight increases cannot be justified without an analysis of highway accident data showing that a safety need exists for the proposed increase. The accident impact speed data currently available with which to perform a benefit analysis of higher speed requirements are dependent on subjective human evaluation. Speed estimates in existing data files are thought to be unreliable because they are formed by witnesses or by accident investigators having varying degrees of experience.

The lack of a sound data base with which to evaluate the need for higher speed performance requirements further underscores the need for a large scale crash recorder program to evaluate the actual crash dynamics. The initial results of crash recorder analyses have indicated that impact speeds estimated by police and accident investigation teams are consistently higher than the speed change noted by the recorder.

Ford is currently engaged in a research project under DOT contract to define the performance parameters of a 3000 pound safety vehicle which will be practicable to manufacture in the mid 1980's. We believe this research will be of value in evaluating future motor vehicle safety needs in the area of higher speed protection. This project is scheduled for completion in April, 1975.

We, therefore, recommend that NHTSA's efforts in the area of higher speed occupant crash protection be concentrated on developing an accurate data base from which the Administration can determine,on an informed basis, the safety need, if any, for a barrier crash test and identify appropriate and practicable test speeds.

At the present time we can only conclude that adopting the proposal advanced in this notice would have the certain effect of increasing weight and vehicle size (with the attendant adverse effects on fuel and material consumption) and consumer cost. The amount of benefit to be gained is only speculative.

If we can be of further assistance in explaining our position, we will be available at the Administration's convenience.

"Respectfully submitted,

J. C. Eckhold
Director
Automotive Safety Office

bgw

Attachments

ATTACHMENT
September 19, 1974

HIGHER SPEED PROTECTION REQUIREMENTS
DOCKET 74-15; NOTICE 1

COMMENTS OF FORD MOTOR COMPANY

Ford Motor Company, with Offices at The American Road, Dearborn, Michigan 48121, as a manufacturers of motor vehicles, is commenting on the Advance Notice of Proposed Rulemaking concerning Higher Speed Protection Requirements published in the Federal Register on March 19, 1974 (39 Fed. Reg. 10273).

The Notice states that the Administration is considering amending Federal Motor Vehicle Safety Standard No. 208 (FMVSS 208) to include a 45 or 50 mph frontal crash requirement with a suggested effective date of Septmber 1, 1980.

In our evaluation of the Administration% proposal, we found we were impaled by the lack of adequate factual information. Analysis of the available accident data lead us to the conclusion that such data are not sufficiently reliable to assess safety need.

Review of the public record on FMVSS 208 did not disclose the existence of technology which would show that a practicable vehicle could be designed to meet the frontal impact requirements of that Standard at 50 mph. The domestic ESV's, including the one built by Ford, represent the most comprehensive attempts to comply with such a requirement and all of them failed in that endeavor.

Nonetheless, we have gained some insight into the problem and have prepared the following comments based in part on engineering judgment, relying heavily upon theoretical studies.

Technology

The Administration states in the Notice that based on research which is extensively documented in the Docket on FMVSS 208, it is of the opinion that technology has advanced to the point where protection can be offered in crashes equivalent to those into a fixed barrier at more than 40 mph. We have examined the public record concerning FMVSS 208 and have found no evidence that the Administration has ever conducted the complete test series required by FMVSS 208 even at 30 mph, much less at 45 or 50 mph.

Technogy (Cont'd)

None of the domestic experimental safety vehicles built under DOT contracts met the performance requirements of FMVSS 208 at 50 mph. These vehicles exceeded the 4000 pound weight objective by between 1000 and 2000 pounds. One such vehicle even used unconventional lightweight materials in an effort to minimize weights. These materials are generally impractical for high volume automotive use because of supply limitations, high cost and lack of adequate manufacturing technology.

More recent higher speed research by NHTSA contractors has concentrated on maintaining passenger compartment integrity independent of programs to develop restraint systems* Advanced structures have not been evaluated in combination with advanced restraint systems in a 50 mph fixed barrier impact test series which would otherwise conform to FMVSS 208 although the intent to do so has been expressed in requests for contract proposals issued by NHTSA.

This was noted by Dr. Patrick Miller of Calspan Corporation in his statement before the Senate Commerce committee on February 21, 1974. He stated that "although impressive structural performance has been demonstrated during frontal collisions, we have not yet developed restraint systems which could take advantage of these advances".

Another problem which has not been adequately considered is the possibility of adverse consequences on occupants of vehicles designed for a 50 mph barrier impact when they are involved in lower speed impacts. The possibility exists that due to increase in vehicle stiffness the injury level in low speed collisions will Become worse.

Many of the crash tests have been conducted at test weights substantially less than that required by FMVSS 208. Under DOT Contract HS-257-2-461, 'Frontal and Side Impact Crashworthiness-Compact Cars" the contractor conducted the crash test without any dummy occupants and with the vehicle weight 700 pounds under that required by FMVSS 208. The effect of added weight is to place even greater demands upon the vehicle structure and, thus, to produce substantially different results.

Further, our review of structural integrity research under NHTSA contracts indicates that these efforts have not been directed toward designs which are practicable in high volume production. The usefulness of the resultant designs for commercial marketing has been inadequate in most cases. For example, the domestic ESV's were five passenger sedans with the occupants tightly packaged while the exterior

Technology (Cont'd)

dimensions were equivalent to current vehicles capable of carrying six passengers. One NHTSA contractor raised the body of a Pinto six inches higher off the ground and moved the driver four inches into the rear passenger space. (DOT Contract HS-113-3-746, "Crashworthiness of Subcompact Vehicles")

We anticipate that the structural modifications introduced to meet the 50 mph fixed barrier impact requirement would aggravate any existing car to car impact compatibility problems. The stiffer frontal *structure* and greater mass would have an effect in frontal, rear and side impacts.

Size and Weight Effects

There is only minimal data and limited experience with vehicle designs needed to approach a 45 or 50 mph fixed barrier frontal impact requirement. Therefore, we have attempted to extrapolate data from existing cars to determine the size and weight effects of the Administration's proposal. The results of Ford's and other domestic ESV programs, along with additional Ford research, were used even though the ESV's did not meet the occupant protection requirements of FMVSS 208 at 50 mph and exceeded the vehicle weight objective by large margins.

The test data used as a basis for the engineering assumptions and projections were gleaned from recorded force and acceleration measurements upon various anthropometric test devices. Though such data was found to lack repeatability, it nevertheless was averaged and used for directional guidance.

Simplified analytical techniques were used along with assumed performance parameters for advanced restraint systems to derive an estimate of the size and weight increases necessary to meet the proposal.

For purposes of this analysis, the parameters for an advanced air bag system and an advanced belt restraint system were hypothesized to represent restraint systems which are not currently available but which may be possible by September, 1980.

The results of numerous barrier crash tests were examined to evaluate the performance of various experimental and production belt and air bag restraint systems. Values for effectiveness time, rate of deceleration onset, and equivalent uniform deceleration or "square wave" deceleration were then determined. The key criterion was the 60 g deceleration limit of FMVSS 208. We concluded that for an

Size and Weight Effects (Cont'd)

advanced belt restraint system, a deceleration curve with an effectiveness time of 20 milliseconds, a uniform onset rate of 1200 g/see, and a constant deceleration of 40 g's gave an idealized representation of the deceleration which could be produced on the chest of an anthropometric test device. For an air bag, the values of 40 milliseconds effectiveness time, 1500 g/see and 48 g's were determined. The deceleration levels represent the square wave that would simulate the average of the peaks and valleys of a dynamic curve in which the peaks would still remain under the 60 g limit of FMVSS 208. Onset rates and effectiveness times were chosen based on predicted future system performance capabilities.

The advanced belt system would include a crash sensor and a preloader device and possibly a load limiting webbing material. The advanced air bag system would require developing improvement to present systems to achieve effectiveness within 40 milliseconds.

The restraint system parameters were used with a simple mathematical model consisting of two point masses representing vehicle and occupant. Idealized occupant stopping distances were determined and then compared with the available vehicle crush and interior occupant space. The vehicle deceleration necessary to produce the assumed occupant deceleration was also computed.

The output of the simple mathematical m-cl thus gives an indication of the amount that a vehicle must be lengthened or stiffened to approach a 45 or 50 mph barrier impact requirement. The length and stiffness increases were used to determine weight effects using engineering judgment based on Ford experimental results and ESV experience, and a review of the ESV's designed by others.

One particular assumption included in the length calculations is that 65% of the added length will actually crush during impact. Deformed structure would occupy the remaining 35% of space. The frontal area occupied by relatively incompressible components such as the engine are considered unavailable for vehicle crush. However, the space occupied by the engine was also considered available for the deformed structure. For purposes of this analysis, length added to the vehicle was considered totally usable for computing crush distance up to the point where the 65% efficiency level was reached. After that point, 1.54 inches of vehicle length were added for each inch of crush length needed.

Size and Weight Effects (Cont'd)

The resultant length increases, stiffness, and weights are shown in Fig. 1 for a vehicle with a Ford size passenger compartment and Fig. 2 for one with a Pinto size passenger compartment.

The Ford size car with an advanced air bag system intended to meet a 50 mph impact level would be over 16 inches longer and an estimated 530 pounds heavier than the current Ford. The same car with an advanced belt restraint would only be 2.4 inches longer than the 1974 model but would be nearly 900 pounds heavier.

The Pinto size car with an advanced air bag system intended to meet the same 50 mph requirement would become 37 inches longer and an estimated 600 pounds heavier than the 1974 version. Under the assumptions for the advanced belt restraint, the Pinto would be 18 inches longer and 630 pounds heavier than the existing car.

Front end structural stiffness would have to be increased substantially for both cars with either restraint system.

Lesser, although dramatic, weight increases would result on both Ford and Pinto size vehicles as shown in Figures 1 and 2 if a 45 mph barrier impact goal were established.

These weight increases are estimates for meeting only frontal impact requirements. No provision has been made in this estimate for increased side, rear and roof structure which we anticipate would be necessary to meet the existing levels of such Standards as FMVSS 214, Side Door Strength, FMVSS 216, Roof Crush Resistance and FMVSS 301, Fuel System Integrity. Structual modifications would be necessary to withstand the increased static or dynamic test loads imposed as a result of the weight added to the vehicle to meet the increased frontal impact speed. The weight increase resulting from these side, rear and roof structural modifications would cause further changes to be made in the frontal structure to meet frontal requirements. These effects would be more pronounced on small cars under 3500 pounds curb weight due to the provisions regarding curb weight in FMVSS 214 and FMVSS 216. Neither is there provision in these weight estimates for revision or deletion of any other standards.

Size and Weight Effects (Cont'd)

The weight and length additions shown in Figures 1 and 2 were derived, in part, using simplified analytical techniques which do not fully consider the dynamic interactions of vehicle structure, restraint system and test device. They represent minimum levels of vehicle modification which we believe would be necessary to approach the frontal impact performance levels of FMVSS 208 at 45 and 50 mph. Restraint system performance parameters were chosen which we believe are possible by 1980, but do not represent any system which we currently have available. Vehicle structures with the necessary frontal crush characteristics would have to be developed. Objective, repeatable conformance demonstration procedures for FMVSS 208 have yet to be developed. We therefore consider these estimated weight and length increases to be minimum levels.

The weight increase shown in Fig. 1 includes that due to structural additions to meet the higher barrier speed requirement plus added weight to upgrade such areas as engine, brakes, suspension and steering. Weight estimates for these other systems were determined by increasing their weight in proportion to the increase in structural weight. This was done by determining the portion of total vehicle weight due to the other systems for several large size vehicles as shown in Fig. 3. The portion of total weight contributed by each system was found to remain fairly constant. The increased weight of these systems was computed by an iterative process based on the added structure weight. This process would add weight to the various supporting systems for each pound of crashworthiness structure added. We realize that in a practical sense weight additions occur in discrete increments.

A similar analysis was conducted for smaller size vehicles to determine the weight additions for a Pinto. (See Fig. 4) .

cost

We have not determined the cost effect of the proposal, but it is obvious that addition of this amount of weight will result in substantial vehicle cost increases. The engineering and investment cost to redesign all of our vehicles to attempt to meet a 45 or 50 mph requirement would be staggering.

Timing

The vehicle modifications required to meet a 45 or 50 mph barrier impact requirement are so extensive that we would be required to redesign all of our affected vehicles. After a final rule of this type is established, technology

Timing (Cont'd)

is available, and practicability is achieved, it would take approximately three years to redesign and retool a single car line family.

Ford normally cannot develop more than two totally redesigned car line families in the same model year due to manpower and facility limitations and available capacity within the tooling industry. It would require a total of four additional years to introduce new designs of all existing passenger car models. However, Ford has never before undertaken a task of this magnitude. Even this cycle is optimistic as it is unlikely the tooling industry could contain the magnitude of such programs if all domestic automobile manufacturers found it necessary to implement similar redesigns.

On the basis of our analysis to date, we could not meet a Septmber, 1980 effective date for all cars, even if the means of meeting the proposed requirements were fully developed. Due to the uncertainties that now exist, we cannot determine whether or not we can meet this date even on one car line.

A new car body and chassis design is produced for a minimum of three years and in many instances can exist for eight years before a major redesign. Therefore, the redesign program that would be required by the proposal would probably obsolete relatively new car lines before the end of their normal cycle with additional cost consequences.

The precise timing effects of the Administration's proposal have not been determined. Small cars would cease to exist as they are known today and large cars might well become impracticable due to increased size. We do not know what vehicle model mix the market would support if it is artificially constrained by a requirement which has such a pronounced effect on vehicle size.

Accident Data Analysis

Ford and others have previously noted the unreliable nature of reported accident speeds available for analysis. The source of data errors and some of the methods which have been used to adjust these data are shown in Exhibit I. Recent crash recorder results have confirmed that reported crash speeds are usually too high.

Accident Data Analysis (Cont'd)

Twenty accident cases involving vehicles equipped with crash recorders were summarized in SAE Paper 740566 by S. S. Teel et al[1] of the National Highway Traffic Safety Administration (NHTSA). The results of an analysis comparing each case vehicle's velocity change, as reported by the police and/or an accident investigation team, are summarized below. The impact speeds used in this analysis and their differences are contained in Exhibit II.

The accident cases in Teel's paper which contained the necessary information were used to construct a sample of the population of differences between velocity "changes estmated by an accident investigator and the velocity change experienced by the vehicle, as reflected by the crash recorder. The sample of 22 differences as tested for normality using the Kolmogorov-Smirnov test[2] and the hypothesis that the population of impact velocity change differences is normally distributed could not be rejected. Although our sample of accident cases is small, it indicates that the distribution of the difference in estimates is a bell-shaped curve centered at 14 mph (the sample mean) with an estimated standard deviation of 11.9 mph. Using these figures, we are 95% confident that ten percent of the reported impact speeds overestimate the true change in velocity by at least 35 mph while one-quarter of them overestimate the true change in velocity by at least 25 mph.

An interval which contains the true mean difference between the estimated and the recorded velocity change of a vehicle in an accident, with 99% confidence, was constructed using the Students-t distribution. This interval, 7.1 mph < Mean Difference < 21.4 mph, indicates that, on the average, accident investigators can be expected to overestimate accident impact speeds by from 7 to 21 mph. Our

- - - - -

1 Teel, S. S., Pierce, S. J., and Lutkefedder, N. W., "Automotive Recorder Research --A Summary of Accident Data and Test Results", SAE 740566, 3rd International Conference on Occupant Protection, July, 1974.

2 Lilliefors, H. W., "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown", JASA, June, 1967.

Accident Data Analysis (Cont'd)

accident sample also indicates that impact speeds can be overestimated by as much as 40 mph. These large overestimates do not depend on the magnitude of the crash recorder velocity change.

As an alternative statistical test, a non-parametric test, the Wilcoxon Matched-Paris Signed-Ranks Test, also indicates that estimated impact speeds from accident investigators are positively biased. Based on crash tests, Teel concludes that changes in velocity reported by crash recorders are accurate to within ± 2 mph. Therefore, as a conservative approach, the differences between the estimated and the recorded changes in velocity in Exhibit II were reduced by 5 mpht ad the Wilcoxon test was re-run to determine if the velocity differences could be due to the crash recorder accuracy. The results still indicate that impact speeds estimated by police and accident investigators are too high.

The lack of a sound data base with which to evaluate the need for higher speed performance requirements further underscores the need for a large scale crash recorder program to evaluate actual crash dynamics.

EXHIBIT I
September 19, 1974

# THE TREATMENT OF RECORDED IMPACT SPEEDS

## - A Summary-

Methods which have been used to deal with reported impact speeds from the ACIR accident case file are summarized below.

A. Cooke, Conrad H., "Safety Benefits of The Occupant Crash Protection Standard", January, 1971.

Cooke reduced all reported traveling speeds by 10 mph to obtain his estimated impact 'speeds.

B. Mela, Donald F., "A Source of Substantial Error In Estimating The Distribution of Traveling Speed For Accident-Involved Vehicles .*O ", DOT, Septmber 3, 1968.

Mr. Mela stated that, by using the estimated impact speeds to determine speed distributions, "the fraction of vehicles in the speed ranges 20-30 mph and 70-80 mph is overestimated by a factor of 3, and the fractions below 20 mph and above 80 mph are overestimated by a factor of 17". If this statement is true, then it suggests that some variable type of correction factor (and not a constant 10 mph as Cooke used) be applied to the estimated impact velocities in the ACIR file.

C. White, S. B., Nelson, C., "Some Effects of Measurement Errors in Estimating Involvement Rate as a Function of Deviation from Mean Traffic Speed", Journal of Safety Research, Volume 2, June, 1970.

White and Nelson show that even if errors in estimation are non-systematic, an overestimate of high-speed frequency would be found. That is because any error of measurement always serves to inflate the variance of the distribution of reported values, regardless of the nature of the data. Thus, reported variance (i.e., the mean-square deviation from the mean) is equal to the sum of "true)' variance and "error" variance. White and Nelson point this out, in suggesting that high speed estimates would tend to be exaggerated. They state that "errors in estimating speeds of accident-involved vehicles causes the

involvement rate, when plotted as a function of the speed deviation, to be U-shaped -- overestimated for large derivations (from the mean) and underestimated for small deviations". White and Nelson refer to traveling, not impact, speed, but the principle is the same in either case.

D. Grush, E. S., Henson, S. E., and Ritterling, o. R., 'Restraint System Effectiveness", Report No. S-71-40, Ford Motor Company, September 21, 1971.

In this report, ACIR impact speeds were converted to barrier-equivalent velocities. The following factors were considered in the conversion: the estimated relative closing speed; the weight differential; a center of gravity adjustment; and an accident location adjustment. A second method of obtaining the barrier-equivalent value for each accident-involved vehicle was based on photographs of the vehicle damage and the study showed that this latter method produces better results.

E. Mason, R. R., D. W. Whitcomb, "The Evaluation of Accident Impact Speed", CAL Report No. YB-3109-V-1, August, 1972.

This report presents several formulas, one for each type of vehicle impact, which can be used to estimate a vehicle% impact speed. It provides some insight into how Calspan may estimate impact speeds.

EXHIBIT II
September 19, 1974

IMPACT VELOCITY CHANGES

| Recorder Number | Crash Recorder Velocity Change (mph) | Accident Investigator Estimated Velocity Change (mph) | Difference (mph) |
|---|---|---|---|
| 1086 | 20 | 60 + | + 40 |
| 485 | 15 | 50 | + 35 |
| 485 | 15 | 50 to 60 | + 35 |
| 642 | 10 | 30 | + 20 |
| 322 | 5 | 25 | + 20 |
| 335 | 6 | 25 to 30 | + 19 |
| 641 | 13 | 30 | + 17 |
| 694 | 9 | 25 | + 16 |
| 596 | 10 | 25 | + 15 |
| 596 | 10 | 24 to 26 | + 15 |
| 596 | 10 | 25 | + 15 |
| 641 | 13 | 25 to 35 | + 12 |
| 642 | 10 | 22 to 25 | + 12 |
| 306 | 19 | 30 | + 11 |
| 463 | 19 | 30 | + 11 |
| 463 | 19 | 30 | + 11 |
| 485 | 15 | 25 | + 10 |
| 25 | 18 | 25 to 35 | + 7 |
| 352 | 15 | 22 | + 7 |
| 463 | 19 | 20 | + 1 |
| 94 | 11 | 5 to 8 | - 6 |
| 352 | 15 | 5 | - 10 |

- - - -

Reference: Teel, S. S., Place, S. J. and Lutkefedder, N. W., "Automotive Recorder Research -- A Summary of Accident Data and Test Results", SAE 740566, 3rd International Conference on Occupant Protection, July, 1974.

FORD

SUMMARY: LENGTH AND WEIGHT INCREMENTAL EFFECTS FOR FRONTAL IMPACT

| SPEED | RESTRAINT SYSTEM TYPES | AVERAGE g's S.W.E 1/ | WHEELBASE LENGTH INCREASE 2/ | VEHICLE WEIGHT Additions (LBS.) STRUCTURE | RESTRAINT SYSTEM | BUMPER SYSTEM | ENGINE SYSTEM | DRIVELINE SUSPENSION AND BRAKES | FUEL SYSTEM INCLUDING FUEL | STEERING SYSTEM | TOTAL WEIGHT INCREASE | CURB WEIGHT (LBS | INCREMENT OVER BASE (%) | FMVSS 208 TEST WEIGHT (LBS) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 30 | Product ion | 11.3 | ← | Base | | | | | | | → | 4400 | Base | 5600 |
| 45 | A | 19.3 | 6.0 | 182 | 20 | 21 | 53 | 69 | 6 | 7 | 358 | 4759 | 8% | 5958 |
| 45 | B | 22.5 | 0 | 283 | 20 | 31 | 81 | 104 | 9 | 10 | 538 | 4938 | 12% | 6138 |
| 50 | A | 19.6 | 16.2 | 282 | 20 | 31 | 80 | 104 | 9 | 10 | 536 | 4936 | 12% | 6136 |
| 50 | B | 26.0 | 2.4 | 483 | 20 | 52 | 134 | 172 | 14 | 17 | 892 | 5292 | 20% | 6492 |

Restraint Type A: Air bag - 40 msec. effectiveness - 1500 g/ sec. onset - 48g max.

Restraint Type B: Seat belt with sensor and preloader - 20 msec. effectiveness - 1200 g/SeC. onset 40g max.

1/ Square Wave Equivalent of vehicle deceleration pulse based on impact speed and total crush *distance.*

2/ Crush length increases in excess of 5 inches are adjusted by a 65% efficiency factor

Figure 1
September 19, 1974

PINTO

SUMMAR EN TH AND WEIGHT NCREMENTAL TS FO FRON AL MP

| SPEED | RESTRAINT SYSTEM TYPE | AVERAGE g's S.W.E. 1/ | LENGTH INCREASE 2/ 3/ | VEHICLE WEIGHT ADDITIONS (LBS.) | | | | | | | | CURB WEIGHT (LBS.) | INCREMENT OVER BASE (%) | FMVSS 208 TEST WEIGHT (LBS.) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | STRUCTURE | RESTRAINT SYSTEM | BUMPER SYSTEM | ENGINE SYSTEM | DRIVELINE SUSPENSION AND BRAKES | FUEL SYSTEM INCLUDING FUEL | STEERING SYSTEM | TOTAL WEIGHT INCREASE | | | |
| 30 | Production | 14.4 | ← Base → | | | | | | | | | 2457 | Base | 3300 |
| 45 | A | 18.4 | 23.5 | 188 | 20 | 23 | 57 | 83 | 8 | 7 | 386 | 2843 | 16% | 3686 |
| 45 | B | 25.0 | 5.7 | 227 | 20 | 27 | 69 | 99 | 10 | 8 | 460 | 2917 | 19% | 3760 |
| 50 | A | 18.9 | 37.2 | 305 | 20 | 36 | 89 | 130 | 13 | 10 | 603 | 3060 | 25% | 3903 |
| 50 | B | 24.7 | 18.0 | 319 | 20 | 37 | 94 | 136 | 13 | 11 | 630 | 3087 | 26% | 3930 |

Restraint Type A: Air bag - 40 msec. effectiveness - 1500 g/sec. onset - 48g max.

Restraint Type B: Seat belt with sensor and preloader - 20 msec. effectiveness - 1200 g/sec. onset - 40g max.

1/ Square Wave Equivalent of vehicle deceleration pulse based on impact speed and total crush distance.

2/ Crush length increases in excess of 5 inches are adjusted by a 55% efficiency factor

3/ 2.8 inches can be added to front overhang without wheelbase increase (19° minimum approach angle). (wheelbase increase is 2.8 inches less than length increase)

Figure 2
September 19, 1974

September 19, 1974

Figure 3

WEIGHT OF VARIOUS VEHICLE SYSTEMS
AS A PERCENTAGE OF TOTAL WEIGHT

| | RSV* | TORINO | FORD | LINCOLN | AVERAGE |
|---|---|---|---|---|---|
| Curb Weight: | 3000 | 4030 | 4398 | 5373 ! | |
| Percentage of Curb Weight: | | | | | |
| Bumper Systems | 6.0% | 1 5.9% | 5 4% | 5 [illegible]% | 5.8% |
| Engine | 15. 6% | 14• 2% | 15• 8% | 15 0% | 15 .0% |
| Suspension Driveline Brakes | 21. 3% | 19.8% | 18. 5% | 17• 5% | 19.3% |
| Fuel System: | To maintain the current Ford vehicle range fuel system weight should be increased at-the rate-of .01415 lb. per lb. of added vehicle weight. The fuel tank weight is approximately 17% of the total fuel system weight. | | | | 1.6% |
| Steering | 2.0% | 1.5% | 2.3% | 1.8% | 1.9% |
| **TOTAL:** | **46.5%** | **43.0%** | **43.6%** | **41.5%** | **43.6%** |

- - - - -

*RSV figures are an average of 10 Unitized vehicles with curb weights from 2000 to 3300 lbs.

Figure 4

WEIGHT OF VARIOUS VEHICLE SYSTEMS
AS A PERCENTAGE OF TOTAL WEIGHT

| | PINTO | MUSTANG | MAVERICK | GRANADA | AVERAGE |
|---|---|---|---|---|---|
| Curb Weight: | 2457 | 2753 | 2831 | 331.9 | |
| Percentage of Curb Weight: | | | | | |
| Bumpers | 6.1% | N. A. | 6.0% | 5.7% | 5.9% |
| Engine | 14• 0% | 14. 6% | 14• 996 | 15• 9% | 14• 9% |
| Suspension Driveline Brakes | 21• 3% | 21• 7% | 22• 1% | 21• 3% | 21. 6% |
| Fuel System: | To maintain the current Pinto vehicle range fuel system weight should be increased at the rate of .01415 lb. per lb. of added vehicle weight. The fuel tank weight is approximately 17% of the total fuel system weight. | | | | 2.0% |
| Steering | 1.7% | 1.5% | 2.0% | 1.7% | 1.7% |
| TOTAL: | 45.1% | | 47.0% | 46.6% | 46.1% |

APPENDIX C

# STATEMENT AT THE AUTOMOBILE COLLISION DATA WORKSHOP

B.J. Campbell
HIGHWAY SAFETY RESEARCH CENTER
University of North Carolina

January 17, 1975

STATEMENT BY B. J . CAMPBELL
University of North Carolina
Highway Safety Research Center

Presented at
Automobile Collision Data Workshop

Sponsored by
Economics & Science Planning

In acquiring automobile accident data several approaches are used in the U.S.: First, are intensively investigated accident crashes of which several thousand have been collected. The advantage of this approach is that the cases are extremely detailed with photographs and good injury data. The most important disadvantage is that by virtue of the changing sampling criteria and the small sample size, the ability to generalize these few cases to the population is restricted heavily.

I believe too much reliance has been made on this type of data for guiding NHTSA decisions. It leads one to situations in which too much is made of a small number of cases. For example, in interpreting the 35 or 40 crashes in which air bags are present some feel the crashes support air bags because relatively few moderate or serious injuries occur. However, what if these air bag cases were matched with several hundred cases in which no protective systems are used at all (i.e. no belt or bag)? What if one found pretty much the same proportion of injuries in both series? Wouldn't that suggest that 40 cases is just not enough?

Second is an approach called the tri-level system. There the samples are larger, but the negative aspect is that the reporting threshold is based on accident severity which results in eliminating certain cases in which safety belt and perhaps other safety device effectiveness is greatest.

Third, and at the other extreme from individual case studies is the attempt to use an entire state accident data system as the basis on which to do research and make decisions. The biggest advantage in this case is the perspective gained from very large sample sizes and the ability to partition and control the data. But on the negative side many such systems contain too few content variables of interest. The quality of reporting may be poor and the injury data is crude.

In my opinion a crucial need in the field of crash injury is the means to forge a meaningful link between laboratory test crash data and events as they occur in the field. Much can be gained from laboratory sled and full-scale crash tests involving dummies, Cadavers or even live subjects, and also much can be gained from the study of actual crashes on the highway. But each lacks a significant variable.

In the staged crashes in the laboratory, telemetric procedures are used for recording data and one can specify in considerable detail the physical system in which the crash occurs--the "g"-forces, the rate of onset, delta "v" etc. But when one is forced to use nonhuman subjects then one is left in the situation of knowing a great deal about the physics of the crash but knowing little of the actual injuries that might have occurred in such a crash. On the other hand, in real world automobile crashes one can learn about the actual outcomes in terms of survival and injuries, but the input variables mentioned before are unknown.

The need to link these two systems is apparent. Engineers who design protective systems need to know about stopping distances, forces, decelerations, etc. But knowing these things is of too little help unless one has a way to relate them to real world injuries. An

illustration of the need for this data link is the NHTSA analysis conducted in connection with the air bag. This NHTSA analysis initially indicated that lap and shoulder belts would only reduce fatalities by 35-40 percent, and that lap belts alone would be of almost no benefit at all in reducing injuries. These conclusions were presumably based in large measure on results of crash tests involving cadavers and dummies. The problem is that these conclusions disagree sharply with studies of tens of thousands of crashes that have occurred on the highways. Studies from all over the world indicate that in actual crashes injuries are reduced by lap belts, and that lives are saved, and that the degree of lifesaving is much higher than 35 or 40 percent NHTSA has indicated.

It is the very occurrence of this type of disagreement that shows that the analysis system in each sector (laboratory vs highway) by itself is inadequate and that means must be found to bridge the gap. The primary advantage of a crash recorder program would be a means to forge this link between the two data systems. It would finally be possible to gather data on a few thousand actual highway crashes in which crash conditions, the decelerations, the forces, the amplitudes and so forth would be knowable as well as the injury.

By using these several thousand crash recorded events as a calibration standard it would be possible to work outward to the hundreds of thousands of other actual crashes in which recorders weren't available, and the thousands of lab tests in which recorders are available but human injury is not.

It is not necessary to have an "infinite" number of crash recorders in the field, only enough to validate other approaches. I personally

do not see the crash recorder program as an end in itself but one which would support and validate other types of crash studies.

My remarks do not suggest the level of detail needed from the crash recorder, but in any case, the program will be expensive. For six million dollars one could equip 100,000 cars with crash recorders that cost $60 each. It would also be possible to equip more cars with a simpler, less costly crash recorder.

It is for others to determine the needed complexity of the crash recorder. Perhaps it is not necessary to have a crash recorder that records force time histories in three dimensions. Maybe vertical accelerations can be sacrificed.

Perhaps it would also be useful to consider a "tri-level" crash recorder program; this could involve a modest number of cars equipped with a very complex recorder and a larger number of cars equipped with a simpler, less expensive recorder system.

As a prelude to the program it might be appropriate to have a research project to synthesize past laboratory crash data to try to agree what measure in the field is the one that would account for the most injury variance. Would it be impact speed, barrier equivalent velocity, delta "v" or what?

The crash recorder, of course, is not the only need in studying and understanding actual crashes. Much better and much larger collections of highway crash cases are also necessary. I stated my belief that too much reliance has been placed on the small number of intensively investigated crashes. This country needs a multi-state data collection program which would accumulate records on 600,000 to 1,000,000 accident-involved vehicles per year. This would require three to five states the size of North Carolina.

For a surprisingly small additional cost it would be possible to collect that many cases with reasonably good deformation data, an operationally defined injury scale, vehicle identification numbers, belt usage, and various file linkage numbers to cross-link accident data and driver history files, road data, etc. It is extremely important to have this quantity of data in order to get timely answers to questions. If a safety device has gone astray or a dangerous car is coming onto the market--we need to know it soon--not after ten years.

I would be pleased at some future time to discuss some of the characteristics such a multi-state data system should have.

APPENDIX D

# SUMMARY OF REMARKS AT THE AUTOMOBILE COLLISION DATA WORKSHOP

Lawrence Patrick
Wayne State University

January 16, 1975

**WAYNE STATE UNIVERSITY**

BIOMECHANICS RESEARCH CENTER
COLLEGE OF ENGINEERING
5050 ANTHONY WAYNE DRIVE

DETROIT, MICHIGAN 48202
TELEPHONE: (31 3) 577-3835

January 20, 1975

Mr. Howard Gates
Economics & Science Planning
1200 18th Street
N.W. Washington, D.C. 20036

SUBJECT: ESP Meeting, January 16 & 17, 1975

Dear Mr. Gates:

As requested at the captioned meeting, I am enclosing herewith prints of the slides I used in my presentation together with a brief summary of my remarks. In the interest of brevity, the remarks are presented in outline form.

PREMISE

1. The only valid way to establish safety needs for automobiles is through examination of field data.

2. The only valid way to evaluate the effectiveness of safety measures is through analysis of their effect on accident data.

CONCLUSION

Accident data are essential.

CRITERIA FOR DATA COLLECTION

1. Sufficient data must be obtained for statistical analysis. Collection of accident data is expensive so it must be optimized for the number of variables, depth of study, and type of collision to minimize the cost per accident. The present MDAI studies cost approximately $2500.00 apiece, and include greater detail than is necessary. With modification of the collection procedure accident data in sufficient depth should be available at a cost of under $400.00 per case. Other data should be gathered on a large sample basis in even less detail at a considerably lower cost.

2. Complete injury data must be included in the accident data. Sex, age, weight, height, and general physical condition are all important factors in analyzing accident data. The type and degree of injury of each occupant including the minor bruises and abrasions and going through the severe bone and soft tissue damage are required. It is important to have complete data on the restraint systems used and the interior components of the vehicle that caused the injury.

3. Complete vehicle crash data are essential to permit an estimate of the collision severity. The crash data in addition to the usual photograph should include measurements of vehicle deformation. A standard means of recording deformation of the vehicle would be beneficial.

4. Reference collisions are required to establish severity of the accident from the crash data and deformation measurements. Eventually the reference collisions and deformation data can probably be replaced by a data recorder. The data recorder should be relatively simple and the cost should be low enough to permit installation in all vehicles. A crash severity signature is required which gives crash severity in the most meaningful terms. This does not necessarily require triaxial acceleration time histories. The Barrier Equivalent Velocity that has been used extensively is not necessarily the best measure of severity, but is one that has been used extensively and should continue to be used until a better measure of severity is developed.

DATA ANALYSIS

1. Standardized injury and deformation reporting is essential to keep the results of investigations by different groups in different parts of the country on a uniform basis. The AIS scale and the VDI should be considered for the immediate future and utilized until a better scale is devised.

2. The effect of sex, age, weight, size, position in vehicle, direction of impact, restraint systems etc. should be established. This will permit an accurate judgement to be made of the area of safety improvement that should be stressed.

3. Probability of injury as a function of collision severity is essential. It should be recognized that some individuals are going to be injured severely at low severity clue to inherent weaknesses. Fundamentally, it is necessary to protect the maximum number of people from the maximum number of exposures. From a design standpoint, it is essential to establish an acceptable degree of injury under the most severe collision conditions. It is recommended that the AIS-3 injury be the maximum acceptable injury with no injury as the ultimate goal.

EXAMPLE: WSU-VOLVO STUDY

1. The WSU-VOLVO study was divided into four major divisions as follows:

a. Accident Investigation - complete injury data including the AIS rating and complete vehicle deformation measurements.

b. Staged Collisions - complete deformation data in terms of impact speed.

co Simulation Tests - records of injury criteria as a function of simulated speed.

d. Analysis - Injury data related to severity and test data.

The accident investigation was conducted by the Volvo investigation team with special instructions to meet the requirements of this study. The staged collisions included frontal force, barrier, pole, and car to car collision. The collision simulations were made in the laboratory in a modified Volvo automobile with instrumented dummies as the occupants using the same stopping distance and deceleration pulse as measured from the staged collisions.

2. Accident criteria established to minimize the number of variables include:

a. Frontal force collisions only.

b. Belted front seat occupants (one or more).

c. No unbelted rear seat passengers or other heavy objects in the rear seat.

d. No external secondary impact of substantial severity.

3. With these stipulations, a total of 128 accidents were investigated with 169 occupants in a two year period. During this time there were eleven staged collisions at Volvo and 72 simulated tests at Wayne.

4. Figure 1 is a plot of the injury as a function of Barrier Equivalent Velocity with three injury areas for each occupant. As noted from the legend, the data are divided into head, neck, and chest injuries for each occupant with the driver and right front passenger position differentiated. The figures at the bottom of the graph refer to the number of body areas at each velocity for which there were no injuries. It is important to note that AIS-3 injuries were found at velocities ranging from 10 to 53 mph with the major number clustered at about 30 mph.

Figure 2 is a bar graph showing the distribution of injury as a percent of the number of occupants in 10 mph increments. At the 0 to 9 mph level approximately 90% of the occupants had no injury and the remaining 10% sustained only minor injuries. In the 50 to 59 mph range all occupants had some injury with one third having the AIS-1 injury and two thirds having AIS-3 injury. It is obvious that as the BEV increases the injury also increases.

Figure 3 is a sketch of the rib cage with rib fractures and sternum fractures illustrated. In the field study all of the rib fractues occurred on the inboard side which is the side which the belt applies the force to the ribs. The fractures have all been put on one side although in the field there were fractures to the driver and passenger and consequently they were on both the left and the right side of the rib cage.

5. The accident investigating team carefully measured the deformation of the vehicle at six different points on the front as shown in Figure 4. A computer program was developed to record the six deformation measurements in graphical form. Figure 5 shows the deformation for the staged barrier collisions. These were all normal frontal force collisions and consequently the deformation is symmetrical. Figure 6 shows the same data obtained from the measuring fixture in the field accident study. It will be noted that in this figure the impacts are to poles and/or asymmetrical impacts which result in a different pattern than the barrier results. It was necessary to interpolate the field data to provide the closest BEV for the analysis. It is felt that the overall barrier equivalent velocity assigned to each collision is considerably more representative of the collision severity than in previous studies.

6. Figure 7 shows the rib fractures for male and female as a function of velocity. It should be noted that the age of the occupant should be included as another variable. However, the figure shows that the female has a greater number of ribs fractured than the male.

7. Figure 8 is a graph of cumulative injury risk as a function of abbreviated injury scale with velocity as a parameter. The data are plotted for the 10 mph increments. The dash lines indicate that the data are extrapolated with insufficient data for an exact definition of the curve. However, the data show a distinct family of curves. Additional data is required to delineate the curves with greater accuracy. The same data are shown in Figure 9 with abbreviated injury as a function of barrier equivalent velocity. This graph permits an estimate of the likelihood of injury in a given frontal force collision.

AMOUNT OF DATA REQUIRED

1. The collection of accident data requires a substantial amount of data with extreme accuracy desirable but not necessary. For example, there is no need to have a collision severity to within plus or minus "one mile per hour". This is especially true since we really don't know what the barrier equivalent velocity means or whether some completely different severity index should be used. With the large number of variables including impact velocity, impact direction, rigidity of vehicle, rigidity of object struck, location of impact on car, occupant location, occupant age, sex, height, weight, physical condition, tolerance

to acceleration environment, posture, vehicle interior design, and restraint systems, it is more important to have a substantial amount of data with reasonable detail rather than a small number of cases that have been investigated to a great depth.

2. With the large number of variables it is necessary to have a large number of recorders in the vehicle population in order to obtain a reasonable number of accidents with the recorders in the car. The most desirable situation is one in which each car manufactured is equipped with a recorder installed at the factory.

CRASH RECORDER REQUIREMENTS

1. The crash recorder should be installed in a large number of vehicles. Consequently, it must be low in cost.

2. The recorder does not have to be ultra-accurate (such as plus and minus one percent on the acceleration and time scale), since the analysis will be based on a large amount of data rather than a small sample which would require the greater accuracy.

3. The crash recorder should be based upon a "severity index" that has yet to be developed depending upon the injury potential to the occupants. Such a recorder could be an integrating accelerometer with electronics to perform necessary operations on the accelerometer output to provide the severity index. Other means that might be satisfactory include fracture of a number of elements in the accelerometer or the deformation of an element in the accelerometer. The exact function to be measured and the method of measuring it has to be developed.

4. The crash recorder should be developed in conjunction with the data analysis group to insure maximum utility from the installation of the recorder.

5. The recorder should be sealed to prevent tampering and to guarantee that when the record is interpreted it has not been damaged prior to being collected by the investigator. It should be designed to give a record for a collision in excess of some predetermined severity such as a 10 mph barrier equivalent or greater. This will avoid the danger of having a recorder in multiple crashes which could confuse the data or give false results. Obviously the recorder must be rugged enough to withstand the collision without damage.

I believe that you or Dr. Goldmuntz requested a copy of my curriculum vitae and list of publications. They are enclosed.

I thoroughly enjoyed the meeting on January 16th and 17th and feel that it was productive in that I learned considerably from it. Hopefully, the goals of the meeting will be achieved. Bob Cromack has the preliminary writeup that we came up with during our working

lunch on Friday. He is going to have it typed up and sent to the rest of us (Brian O'Neill and David Morganstein). We will review it and approve or modify it for final submission.

An invoice for my expenses is enclosed in accordance with our agreement.

It was a pleasure to work with you on this program. If I can be of any further assistance, please don't hesitate to call on me.

Sincerely,

L. M. Patrick
Professor

LMP:ldd
ENCLOSURE

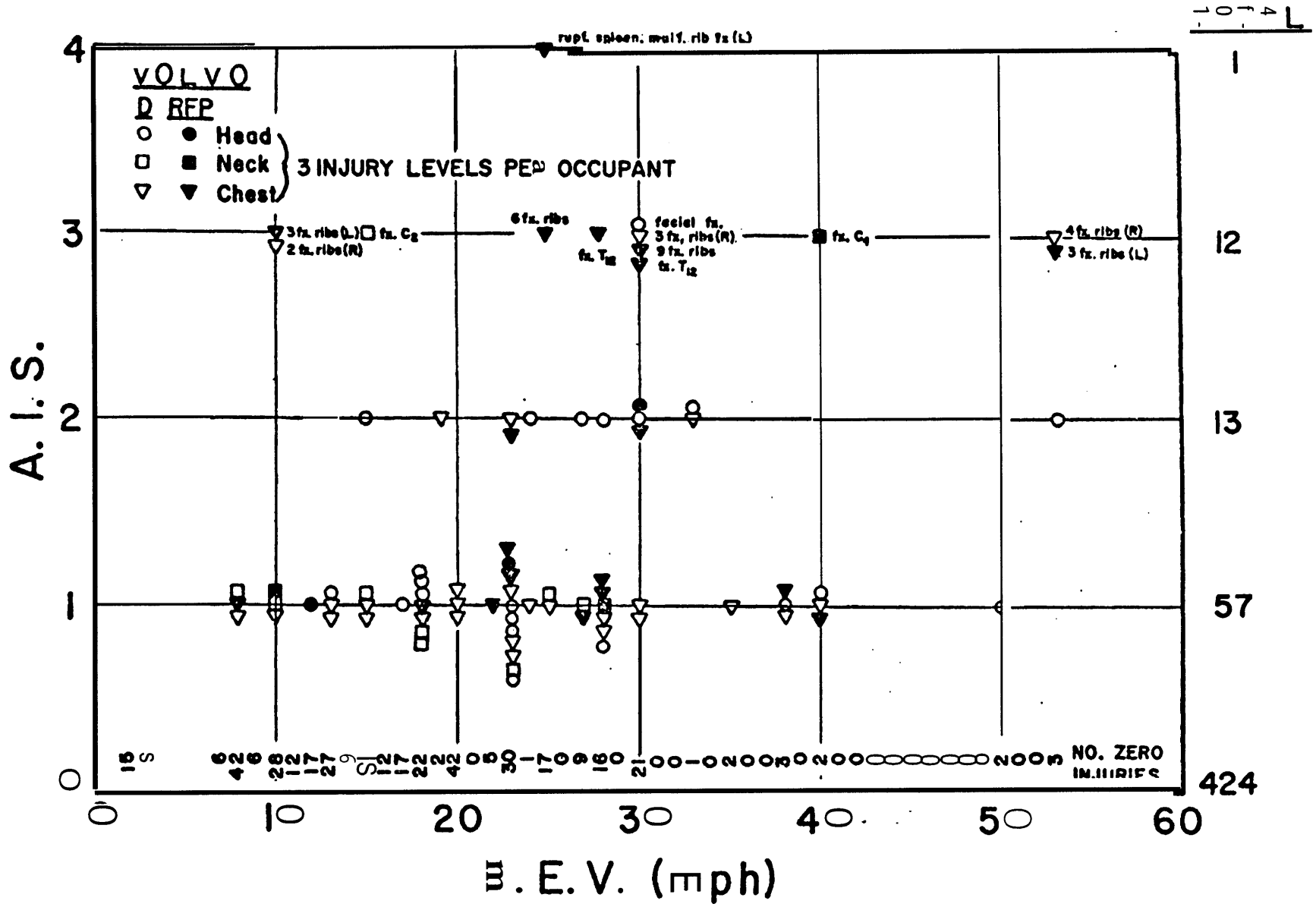



FIGURE 1: ACCIDENT INJURIES IN TERMS OF AIS AS A FUNCTION OF BARRIER EQUIVALENT VELOCITY WITH THE MOST SEVERE INJURY TO HEAD, NECK, AND CHEST INCLUDED.

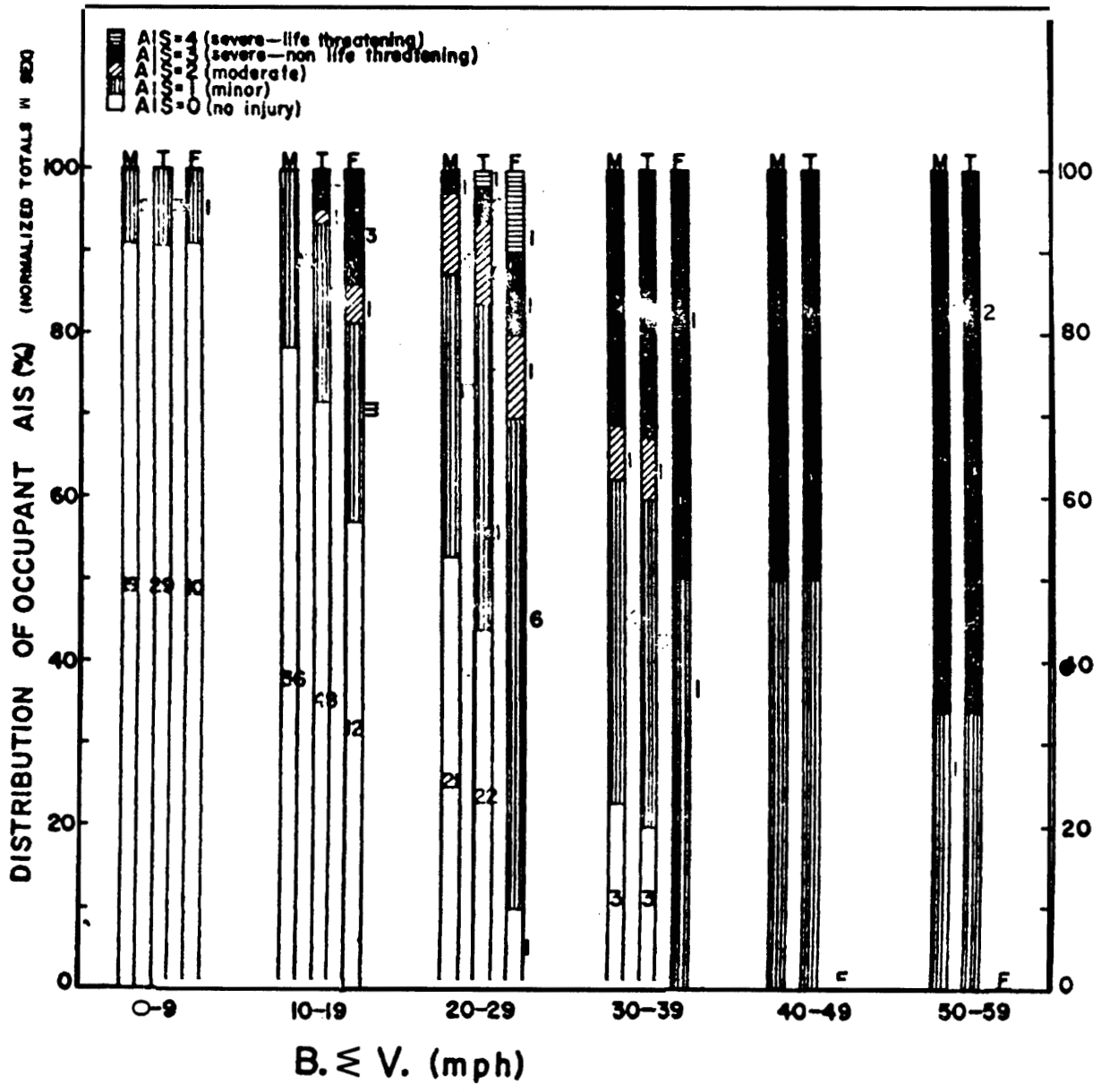



FIGURE 2: DISTRIBUTION OF SINGLE MOST SEVERE OCCUPANT INJURY (AIS=
NORMALIZED BY CATEGORY FOR EACH VELOCITY INCREMENT

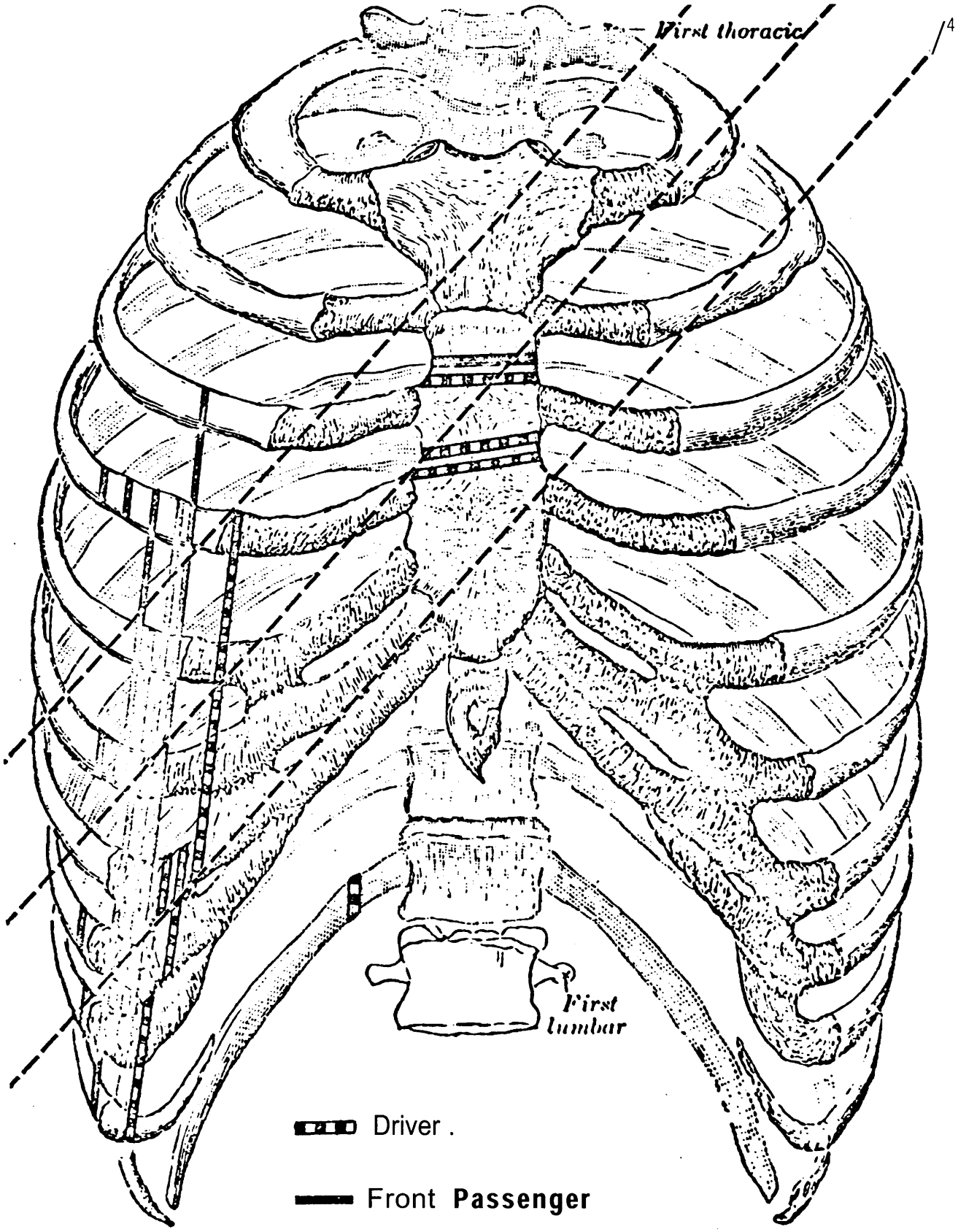
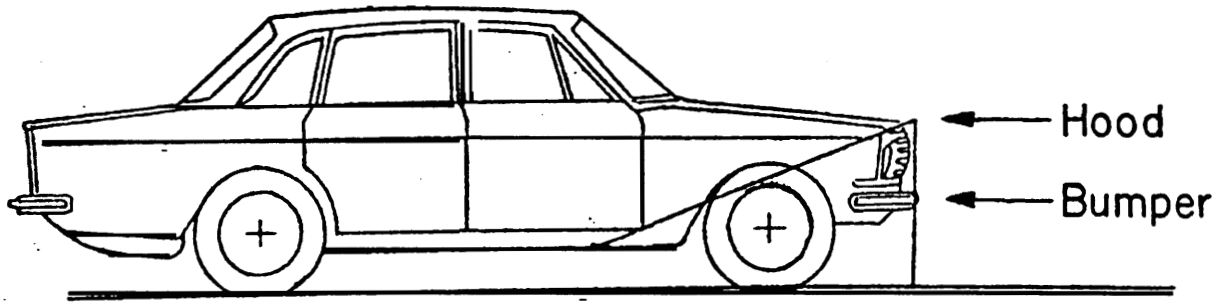



FIG. 3: ALL RIB AND STERNAL FRACTURES (WITH PASSENGER INJURIES TRANSFERRED TO THE DRIVER'S SIDE)

Vertical




Horizontal

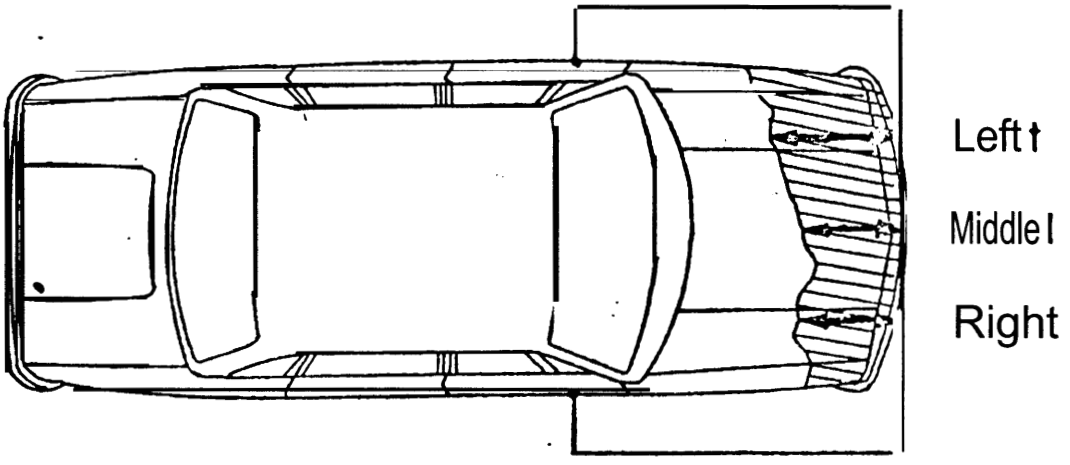



FIGURE 4 : DIAGRAM SHOWING MEASUREMENTS TAKEN WITH FRONT END DEFORMATION FIXTURE.

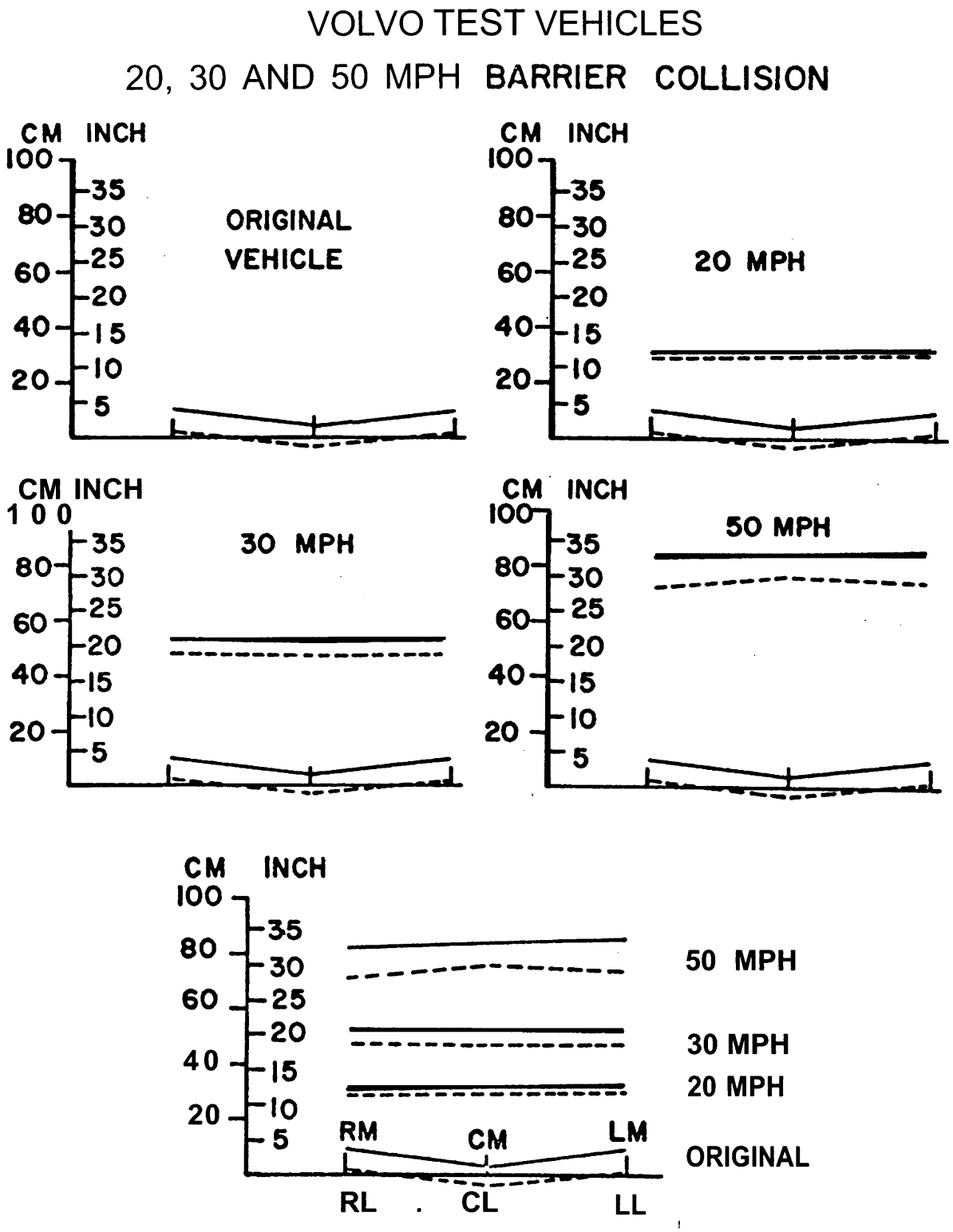
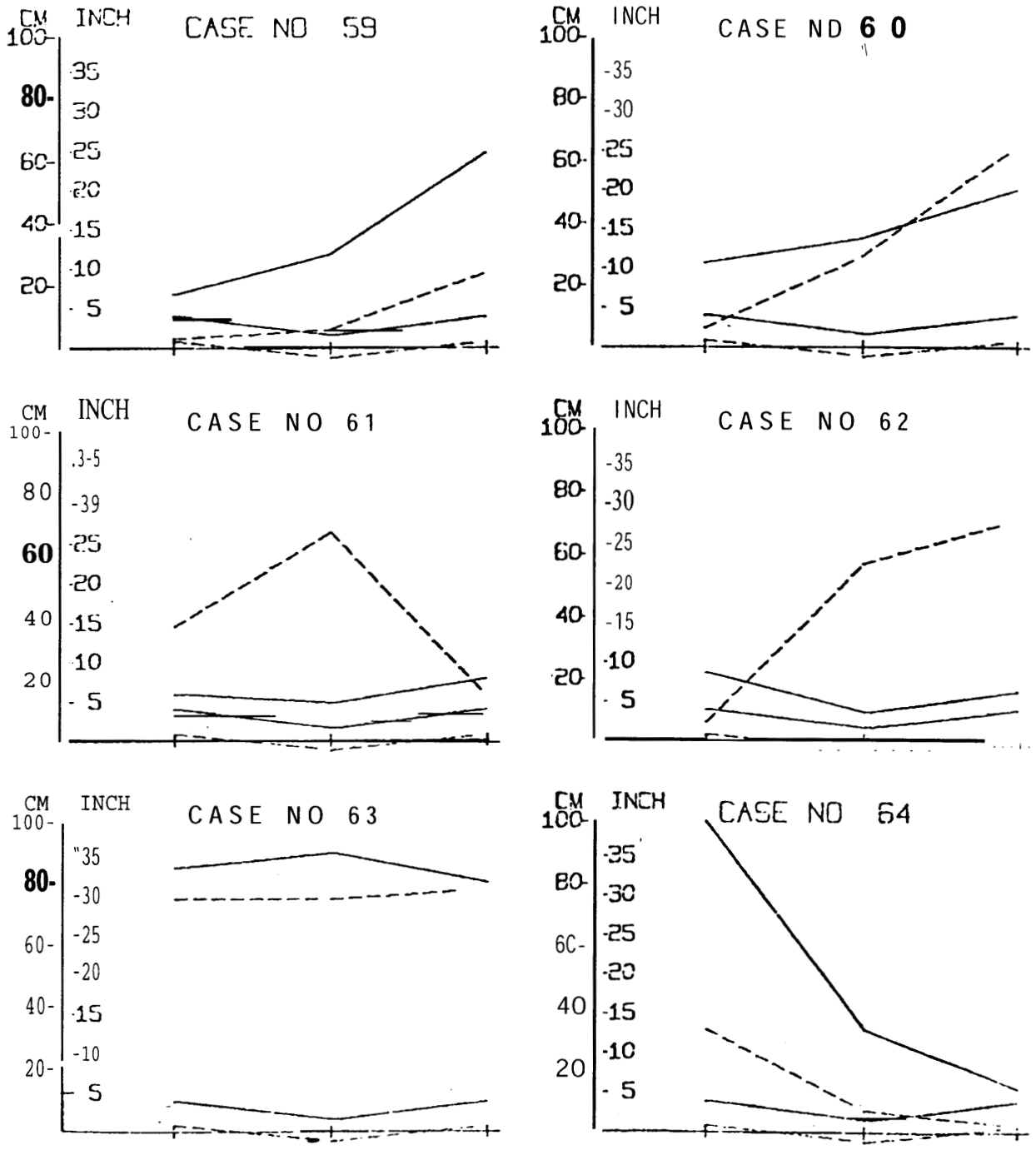



FIGURE 5: HOOD AND BUMPER DEFORMATION FROM STAGED FRONTAL BARRIER COLLISIONS.




'FIGURE 6 : " DEFORMATION DATA FROM EIGHT OF THE ACCIDENT CASES .

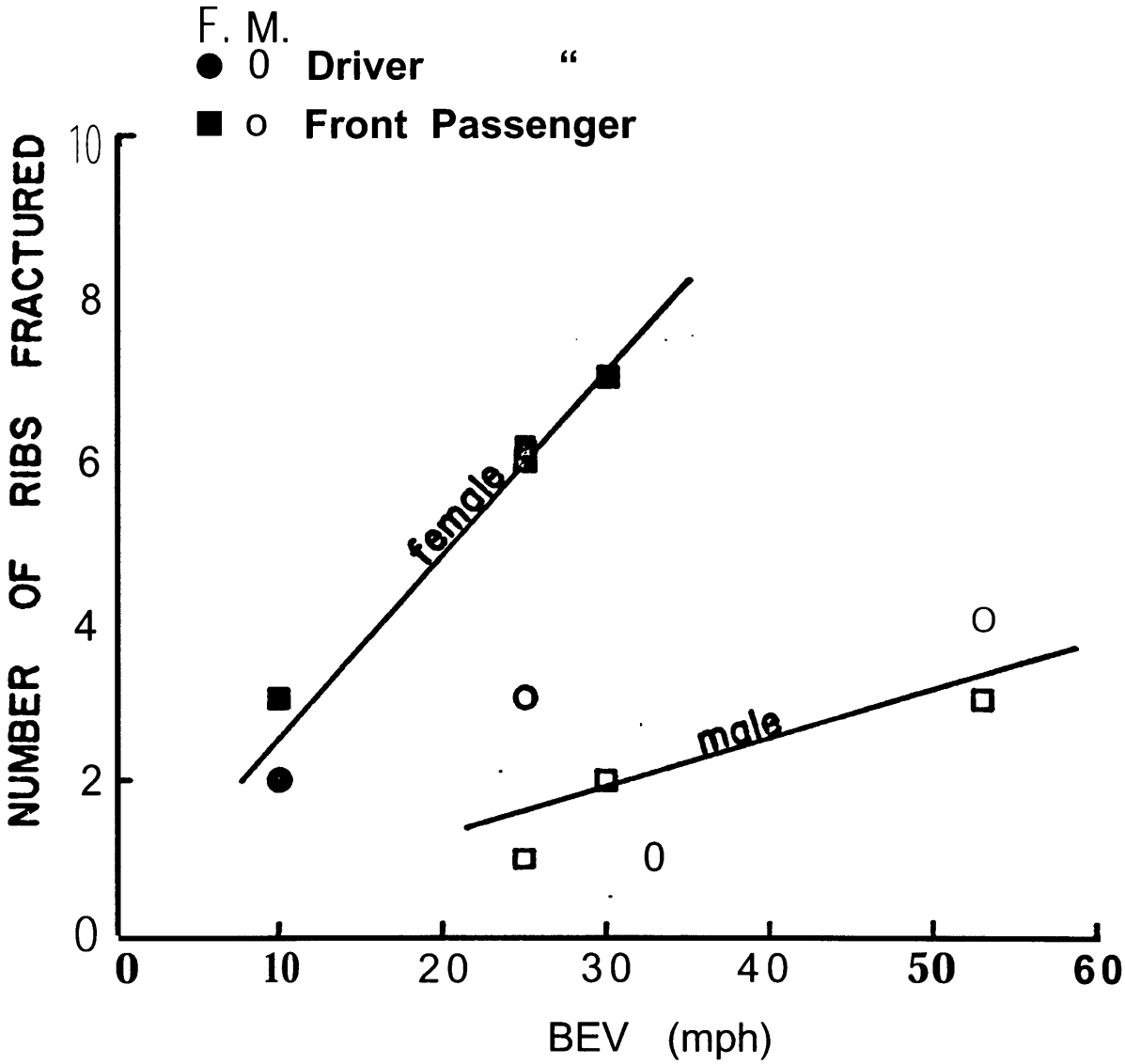



FIGURE 7 : RIB FRACTURES FOR MALE AND FEMALE OCCUPANTS AS A FUNCTION OF BARRIER EQUIVALENT VELOCITY.

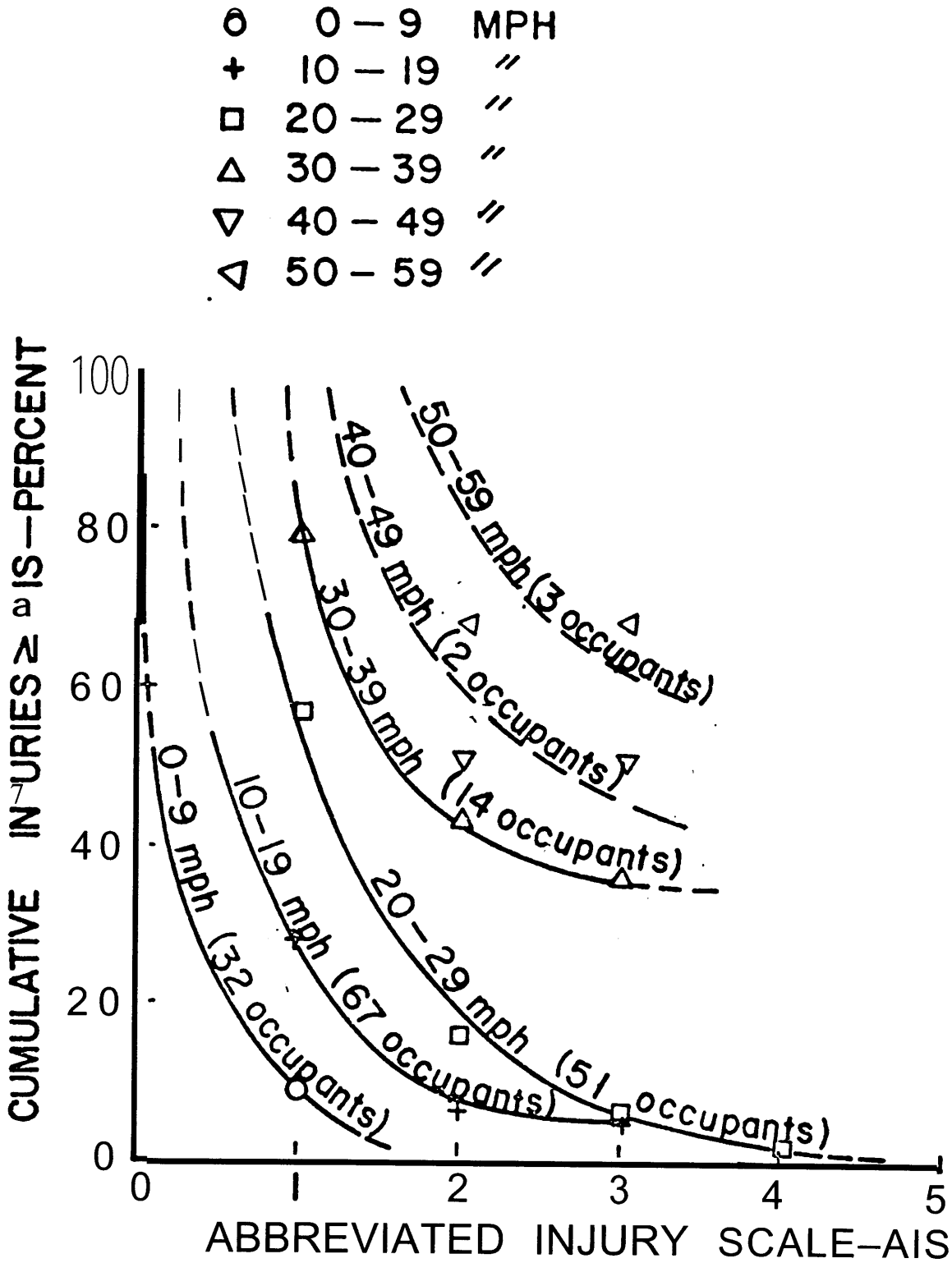



FIGURE 8: PERCENT OF CUMULATIVE INJURIES, EQUAL TO OR GREATER THAN, A GIVEN AIS LEVEL FOR 10 MPH INCREMENTS .

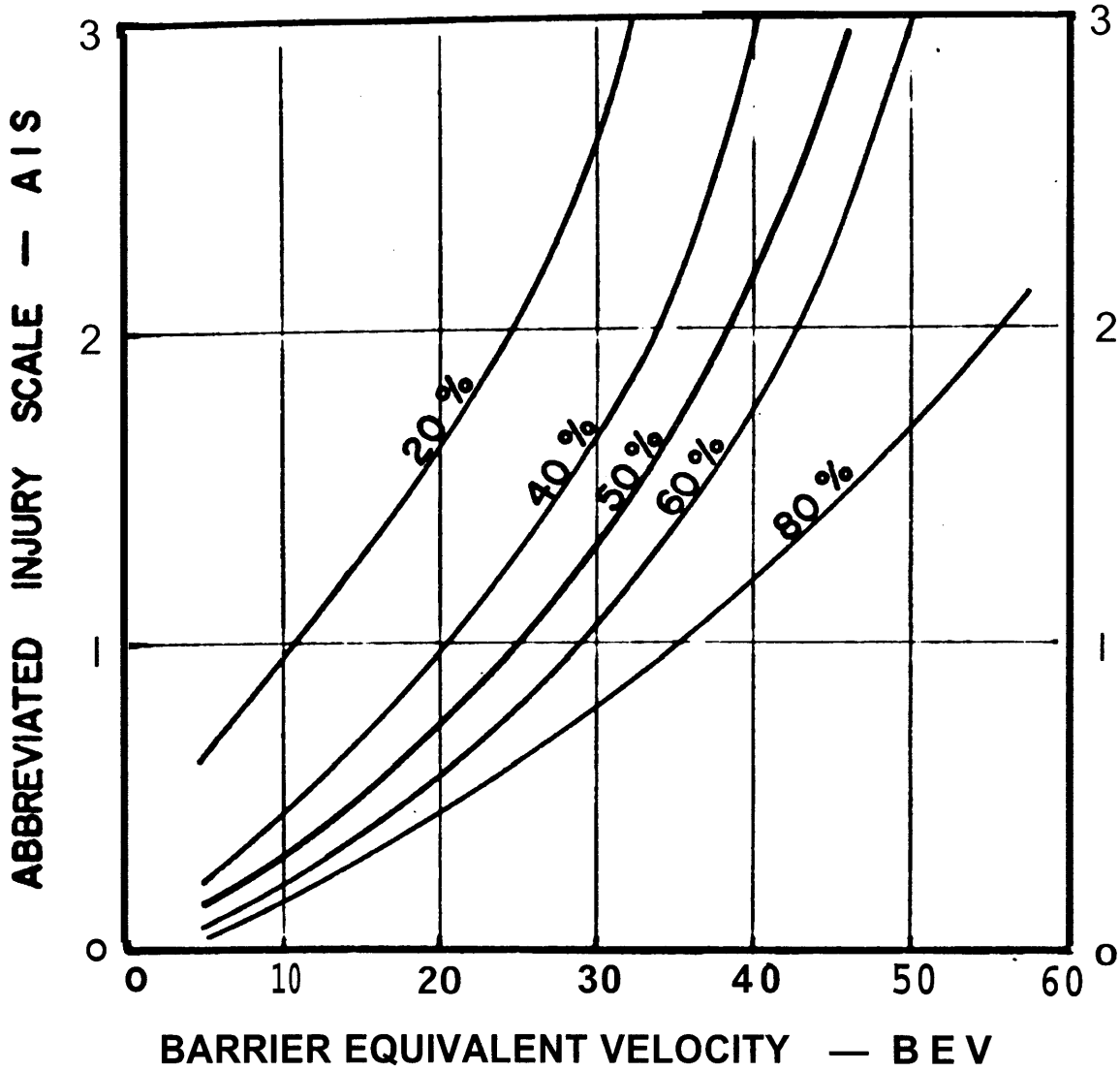



FIGURE 9 : INJURY AS A FUNCTION OF BARRIER EQUIVALENT VELOCITY FOR CUMULATIVE PER CENT OF INJURY

APPENDIX E

LETTER FROM RICHARD WILSON
GENERAL MOTORS CORPORATION

February 4, 1975

General Motors Environmental Activities Staff
Safety Research and Development Laboratory

General Motors Proving Ground, Milford, Michigan 48042

February 4, 1975

Dr. Lawrence A. Goldmuntz
Economics and Science Planning
1200 18th Street, N.W.
Washington, D.C. 20036

Dear Dr. Goldmuntz:

You are to be complimented on your recent Automobile Collision Data Workshop. The free interchange of ideas from such a wide cross-section of data gatherers and data users should be most useful as you formulate your recommendations to the Office of Technology Assessment. I was happy to participate and hope the following comments and the attached material will add to your study.

GM believes there is a need for better accident data so that the true benefits of safety standards can be assessed along with their cost of implementation. This applies to current standards just as well as it does when considering future rulemaking. The value of better data is to improve vehicle safety and to decrease the risk of making an incorrect decision on a standard. The incorrect decision may result in enacting or failing to rescind a standard which is not cost beneficial, or, on the other hand, rescinding or failing to enact a cost beneficial standard. NHTSA should move ahead only with those standards on which they have sufficient information to support a favorable benefit/cost relationship.

You specifically asked for an estimate of the "potential societal cost of not having better accident data". One way to look at this is to consider that the cumulative cost to the consumer for safety standards to date is estimated to be approximately $245 per car (exclusive of bumper provisions). An additional $250 per car are forecast if proposed new safety standards take effect. This $495 per car total related to current and proposed safety standards (bumper standards would be a further addition) translates to about $5 billion per year if applied to production rates of 10 million cars per year. The need for reliable benefit data against which these costs can be evaluated is urgent. Accident data is one source for such information.

Basic requirements for a better accident data system have been presented before. GM has discussed NADS* and the University of Michigan Highway Safety Research Institute has presented SIR**. Other plans may be

* National Accident Data System - Paper by Terry and Schneider given at GM's June 1973 Automotive Safety Engineering Seminar (copy attached).
** National System for Collecting Multipurpose Accident Data - paper by O'Day given at the June 1974 Experimental Safety Vehicle Conference.

forthcoming from your workshop. While exact data system costs have not been formally worked out, they likely are in the area of 10 to 20 million dollars a year. If better accident data could increase the benefit/cost of safety standards by even a few percent (one percent of the above $5 billion would represent $50 million), the $10 to $20 million government investment per year seems very reasonable.

As a specific example, we estimate the cost of continued use of side guard beams, needed to meet MVSS 214, to be about $10 to $12 per car. Applying this cost to 10 million cars per year, this single item of standard represents a total amount to the consumer of $100 to $120 million per year. And yet, the current state of accident data does not even allow a determination of whether side guard beams have had any benefit or not. Again, $10 to $20 million per year for better data seems a minimum expenditure when viewed as a critical ingredient guiding the public's investment of billions of dollars in the costs of their cars.

I hope your project will pull together our country's need in the accident data area. We are convinced there is a need for this type of better decision-making information. I look forward to your final report.

Very truly yours,

RAWilson

R. A. Wilson
Engineer-in-Charge

***RAW/clw***
Attach.

# National Accident Data System

C. Thomas Terry – Section Engineer
Richard W. Schneider – Senior Project Engineer
Safety R & D Laboratory
GM Environmental Activities Staff

Field accident data which reflect what is truly happening m the field today are necessary (1) for the automobile industry to **evaluate performance** and **guide future designs and (2) for the NHTSA to evaluate standards and guide future rule making. This type of data system is not available now. The multilevel system recommended by GM to accomplish this would use the expertise already available in many of the NHSTA-Sponsored multidisciplinary accident investigation teams. The system consists of several study areas which include exposure data and levels 1, 2 and 3 accident data. Another requirement of the system would be a central facility** which **would process the data and make it available to both NHTSA and industry.**

On June 12, 1970, at a Data Accident Investigation workshop* in Brussels, Belgium, GM outlined why field accident data is needed by automobile manufacturers. These needs to collect accident data are:

## Data Needs

1. Evaluate present safety system% ,
2. Predict performance of proposed safety systems.
3. Identify problem areas & evaluate solutions on cost/benefit basis.
4. Estimate human tolerances to impact

### *1. Evaluation of Production Safety Systems*

Early accident investigators saw the results of automobile accidents and identified those vehicle components which were producing frequent and severe types of trauma. This early work supported the introduction of items such as the high penetration resistance (HPR) windshield in 1966 and energy absorbing steering columns in 1967. These investigators were able to measure the relatively large performance improvements of those safety systems. More subtle changes in safety performance can be found only by data collection programs that are refined enough to exhibit statistical trends. For example, it is generally agreed that further changes made to the present windshield will result in a smaller improvement in injury reduction compared to that made in 1966. Measuring this potential change in performance will require a sophisticated accident data collection program.

* R. A. Wilson & C.T. Terry, NATO Accident Investigation Workshop, FIELD ACCIDENT RESEARCH – GM's APPROACH, unpublished presentation, Brussels, Belgium, June 12, 1970

### 2. *Prediction of Proposed Safety Systems*

Before implementing *any* change to safety systems already in the field, the performance of the new safety systems must be predicted. This is the second principle way in which accident data is used.

If the prototype safety system is an improvement on a production item such as the current windshield, then the field data gathered in evaluating the current windshield's performance is used as the injury pattern baseline. The modified system is then tested in the laboratory to compare its performance with the present system. This laboratory comparison provides data to subjectively project how the new windshield might modify the present injury pattern in the field. In this way, the prediction can be made with some confidence as to the performance in the field of the proposed new system.

If a completely new safety system, such as the air cushion restraint system is proposed, the injury patterns which the new system could somehow influence must be identified. In the case of the air cushion restraint, available accident data might be used to identify the injury patterns in frontal collisions where the air cushion is envisioned to be most useful. The air cushion's effectiveness, as determined from laboratory tests, could then be used to predict how the present injury patterns could be modified by the introduction of this new restraint system.

### 3. *Identification of Problem Areas and Evacuation of Proposed Solutions on a Cost/Benefit Basis*

***This*** identification of problem areas requires an over-view of the total injury picture. The over-view consists of the frequency of particular injuries caused by various components and the severities of these injuries. The areas where the most improvement can and should be made are generally where the highest frequency of most severe injuries occur. A relationship between frequency and severity should indicate the areas of high payoff – those areas where the most good can be done. Once these high payoff areas are identified, the priorities of safety development can be established by cost/benefit studies.

As solutions to the more obvious problem areas are incorporated, the identification of the less obvious problem areas becomes more difficult. To identify the less obvious problem areas will require incorporating even more rigorous data collection programs. It may be possible that a point of diminishing returns will be reached. That is, the time and cost of acquiring even more detailed information may not justify the insignificant amount of improvement made from the data derived. To reach this point is a noble goal indeed.

4. ***Estimation of Human** Tolerances to Impact*

The three uses of the field accident data discussed above are specifically aimed at changing the design of the vehicle to reduce the frequency and severity of injuries. A different use of the data is to isolate particular accident situations so that information concerning human tolerances to impact can be generated.

Occasionally, from a large source of accident data, a particular occupant injury in a well-defined automobile accident situation can be attributed to a particular vehicle component. When this infrequent situation arises, and the mechanism of injury is understood, correlation of the accident or "field experiment" with a similar laboratory experiment is attempted. If the "field experiment" can be correlated to the laboratory, the occupant's impact situation might be quantified and the human tolerance to a particular type of trauma can be estimated. For example, an instrument panel may be identified as the cause of a particular type of head injury, A series of similar instrument panels are impacted in the laboratory until the damage to the instrument panel in the accident case is reproduced. The forces and accelerations to produce the damage in the laboratory are then correlated to the injury produced in the field, In this way, the human tolerance is quantified for this particular type of injury.

These needs remain as valid today as they did three years ago. Further mentioned were the qualities of a good field accident data system:

## Data Qualities

1. Rapid feedback
2. Random data sample
3. Current model data
4. Data compatibility

*1. Rapid Feedback*

A prime goal in automotive safety is the reduction of injuries and deaths due to automobile accidents. The more injuries prevented and lives saved, the better the job is done. improved safety systems must be incorporated as rapidly as practicable to achieve this goal. An orderly implementation of improved safety systems depends in large measure on the collection and assessment of field accident data. Only after a sufficient amount of statistical and in-depth data is collected can problem areas be identified and further improvements be recommended and *implemented.*

***2. Random Data Sample***

Besides the quantity of data gathered, a random sample is essential to insure its quality. Basically, random data is needed so that conclusions aren't erroneously based on the consequences of a unique accident, or limited number of accidents. False accident and injury patterns can be created by generalizing from a small sample of non-random cases. In the past, most sources of accident data have not been random. Most accident investigations typically have been biased by geography, injury level, damage level, or other accident selection techniques. A valid data sample must be representative of the real world.

***3. Current Model Vehicles***

Each year safer automobiles are produced. Measuring these advances in safety performance from one year to the next requires a valid data baseline. it should be realized that resources are limited and it would be virtually impossible to collect enough data on the total vehicle population in one year. The most efficient use of resources is to concentrate investigation on the most useful data source - current model vehicles. Of course, as current model data is collected each year, in time, a data bank will be built which will allow a comparison of newer automobiles with trends based on many years.

***4. Compatibility of Format***

If various data sources are ever to be combined to form large data banks, they must, at least, be in the same basic format. This means that the same information is recorded for each accident and some means of easily combining information from different sources is provided. This is particularly important when in-depth data is being collected because of its inherent complexity.

However, even when it is physically possible to combine data from various sources, it is not always advisable. Each investigator tends to bias his accident selection in some manner such as injury only, rural only, etc. Since the data base for each investigator is usually different, a direct statistical comparison of their data is not advisable.

Again, these characteristics are still desirable today as they were three years ago. There is no known source today which satisfies all of these qualities. The one key quality which bears emphasis is the random data sample. The random data sample criteria implies that the accident cases selected are representative of the national accident experience. This representativeness is critical for sound decision making regarding automobile design and government rule making.

Making decisions with national implications in highway safety using only data from rollover accidents in North Carolina is no more valid than predicting the Gross National Product from monitoring only the construction industry in Utah.

### Current Data Status

In the three years since that NATO workshop, some other factors have become obvious regarding the value of accident investigation.

1. The information received not only can be used by the industry for evaluation and direction, but also can apply to Government at all levels for rule making.
2. Variation in the interpretation of current accident data results from two factors:
   a. Different analysis techniques
   b. Different data sources

Variation of results due to the first cause i.e., different analysis techniques, is healthy and promotes various problem solving strategies to be explored and compared. However, differences due to the second source are generally inefficient and result in problems of interpretation. This problem will remain unsolved until the many various data collection efforts are coordinated so that their results can be combined. This combination into a representative data set will then allow, the safety experts to base decisions on a sound technical basis.

These previously stated needs and system characteristics coupled with the conflicting conclusions which result from the uncoordinated data collection activities around the country have led GM to propose what is called a National Accident Data System.

Before outlining the proposal for such a system, one point should be stressed: the system being proposed is not the best system that theoretically could be designed. In fact, it is several steps away from being an optimum design. But it is also many steps closer to an optimum system than anything that exists today. Rather than wait for that perfect system to be implemented, it is imperative that the obvious contradictory nature of various data sources be eliminated now so that valid cost/benefit studies can be used in achieving the goal of reducing injury and death on the highway. Each change made to the system after it is begun should be directed toward the desired optimal system.

The proposal itself tries to incorporate many of the data collection activities that are now in existence while eliminating other unnecessary ones. But the design is primarily dictated by the desire to establish a coordinated National Accident Data System in a relatively short period of time.

## DATA COLLECTION

The proposed system involves designating certain geographic regions of the country as sample areas where extensive surveying and profiling will be conducted. This is analogous to taking a Gallup Poll of the nationwide accident experience. Since many of the existing Multidisciplinary Accident Investigation (MDAI) teams sponsored by the NHTSA are somewhat randomly located and because expertise already is available from the teams, we are proposing that selected MDAI teams would form the nucleus for the data collection system. This proposal would convert existing MDAI teams into multi-level programs such that each team has the responsibility of coordinating the gathering of the following levels of information within their specific regions:

1. Exposure data (non-accident)
2. Level 1 accident data
3. Level 11 accident data
4. Special accident studies

***Teams which could not reliably supply all these*** levels ***of information would not be included in this program.***

### Exposure Data

Exposure data is profile information on the number and types of people, vehicles and roads in the area. this information is used to define the universe in which the accidents are recorded. Ideally, when all the regions are combined, the exposure should be "representative" of the total United States. Capturing data of this nature allows the various combinations of vehicles/drivers/roads to be described whether in an accident or not. Most of this information is available in existing state operational files. The system should allow specific surveys of additional data to also be conducted. For example, it may be necessary to establish how many miles various age groups drive annually.

### Level I Accident Data

This level requires collecting a standard police report on ***all*** accidents in the region which meet a predetermined severity threshold. An alternate to the standard form would be a form with a common core of information with other elements decided upon by the local jurisdictions. This level of information briefly defines the nature of all accidents in the area. This information, coupled with the exposure data, make possible the computation of accident rates, such as

fatalities/miles driven, accidents/make and model, or accident/driver age. Since the accidents described in this file contain both injury and no-injury cases, computing the probability of an injury occurring is also possible. Definitions or specifications of variables within each region and from region to region ***must*** be consistent. ***This standardization of definitions between regions is imperative, and will provide the program with one of its greatest challenges and one of its greatest advantages over current programs*** Emphasis upon the training of the police investigation people is important for this level of data. Definition of what an accident **is or of what the** various injury levels are must be explicitly stated and uniformly interpreted. Again, flexibility should allow specific, supplemental information to be collected when needed. As an example, the police could be asked to ascertain whether the head restraint was in the "up" or "down" position in a rear end accident.

**Level II Accident Data**

**This** level of data would collect information on all accidents in the region which involve a recent model vehicle and an injury. Information on all vehicles involved in the accident would be required. The injury may in fact occur in an older vehicle which impacted the recent model vehicle. This level data **has been most valuable from** the manufacturer's viewpoint and has historically been the source of injury causation information. Extending the coverage to include older vehicles would allow comparison of vehicles of different ages. in the past, information of this type collected by GM and other ***has led*** to improved vehicle design, examples being HPR windshields and the energy absorbing steering assembly. The information gathered would define the ***injury severity,*** the ***causes of the injury, the accident description,*** a ***measure of its severity,*** and some information relative to the ***cause of the accident.*** This information will allow the assessment of new safety systems as they are introduced such as air cushion restraint or starter-interlock webbing systems. Gathering the data on all accident modes and injuries will allow relevant safety evaluation tests to be specified. By combining this data with the Level I Accident Data, it may be possible to evaluate the relative safety performance of various makes and models of vehicles. The current thinking is that the information would be gathered on a modified version of the GM Field Form by investigators working for the MDAI teams. As with the present Field Form, a series of photographs will be required to supplement the information. The form would be expanded to collect information on pre-crash and post-crash phases of the accident which are not presently addressed on our existing form. This part of the system would also allow extra information to be collected on items of specific interest which are not in the GM Field Form. For example, the investigators may be asked to see if the starter interlock system has been defeated or if it had any effect on the occupant's usage.

Level Ill Accident Data

These special studies are performed to see why **particular** problem **areas exist. The special studies conducted are based on the Level I or Level [1 information already gathered. For example, a special investigation could be undertaken to more closely examine** why a particular class of vehicles for **"over-represented" in a particular type of accident.** The investigation may find that this type of vehicle is popular for owner modification which could result in unstable handling characteristics.

**DATA COMPILATION**

**The next logical question is what to** do with the data after it is collected in its relatively rough form i.e., police reports, GM **Field Form, and photographs.** To keep the interpretation of **raw data** consistent from **area to** area, it is proposed that the data be collected in a central location. At this location, the Level I data would be entered directly into a data bank. The information from the detailed Field **Form and photographs in** the Level II system would be analyzed and the final information entered into an automated data system. By centralizing this function, the number of subjective judgments are made more or less to be consistent because of the relatively few number of people involved. This situation is similar to that which is now used with the General Motors-MIC program, and has been found to be quite satisfactory. We feel the overall quality of data will be enhanced by increasing the consistency of the data. This central facility **would not only provide** common data entry and storage facilities, but would also offer a retrieval system for interested data users.

## PROGRAM IMPLEMENTATION

**Since this program should benefit the industry as well as the Government, it** is recommended that **joint Government/industry support** for the implementation and annual operation of this program be solicited. The industry support could logically be under the auspices of either MVMA or SAE. Specifically, it is felt that the program offers a great opportunity for joint efforts between Government and the industry toward achieving a common goal. There **are** actions required of both industry and **government to implement the proposed program. The program is a national** goal **and therefore should be funded with Federal monies. However, the**

industry should be willing to participate in initiating the program and continue support to the end that the data will be valid and available.

After this program is initiated, data acquisition could begin in less than a year. As shown in Figure 1.

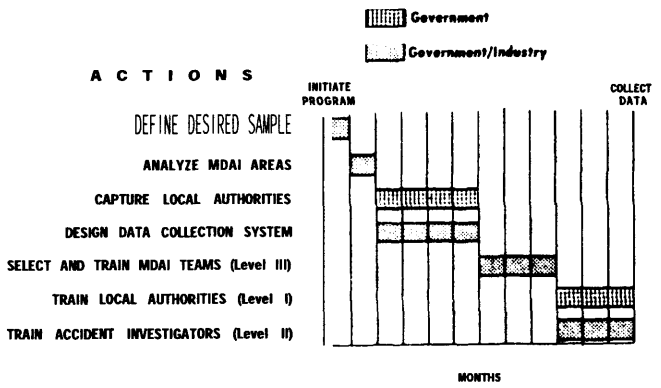



Figure 1

## SUMMARY

Although this system is not a new idea, it is the basic simplicity which is most appealing. The program has been outlined in general terms only, although it has been given much more detailed thought as this general outline was developed. Rather than explore the details at this time, support is being solicited for the overall plan of action in the hope of gaining cooperation from other groups in the detailed planning phases of the program. Again, the payoff from such a system would be high, and achievable in a relatively short period of time.

It is GM's intent to act as a catalyst in the design and implementation of a National Accident Data System and encourage any of you today to accept this challenge with us.

C. Thomas Terry

*C. Thomas Terry is a Section Engineer responsible for the Field Accident Research activity at the safety Research and Development Laboratory located at the GM Proving Ground. His responsibilities include the collection and analysis of field accident data*

*He received a Bachelor of Science Degree in Civil Engineering from Rose Polytechnic Institute, Terre Haute, Indiana, and a Master of Science in Engineering Mechanics from Wayne State University, Detroit, Michigan. Mr. Terry joined General Motors in 1969 and was assigned to the biomechanics area with responsibilities in human simulation and volunteer testing. He was chairman of the SAE Crash Test Dummy Subcommittee during this time.*

*Mr. Terry then joined the Field Accident Research group in 1970 and was promoted to his present position in 1971. In 1972 he assumed the role of Chairman of the Motor Vehicle Manufacturers Data Collection Co-ordinating Subcommittee.*

*Among his publications are:*

*"Radiological Studies of Organ Displacement Due to Vertical Accelerations" presented at the 18th Annual Conference of Engineering in Medicine and Biology, November 1965, Philadelphia, Pennsylvania.*

*"Review of Mathematical Models of Response to Acceleration," presented at the Winter Annual Meeting of the American Society of Mechanical Engineers, November 1966, New York, New York.*

*"A viscoelastic Model of the Human Spine Subjected to +$g_z$ Accelerations," Journal of Biomechanics, Vol. 1, pp 161-168, Pergamon Press.*

*"Field Accident Research–GM's Approach," R. A. Wilson, C. T. Terry, presented at NATO Accident Investigation Workshop, Brussels, Belgium, June 12, 1970.*

*"Benefits of the In-Depth Case Study," presented at 1972 Annual Meeting of Society of Automotive Engineers, January 10-14, 1972.*

*"National Accident Data System," C. T. Terry, R. W. Schneider, GM Automotive Safety Seminar, June 2&21, 1973.*

Richard W. Schneider

*Richard W. Schneider graduated from Grinnell College where he received the degree of B.A. in 1969 and a Masters Degree of Business Administration in 1971. He joined General Motors Proving Ground in 1971 where he was involved with field accident research. Mr. Schneider is currently senior project engineer with the Safety Research and Development Laboratory at the Proving Ground and active in the area of field accident research. He is a member of Operations Research Society of America.*

APPENDIX F

# STATISTICAL RATIONALE FOR THE NUMBER OF AUTOMOBILE CRASH RECORDERS PROPOSED FOR PROCUREMENT AND INSTALLATION BY NHTSA

National Highway Traffic Safety Administration

February 5, 1975
(date of receipt)

# STATISTICAL RATIONALE FOR THE NUMBER OF AUTOMOBILE CRASH RECORDERS PROPOSED FOR PROCUREMENT AND INSTALLATION BY NHTSA

National Highway Traffic Safety Administration
(Received February 5, 1975)

The statistical justification for the number of crash recorders requested by NHTSA depends primarily on the answers to two general questions.

A. If N crash recorders are installed in passenger cars, what number of crashes will be recorded annually in each category, or "cell" of interest? For example, how many frontal impacts with impact speed ($\Delta v$) 30 mph or more will be recorded by the crash recorders?

B. Given the answers to question A, will these numbers provide adequate information on the crash environment. This involves statements about the precision and accuracy of various estimates of rates, proportions or distributions, such as confidence limits or error standard deviations.

Figure 1 summarizes much of the basic factual information needed to answer question A. The figure shows numbers of crashes of various types that would be expected in 1 year from a crash recorder fleet of 100,000 vehicles. The numbers are derived from NHTSA's experience with the current restraint systems study and other accident studies. The estimated recovery rate for crash recorders in accidents is 64 percent; this is a judgment factor on which there are few relevant data.

Figure 1 shows an initial fleet of 100,000 recorders, and it is easy to modify the figure to obtain two other useful sets of figures. If one adds three zeros to each number in the figure, the resulting numbers are estimates of the numbers of crashes occurring with the entire U.S.-automobile population in 1 year. If the numbers are each divided by 1,000, the result is the percent in each category. For example, we can see that about 1.6 percent of the vehicles each year will be involved in towaway crashes from which the recorder is recovered.

Figure 2 illustrates the problem of estimating the cumulative distribution of crash speeds. ("Speed" may refer to any measured value such as $\Delta v$, barrier equivalent velocity (BEV), traveling speed, etc.). The figure shows a "true" distribution function, represented by the solid curve, and an empirical distribution, obtained through the recorder, and represented by the stepped graph. The maximum vertical distance D between the two curves is a random variable. As the number of observations increases, the probability that D will exceed any specified value decreases; i.e., the empirical distribution function approaches the true population distribution function. The following table shows the numbers of observations needed to obtain 80. and 90-percent confidence that the maximum deviation between true and empirical distribution functions does not exceed a specified value.

Figure 1 - Yield from 100,000 Crash Recorders in 1 Year

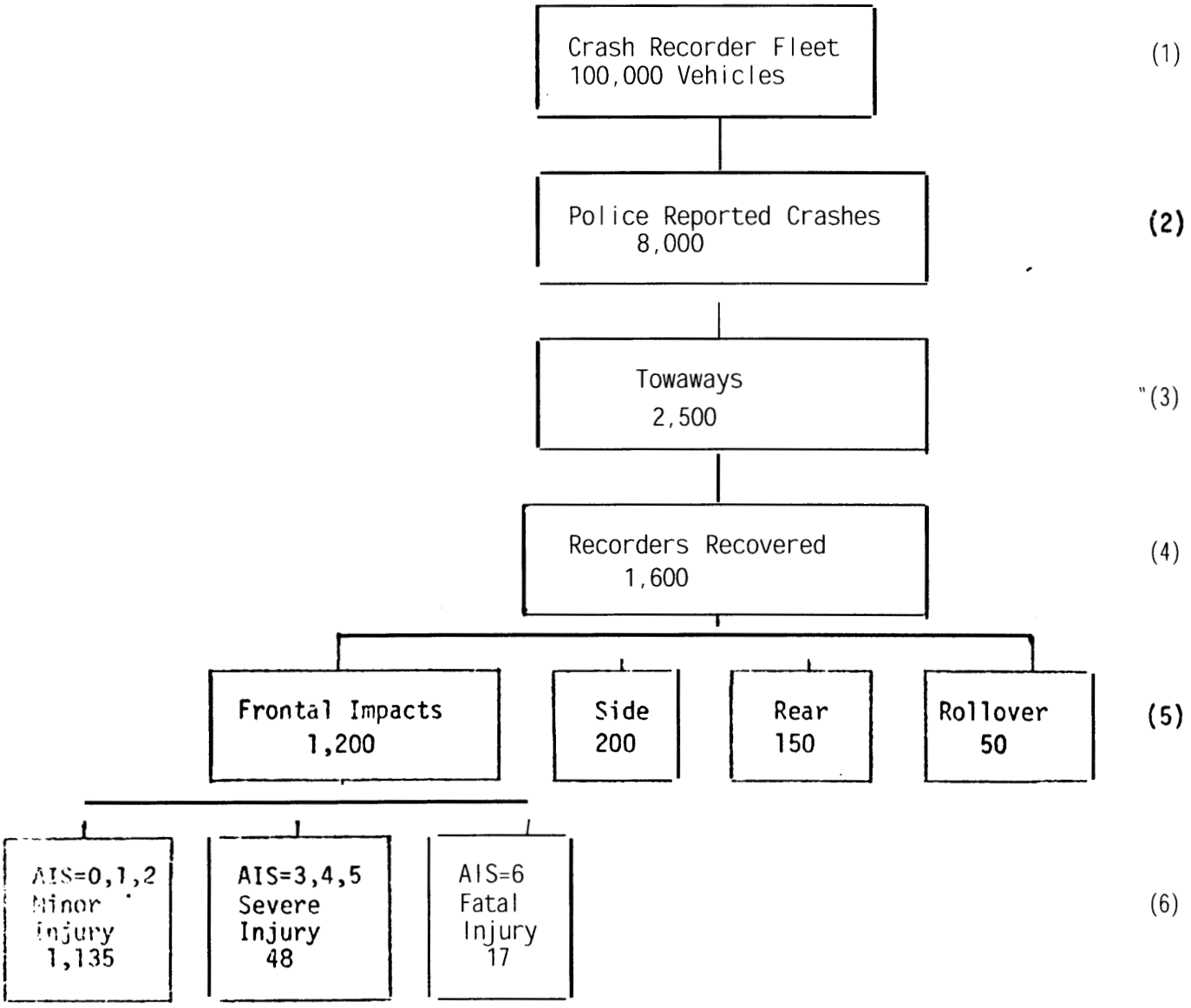


Figure 2 Cumulative Distribution of Crash **Speeds**

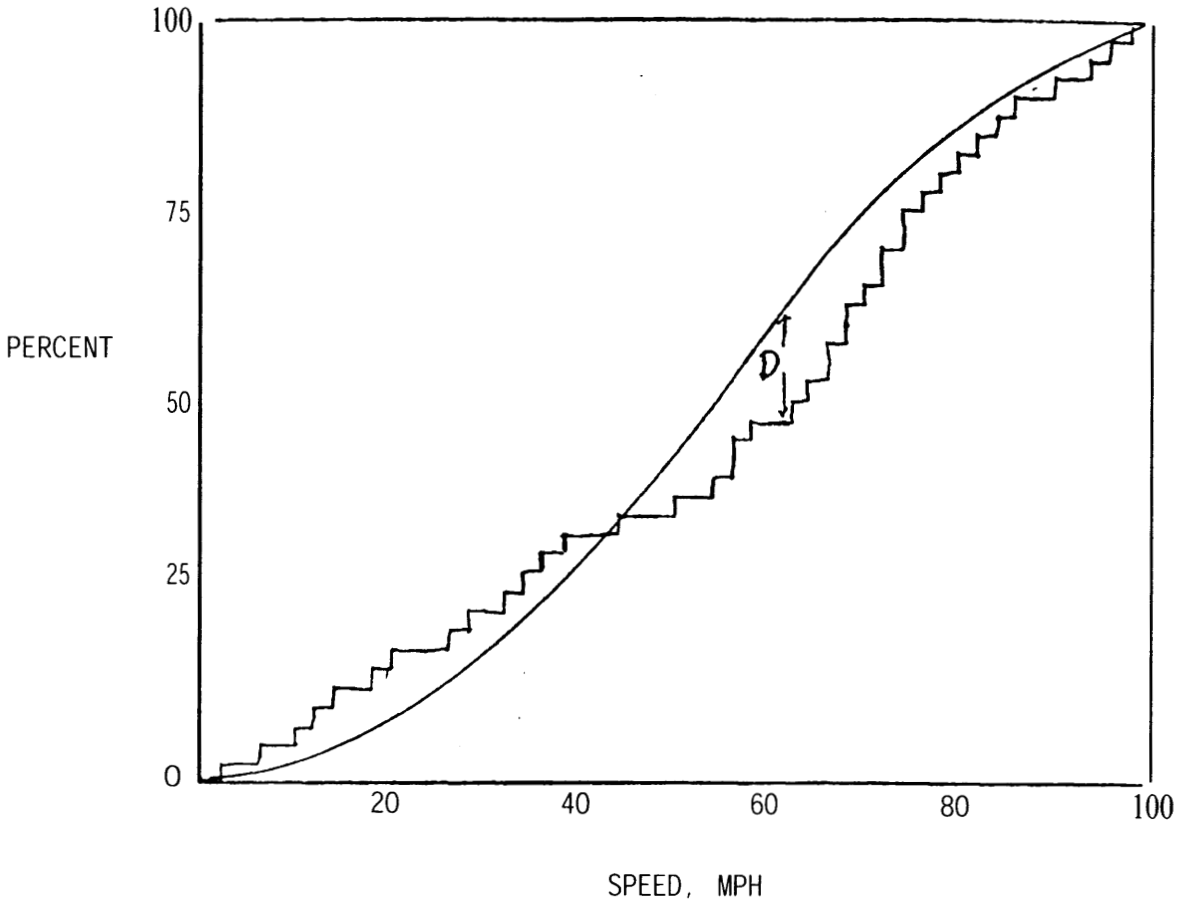


Table 1. Number of Observations Required for Specified Confidence That Maximum Deviation Between Empirical and Hypothetical Distribution Does Not Exceed . Value Shown

| Maximum Deviation | Confidence Level | |
|---|---|---|
| | 80 Percent | 90 Percent |
| .01 | 11,449 | 14,884 |
| .02 | 2,862 | 3,721 |
| .03 | 1,272 | 1,653 |
| .04 | 716 | 931 |
| .05 | 458 | 595 |
| .08 | 179 | 233 |
| .10 | 115 | 150 |

In estimating the fraction of the crashes that fall into a category of interest (e.g., impact speed over 30 mph), we are concerned with the variability of an observed proportion f in a sample from a population in which the "true" proportion is p. In large samples (> 25) the observed fraction is distributed normally with mean p and standard deviation:

$$\sigma = \left[\frac{p(1-p)}{n}\right]^{1/2} = \left[\frac{pq}{n}\right]^{1/2}$$

where n is the sample size. The greatest variability occurs when p = .5, in which case the formula reduces to

$$\sigma = \frac{1}{2\sqrt{n}}$$

So if we specify a probability (confidence level) that the observed results shall not deviate by more than D from the population proportion p, the required sample size can be estimated. Table 2 shows maximum sample sizes required at two confidence levels.

Table 2. Sample Size Required to Estimate a Proportion with Error Less Than D

| Maximum Deviation D From Population Proportion | Confidence Level | |
|---|---|---|
| | 80 Percent | 90, Percent |
| ,01. | .4,107 | 6,767 |
| .02 | 1,027 | 1,692 |
| .03 | 456 | 752 |
| .04 | 207 | 423 |
| .05 | 164 | 231 |
| .08 | 84 | 106 |
| .10 | 41 | 58 |

The preceding material will now be applied to let us reach some conclusions on how many crash recorders NHTSA should purchase and install.

1. To estimate the proportion of fatal crashes at barrier equivalent velocity below a stated speed, close to a million recorders would be needed. From line (6) of Figure 1, we see that these would yield 170 frontal impact fatalities in a year and 510 in 3 years. This would permit us to state, for example with 80-per cent confidence, "the percent of fatalities in frontal impacts in which BEV exceeds a stated speed is x ±3 percent" after 3 years of data collection with 1,000,000 recorders. For deaths in crashes other than frontal, the requirements range from at least six times as great for side crashes to at most 24 times as great (i.e., 24,000,000) for rollover crashes. The costs to determine any of these fatality distributions directly with the crash recorder appear to be prohibitive.

if we use the injury criterion of either fatal or severe injury (AIS > 3), (see line 6, Figure 1) the required numbers reduce by a factor of approximately 4, but are still very high.

2. A *more* limited goal is to determine the distribution of barrier equivalent speeds in crashes by impact type. This information is an essential input for crashworthiness design. In this case, the distribution of BEV's for <u>frontal</u> crashes can be determined quite well in a year to about ±.03 with 100,000 recorders> The error in estimating a single proportion (for example, the fraction of BEV under 30 mph) will be less

than .02 with 80 percent confidence and less than .03 with 90-percent confidence. For side and rear impacts, the BEV distribution can be estimated within ±.80 with 85-percent confidence.

3. Table 1 shows that to reach 80-percent confidence that the distribution of impact speeds observed with a crash recorder is within ±.03 of the "true" distribution function of observed population of crashes, it is necessary to record 1,272 crashes.

The number of recorders needed to be sure of 1,272 recordings depends upon the frequency of the crash type that is of interest. The following table shows the number needed for several crash types of interest. These numbers assure us at the 80-percent confidence level that the maximum error does not exceed ±.03.

| Impact Direction | Severity Level | 1 Year | 3 Years |
|---|---|---|---|
| Frontal | Fatal | 7,490,000 | 2,500,000 |
| Frontal | AIS > 3 | 1,960,000 | 653,000 |
| Frontal | Towaway | 106,000 | 35,000 |
| Side | Towaway | 636,000 | 212,000 |
| Rear | Towaway | 849,000 | 283,000 |
| Rollover | Towaway | 2,546,000 | 852,000 |

4. Another goal of the crash recorder program is to "calibrate" other measures of crash severity. Some cheaper, less accurate, even biased measurements may become very useful if their biases are consistent and if we can estimate their error distributions. For example, we might use vehicle deformation more readily if we know how

to associate a speed with each point on the vehicle damage scale and could determine the expected errors. The situation is analogous to using a ruler that is 1 inch too long. If we knew the "true" values corresponding to the erroneous ones given by the ruler, we would be able to use the ruler and make corrections.

To accomplish this calibration it would be necessary to consider separately vehicles whose deformation characteristics differ substantially. A minimum of four groups would be required, corresponding to various classes of vehicles. Additionally, it is necessary to consider the type of object struck: soft or hard, concentrated or distributed. . Finally, the calibration needs to be done for at least five points on a speed curve, preferably more. There could be a requirement for up to 80 groups of observations or cells (4x4x5).

With a fleet of 100,000 crash recorders, NHTSA could obtain 1,200 frontal crash impact recordings in a year, which is an average of 1,200 ÷ 80 = 15 per group, and many groups would have much less than 15 observations. Over a period of 3 years the average group size would reach 45. If one assumes a 5 mph standard deviation for the inaccurate measurements, then with 15 measurements the mean for each measured point on a speed curve will be determined with 90-percent confidence to within 1.3 mph. For a 10 mph standard deviation in the measurements to be calibrated, the 90-percent limits will lie 12.5 mph from the mean.

Conclusions:

1. Installing 100,000 recorders would permit a reasonably accurate determination of impact speeds for frontal towaway crashes in a year's time. Less accurate determinations of side and rear crash speed distributions for towaways would be available by the end of 3 years. These statements rest on the assumptions that:

(a) The energy crisis and 55 mph speed limit will not reduce the rate of crashes drastically.

(b) NHTSA can find a way to get a representative sample of crashes.

2. With 100,000 recorders, it will be possible to "calibrate" the various proxy measures used by accident investigators with an acceptable degree of accuracy.

3. The recorder program does help provide a basis for rulemaking. The NHTSA rulemaking organization was quite clear in the requirement for data which only recorders can provide. Attached are 4 charts which state the application of recorder data. The standard writers have consistently provided positive support to the recorder program because of the additional dimensions they provide the technical data base upon which standards are based.

U.S. DOT/NHTSA

# AUTOMOTIVE RECORDER RESEARCH PROGRAM

BASIS FOR RULEMAKING

RECORDER: MEASUREMENT OF VEHICLE CRASH SEVERITY

PLUS: ACTUAL DATA ON INJURIES AND FATALITIES

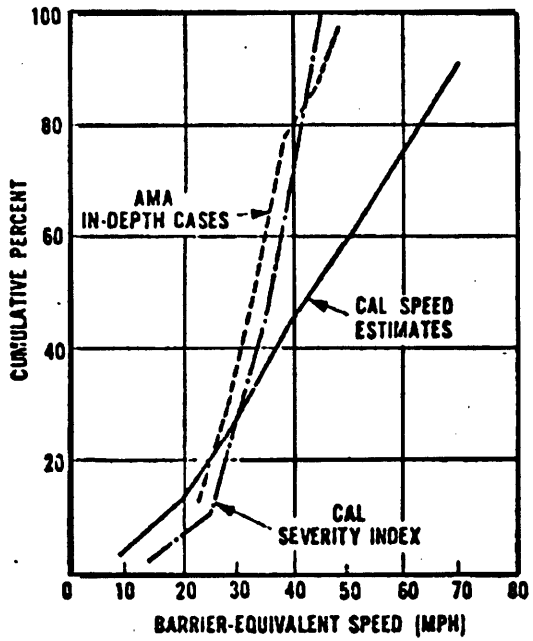
EQUALS: FIRM UNBIASED RELATIONSHIP OF CRASH SEVERITY WITH OCCUPANT INJURIES AND FATALITIES

U.S. DOT/NHTSA

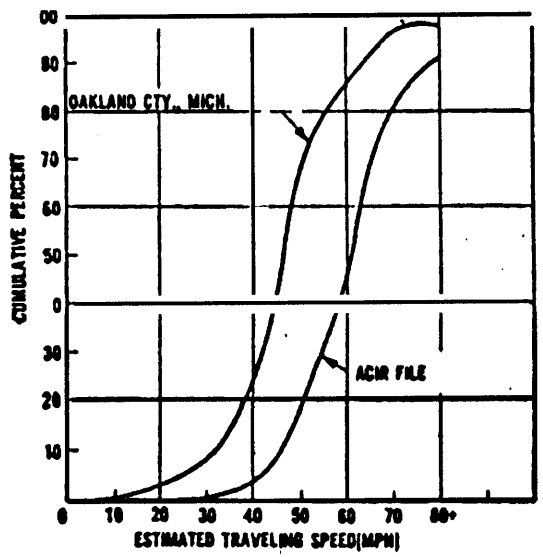
# AUTOMOTIVE RECORDER RESEARCH PROGRAM

## CURRENT RELATIONSHIPS

PERCENTAGE DISTRIBUTION OF BARRIER–EQUIVALENT SPEEDS FOR SINGLE COLLISION FRONTAL FATALITY VEHICLES BY SOURCE OF DATA




CUMULATIVE FREQUENCY DISTRIBUTION OF SPEED OF FASTER VEHICLE IN FATAL MULTIVEHICLE CRASHES

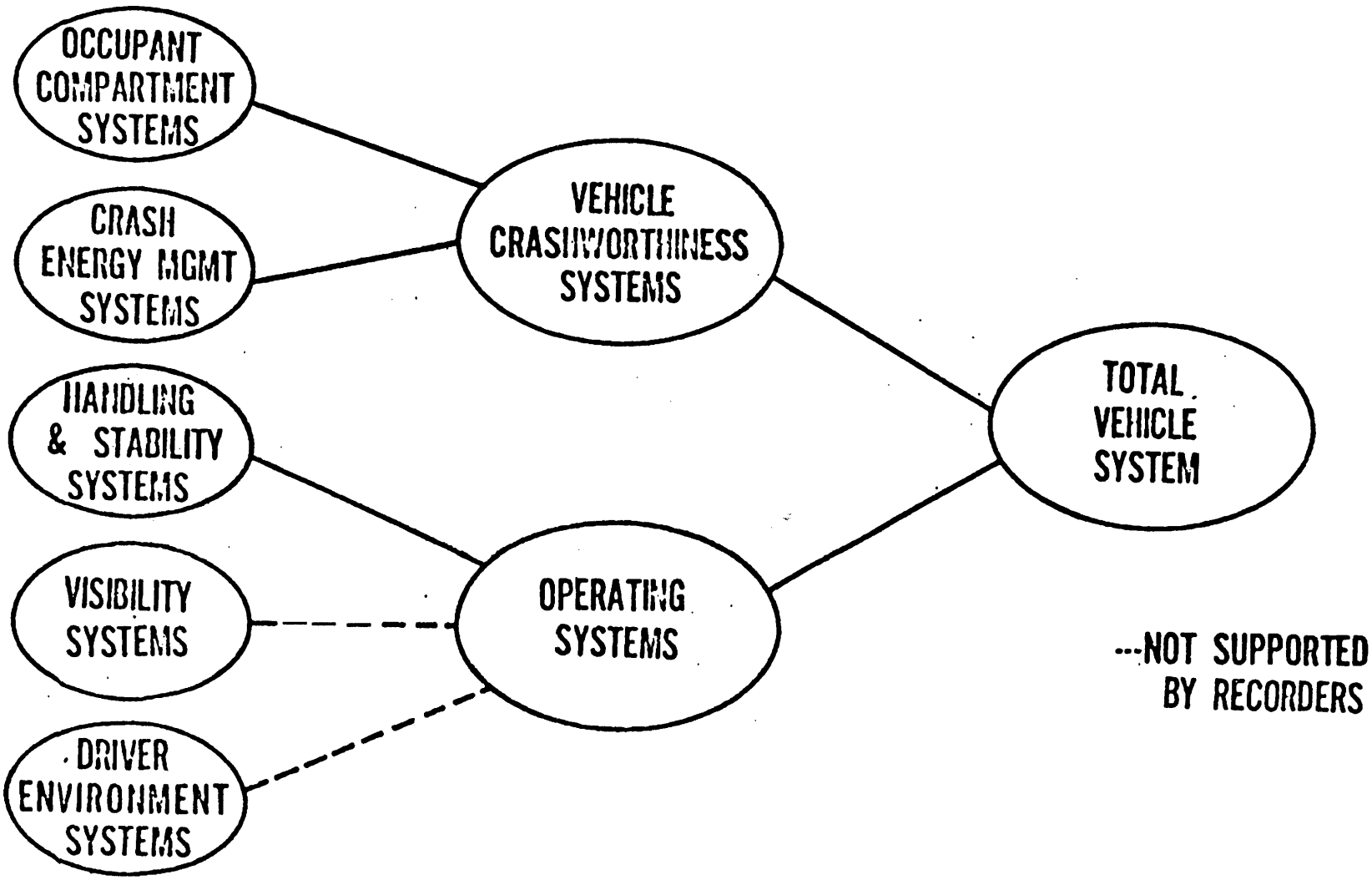



STUDIES TO DATE INDICATE DISAGREEMENT PRIMARILY AT THE HIGHER SPEEDS WHICH CONSTITUTE THE CRITICAL RANGE FOR CRASH SURVIVABILITY RULEMAKING

U.S. DOT/NHTSA

# AUTOMOTIVE RECORDER RESEARCH PROGRAM

## RECORDERS SUPPORT PLANS FOR SAFETY STANDARDS



U.S. DOT/NHTSA

# AUTOMOTIVE RECORDER RESEARCH PROGRAM

## RECORDERS AND TITLE II

AUTOMOBILE CONSUMER INFORMATION STUDY

PART B. CONDUCT A COMPREHENSIVE STUDY AND INVESTIGATION OF THE DEGREE OF CRASHWORTHINESS OF PASSENGER VEHICLES

RECORDER RESEARCH COULD SUPPORT THIS EFFORT DIRECTLY BY:

- PROVIDING CURRENT DATA ON SEVERAL VEHICLE MODELS
- CONFIRMING STUDY RESULTS OVER SEVERAL YEARS

APPENDIX G

# CRASH RECORDERS
# AND
# ALTERNATE METHODS OF DEFINING CRASH SEVERITY

James O'Day
HIGHWAY SAFETY RESEARCH INSTITUTE
University of Michigan

February 8, 1975
(date of receipt)

# Crash Recorders and Alternative Methods of Defining Crash Severity

James O'Day, Highway Safety Research Institute, university of Michigan

Received February 8, 1975

Precise and representative data on highway crashes in the United States have potential value in enactment of standards, design of new vehicles, and in the evaluation of recent safety improvements. Accident data collected to date have been intended to serve many purposes, and one of the consequences of such a multipurpose activity is that it may not solve any specific problem as well or as economically as would an experiment designed specifically for one purpose.

One of the measures desired by many concerning the U.S. fatal accident population is the cumulative distribution of fatalities by crash severity. This has frequently been put in the form shown in Figure 1 with the abscissa being a barrier equivalent speed. It is clear that if we knew the exact crash speed (defined in an understandable and meaningful way) for each fatal crash in the U.S. for, say, one calendar year, the curve plotted from that data would precisely define the population. If we could sample randomly within the same population we could define this curve with a degree of precision which depended on the sample size.

The crash recorders which have been proposed for installation are, of course, not capable of infinite precision nor do they necessarily report the barrier equivalent speed used in the wording of the standard. The test sequences in controlled crash tests reported indicate a 95% error of less than 2 miles per hour in the derived velocity change ($\Delta V$). The sample size required to achieve a precision in the vertical scale to that in the horizontal scale may be computed from a knowledge of the slope using the Kolmogorov-Smirnof test. For large numbers of cases ($N > 100$) the error in percent (95% bound) may be computed from:

$$\text{Error}_{95\%} = \frac{136\%}{\sqrt{N}}$$

For a 2 mph error in $\Delta V$, and a slope of the distribution of approximately 2.5 (percent/mph) the required sample size would" be 740 cases. There would be some gain, of course, in an infinite sample; but a more usual practice would be to define the sample size as above so as to increase the total error only by the square root of two.

SAMPLING CONS1DERATIONS: "

In order for the data for a sample to truly represent the national population, the sample must be properly drawn. If there is a bias in the sample, the output will not be representative. For example, if the mean age of the fatal occupants in the sample were ten years older than the mean age in the U.S. vehicle fatal population-- and with the assumption that 10 years of age were equivalent to 5 miles per hour in fatality probability, the curve of Figure 1 would exhibit a *bias* of the order of 12.5% in a downward direction. There are, of course, a number of other possible biasing factors. If all cars in the sampled group were full size (and the total population contained a large proportion of small cars) the distribution would be affected in the opposite direction. The biases given as examples here are estimates for illustration only, but they are not unreasonable. To get the true representation one must either sample in such a way as to eliminate the biases (e.g., random sampling) or collect enough additional information to adjust the data to correct for unwanted bias.

NUMBER OF INSTALLATIONS NECESSARY FOR 740 FATAL CASES

There are a number of ways of computing the number of installations necessary to compile 740 fatal crashes over some period of time. A simple one will be used here. With approximately 100,000,000 passenger cars in the U.S. and about 40,000 in-car fatalities per year, only one in 2500 passenger cars would have a fatality in it in a year. 740 fatalities, then, would require

1,865,000 installations. If a three year period were acceptable this reduces to approximately 622,000 installations. If a larger error were acceptable (say twice as large), we no longer need 740 fatalities but only 1/4 that number--and the sample could be further reduced to 155,000. So in three years with 155,000 installations there is a potential for defining the desired cumulative curve with a precision on the order of ± 10%. The various options are shown in graphical form in Figure 2.

## DISCUSSION

The statistical considerations above are based on a precise and complete sample. The mechanics of achieving this are not trivial. Placing a number of recorders in a sample of new cars biases the sample against older cars in the general population. And if these new cars were then distributed to the general population a high percentage of recovery would be difficult if not unlikely.

Placing the devices in a fleet (for example by agreement with an insurance company) should increase the probability of recovery--perhaps to a very high value. But this same action is likely to result in a non-representative sample in terms of age, sex, or car size. Adjusting such data to draw inferences to the national population is a questionable practice.

## ALTERNATIVES

A number of crash severity measures can be viewed as alternatives to the crash recorder. None have the advantage of producing a direct acceleration-time trace during impact. But most are applicable in principle to all cars. These include the CDC (Collision Deformation Classification) --a newer version of the VDI (Vehicle Damage Index), the SMAC computer programs developed by the CALSPAN Corporation, comparison of detailed crush measure-

ments on accident involved vehicles with results of instrumented crash tests (as described by Campbell in SAE paper 740565) or by Patrick (in an analysis of Volvo crashes). In addition, the TAD scale as applied by several police agencies is a crude measure of crash severity with the potential for relatively universal employment. Each of these will be discussed briefly below.

The CDC (or VDI)

The CDC was developed as a means of recording crash damage in a simple codable form. It consists of 6 elements--the clock direction of impact, four letter codes indicating the location of the damage (vertically and horizontally) and the general nature of the object struck, and a numeric code (1 through 7) indicating the extent of deformation. An experiment conducted by Cromack at Southwest Research Institute, and reported in an SAE paper, indicates that the CDC as presently defined can, in general, be assigned consistently by a trained investigator. The CDC, however, is not directly convertible into a measure of the crash dynamics because it depends in part on the structural characteristics of the particular car under investigation. Further, it was not developed primarily as a substitute for a measurement of the deceleration characteristics of the crashed vehicle, but rather as a simple codable record of crash damage.

The data elements contained in the CDC, however, when related to a knowledge of the vehicle structure (and perhaps other information about the crash circumstances) could permit a computation of some of the crash dynamics. An experiment could be conducted (largely with existing data) to define the ability of the CDC to predict much of the output desired from crash recorders. If an initial experiment looks promising, a large number of crash recorders in vehicles which are also measured with a CDC could lead to either (1) a calibrated CDC, (2) a redefined CDC which is more useful in the context of defining crash dynamics, or (3) both.

The CDC has the advantage that it can be applied to any accident vehicle after the crash without benefit of additional

instrumentation, and thereby reducing the problem of sample selection. It has the disadvantage, at present, that its capabilities for providing a measure of crash dynamics are not well known, and that these capabilities must depend on better knowledge of vehicle structure than is generally available in the literature.

The SMAC Programs

The MAC development is intended to provide computer assistance to the reconstruction of a traffic' accident. The method involves inputting certain observational and factual data into the computer, and iterating a solution which best fits the final rest positions of the vehicles involved. The iterative computer programs can be run from data acquired with a special observational tool (the SMAC van) or can be run with data taken by manual methods. In the latter instance, in particular, the technique should be applicable to a large number of collision analyses.

The present SMAC programs are limited to the ground plane, and, as a result, are not able to handle certain odd collision configurations-- such as rollovers, or vehicles running down an embankment. To the best of my knowledge the SMAC program output has hot been compared directly with crash recordings, although from some of the remakers at the recent conference I would assume that NHTSA has either started to make such comparisons or has done some. Crash recordings have been used to compute A **V.** This output of the SMAC programs has been validated to some extent.

In addition to the ground plane limitation, these programs are also limited by the accuracy of input data on the structural characteristics of the vehicle. However, the capability exists for removing these deficiencies. The problem of this point seems to be one of choosing the optimum tradeoff of input data requirements and modeling sophistication versus the detail and accuracy of the resulting output.

ments on accident involved vehicles with results of instrumented crash tests (as described by Campbell in SAE paper 740565) or by Patrick (in an analysis of Volvo crashes). In addition, the TAD scale as applied by several police agencies is a crude measure of crash severity with the potential for relatively Universal employment. Each of these will be discussed briefly below.

The CDC (or VDI)

The CDC was developed as a means of recording crash damage in a simple codable form. It consists of 6 elements--the clock direction of impact, four letter codes indicating the location of the damage (vertically and horizontally) and the general nature of the object struck, and a numeric code (1 through 7) indicating the extent of deformation. An experiment conducted by Cromack at Southwest Research Institute, and reported in an SAE paper, indicates that the CDC as presently defined can, in general, be assigned consistently by a trained investigator. The CDC, however, is not directly convertible into a measure of the crash dynamics because it depends in part on the structural characteristics of the particular car under investigation. Further, it was not developed primarily as a substitute for a measurement of the deceleration characteristics of the crashed vehicle, but rather as a simple codable record of crash damage.

The data elements contained in the CDC, however, when related to a knowledge of the vehicle structure (and perhaps other information about the crash circumstances) could permit a computation of some of the crash dynamics. An experiment could be conducted (largely with existing data) to define the ability of the CDC to predict much of the output desired from crash recorders. If an initial experiment looks promising) a large number of crash recorders in vehicles which are also measured with a CDC could lead to either (1) a calibrated CDC, (2) a redefined CDC which is more useful in the context of defining crash dynamics, or (3) both.

The CDC has the advantage that it can be applied to any accident vehicle after the crash without benefit of additional

instrumentation, and thereby reducing the problem of sample selection. It has the disadvantage, at present, that its capabilities for providing a measure of crash dynamics are not well known, and that these capabilities must depend on better knowledge of vehicle structure than is generally available in the literature.

The SMAC Programs

The SMAC development is intended to provide computer assistance to the reconstruction of a traffic' accident. The method involves inputting certain observational and factual data into the computer, and iterating a solution which best fits the final rest positions of the vehicles involved. The iterative computer programs can be run from data acquired with a special observational tool (the SMAC van) or can be run with data taken by manual methods. In the latter instance, in particular, the technique should be applicable to a large number of collision analyses.

The present SMAC programs are limited to the ground plane, and, as a result, are not able to handle certain odd collision configurations-- such as rollovers, or vehicles running down an embankment. To the best of my knowledge the SMAC program output has not been compared directly with crash recordings, although from some of the remakers at the recent conference I would assume that NHTSA has either started to make such comparisons or has done some. Crash recordings have been used to compute A V. This output of the SMAC programs has been validated to some extent.

In addition to the ground plane limitation, these programs are also limited by the accuracy of input data on the structural characteristics of the vehicle. However, the capability exists for removing these deficiencies. The problem of this point seems to be one of choosing the optimum tradeoff of input data requirements and modeling sophistication versus the detail and accuracy of the resulting output.

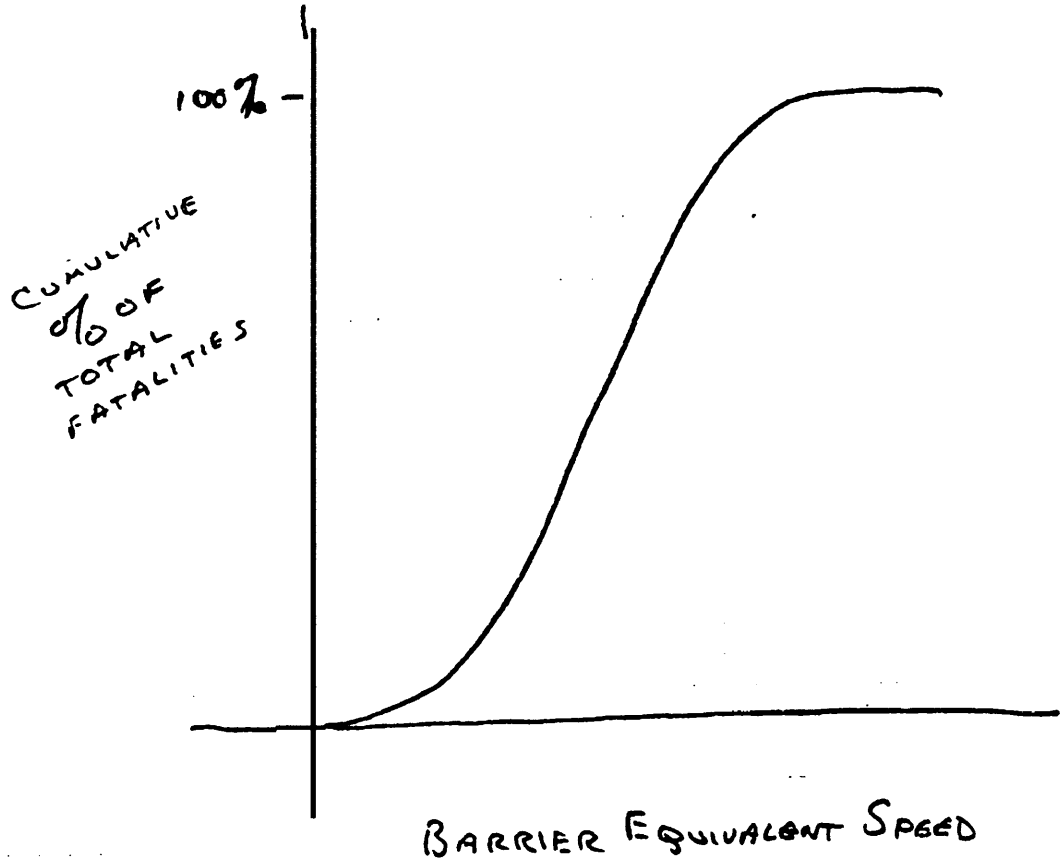



Figure 1: CRASH SEVERITY VS. CUMULATIVE FATALITY %.

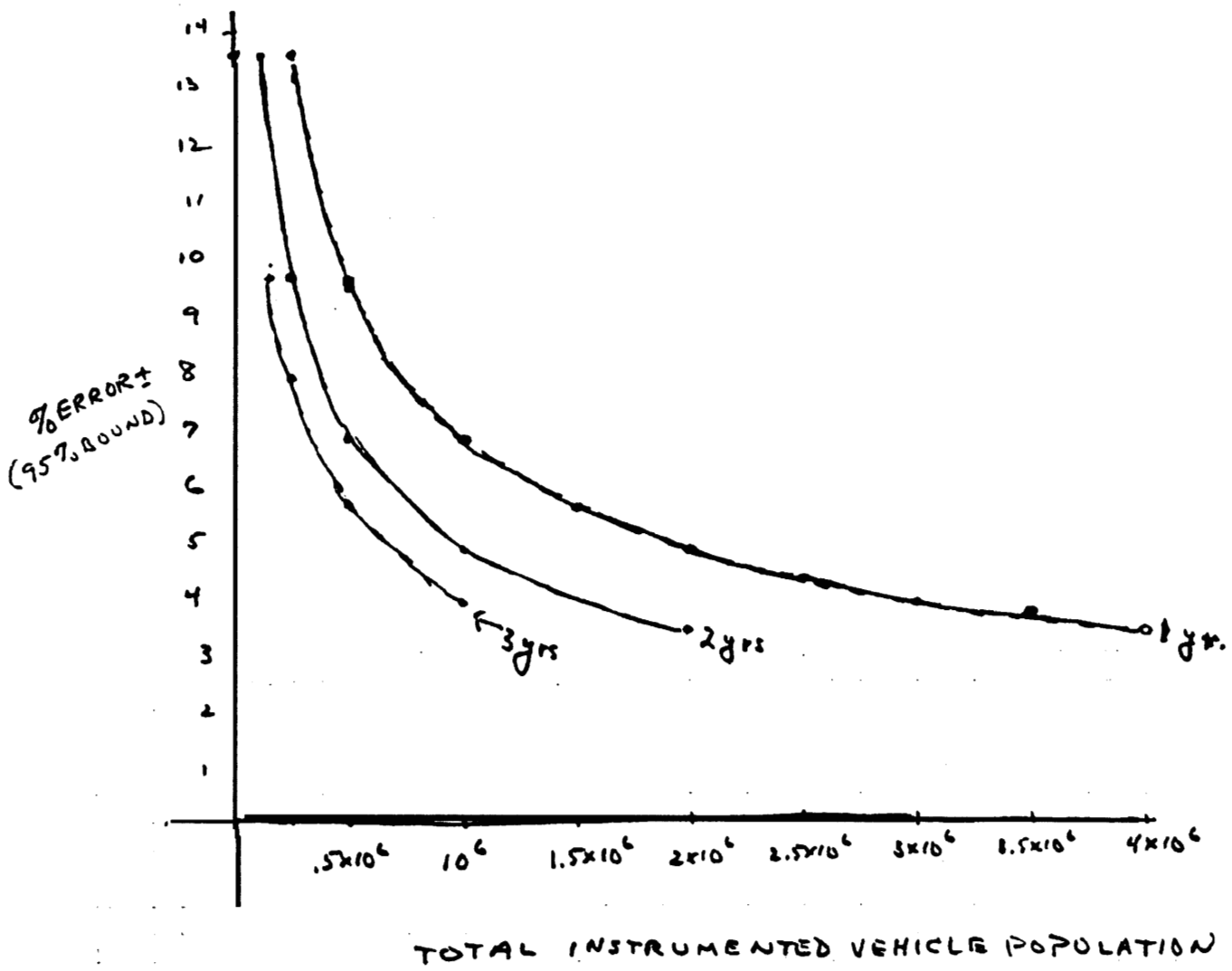



Figure 2: SAMPLE SIZE REQUIRED FOR A GIVEN 95% ERROR BOUND FOR 1, 2, AND 3 YEAR PROGRAMS.

APPENDIX H

# ADEQUACY AND LIMITATIONS OF CURRENT DATA SYSTEMS

Marie D. Eldridge
NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION

January 16, 1975

ADEQUACY AND LIMITATIONS OF CURRENT DATA SYSTEM

Remarks By
Marie D. Eldridge, Director
Office of Statistics and Analysis
National Highway Traffic Safety Administration

AUTOMOBILE COLLISION DATA WORKSHOP CONFERENCE
January 16, 1975 -Rosslyn, Virginia

In the very short time available to me this morning, I'm going to try to give you some highlights about the capabilities and limitations of our current data systems. I will also try to indicate where we see improvements within the near future.

However, before talking about the capabilities or the limitations, we really need to ask "capabilities or limitations for what?" So let's briefly talk about the objectives of our accident data systems.

First, we have to classify and count accidents. We need to determine the frequencies of accidents and classify them by their causal mechanisms, by their injury-producing potentials.

Second, we need adequate measurements of accident consequences, injuries, property damage or broader measures such as societal costs, a much neglected area and subject to great controversy but still one on which utlimately our decisions have to rest.

Third, we need to be able to describe, or model, crash injury mechanisms, that is, to relate the causal mechanisms and injury-producing potential to the actual occurrence of crash in-jury. This is particularly important in predicting the effects of proposed safety countermeasures. We have to

describe functional relationships between numerous factors which at present are considered separately. All of these things that I have mentioned enter into the process of determining the efficacy and the benefits of existing or proposed safety measures.

Let us consider the criteria by which we should assess our crash data collection systems. It is my view that a comprehensive approach and a comprehensive consideration of all the data requirements that combine to give us the needed information is essential. It just won't do to get vey high accuracy in estimating speeds if at the same time the sample of accidents for which we obtain this information cannot he used to generalize and cannot provide us with the proper support for a rule that will apply to the whole country. High accuracy in one part of the data system can easily be nullified by weakness in another and, to quote an old saying there is no need to put a micrometer on the end of the yard stick.

View Graph II

I have listed in this view graph some of the criteria that we may use in assessing crash data systems. There are many ways of doing this but this may help provide a framework for discussing our present systems.

First of all, there is the quality of the data. We are concerned with its representativeness and in our ability to generalize from it to a national crash population. A sample that contains only new cars or only auto fleets is not representative. Frequently, we may have a situation in which sample populations as defined are representative, but in fact, because of missing data or non-returns, we don't get an unbiased sample.

A second criterion is accuracy of information. One of the reasons we are here today is the inaccuracy of certain information that we are now getting in crashes, namely the various speed parameters.

A third criterion is the ability of the system to be responsive and timely. The data need to be collected and processed quickly enough that the information is available before the decision has to be made. The sample sizes have to be large enough that we can have confidence in the decisions based on the results. At the same time we have to concern ourselves with costs arid make tradeoffs between costs and precision. Next there is the breadth or extensiveness of coverage of the information provided by the system in the many parts of our highway safety information matrix. And last but not least the cost efficiency.

View Graph III

If we had a great deal of time we could consider all this at the data item level or individual field level, but even to cover this matrix in any detail will have to be left for possible discussion later in the conference. I will simply mention that under exposure items we have the characteristics of the vehicle occupants and the amounts of driving by various driver types, their characteristics, licensing, training and so on. We have under vehicle exposure the counts of vehicles by type, travel amounts, their conditions, size, etc. The

environmental exposure includes such things as traffic density, speed limits, highway types, design and so on. We could go down this matrix cell by cell and fill in the types of things that need to be considered.

The final and very difficult quality has to do with the cost-efficiency of the data systems. When a decision involves a high cost or an extreme inconvenience, a great deal of effort will generally have to go into the data collection and analysis. However, we also wish to keep our data collection efficient in the sense of not collecting information for which there is no need or employing personnel or equipment more skilled or more accurate than is really necessary.

IV. Now let us turn to the capabilities of some of our current data collection systems. Basically, we have two types of systems. The first is based primarily on the state or local traffic and related records systems. The second type involves special investigative work. The state records are kept primarily for purposes other than safety analysis. However, we utilize their records for the Fatal Accident Reporting System, which is essentially a census or 100 percent sample of fatal motor vehicle accidents and for the planned National Accident Reporting System, which will be a probability sample of all accidents, of a given threshold. The accuracy of the information provided through the State traffic record systems varies of course. In some areas of particular interest to us it is quite poor. Speed causal factors and restraint systems usage, for example, may be misreported or unreported frequently. Timeliness is generally not a problem. It usually takes only a few months before an accident is in the file and therefore accessible to us. As far as the

quantity of information is conerned, the State Traffic Record Systems are likely to provide us with a large number of cases for the more frequent types of accidents and the items of highest interest, but it is surprising how often in other circumstances we run out of data. The most obvious example is in making comparisons between makes and models of cars. When we get to some types of vehicles that are not on the road in large numbers, we have a very hard time collecting enough accidents to have a useful sample. The breadth of the information provided is generally not adequate. Impact speed for example is reported only in one State; traveling speed in about half the States and not for all accidents even in those States. Restraint system usage is not reported in most States and in many where it is reported, it is not reported for uninjured occupants. In-jury information and causal factors are sketchy. Post crash information, societal cost and property damage are usually not in the file.

It has been generally recognized that we can not obtain adequate information to support the standards by relying solely on these basic records oriented data systems. The second type of accident data collection system - those in which specific data collection efforts are sponsored or paid for by either the Federal Government or some other interested organization in the safety field such as MVMA or the Insurance Institute for Highway Safety. In these systems the investigation is likely to be carried out wholly or in part by professional accident investigators, resulting in substantially more extensive information. NHTSA has under way three types of sponsored studies.

First is the Multidisciplinary Accident Investigation teams. These teams do both on-scene and off-scene in-depth investigations. Teams have been performing clinical in-depth studies of selected accidents in the U.S., primarily on new cars, since 1969. The representativeness of the sample that has been produced up to this time is poor. Different teams have been covering accidents most relevant to their special interest. That situation is gradually changing. The accuracy is generally good. Nevertheless, there is considerable room for improvement. We have no capability for getting a time history of the crash forces and accompanying accelerations except through computer simulation such as the SMAC program. At present we have about 6,000 MDAI cases in the file. Many of these were not the result of on-scene investigation. There is detail on most aspects of the accident with the exception of exposure. As a system for producing statistical information needed for supporting our safety standards, the on-scene in-depth investigations cannot be regarded as cost effective. The average cost per case is about $2,000. The cost decreases to about $800 per in-depth case if the on-scene investigation requirement is Eliminated. This does reduce the accuracy of reconstruction of the accident and of course affects the estimate of speed.

At a somewhat lower level of detail NHTSA has developed a system in conjunction with MVNA to collect a probability sample of towaway involvements of new cars in five selected regions of the country primarily for the purpose of evaluating active and passive restraint systems.

Data are assembled from the police report, a doctor's report, photographs, a brief vehicle investigation, and driver interviews performed by field technicians. Data items collected are restricted to those needed for the statistical analysis of restraint systems effectiveness. This is an example of what we may term a Level II study. We expect to make national estimates based on post stratification. The accuracy of the information in the selected data items' should be good, nearly as good as what comes from the multi-disciplinary in-depth investigations. The quantity will be adequate to match the needs for estimating safety belt effectiveness. Because of the small numbers it is not likely to give us what we need for estimating air cushion effectiveness, very soon. As far as the breadth of the file is concerned, it is designed for calculation of crash injury rates and evaluation of restraint systems effectiveness. It does not address exposure or accident causation. Speeds and occupant contact points are not determined. The cost is around $100 per case.

A third type of sponsored system is basically a bilevel investigation or one in which there is a supplementary investigation carried out by police with NHTSA or other funds added to take care of added costs. We have under development a system for sampling pedestrian and bicyclist accidents in several hundred localities. The system is designed to answer questions at the level of detail that we

nced to determine gross behavior and provide some good input for counter-measures. The data to be collected is primarily concerned with the nature and location of pedestrian and cyclist accidents as well as certain other items affecting visibility which would not normally be collected in the state accident reporting system. The cost per case is expected to be high primarily because of the relative rarity of pedestrian and bicyclist accidents and because in order to get an adequate probability sample that will properly represent rural areas, it is necessary to include localities with a very low frequency of accidents. The set-up time in preparing to get the supplemental investigations done in small localities is the same as it is in large localities, but the data rate is low and the total cost is increased disproportionately.

V. As we look ahead to potential improvement in the capabilities of our current systems that may be in sight we are really moving in two directions. The first is to create a national accident sampling system based on a probability sample. We have a contract under way with the Highway Safety Research Institute at Ann Arbor to help develop this system that will include some of the current investigative efforts but provide for sufficient. control of the selection of accidents that we will get a sample from which we can generalize to national crash populations.

The second area in which we anticipate improvements is in determining crash dynamics. These efforts, are of courses of paramount importance to this workshop. The work with the crash recorder is being covered by my colleague, Lynn Bradford. The other approach, SMAC, the computer simulation of the accident dynamics will be dealt with by our representatives from Calspan but I would like to say a few words about our experience with it.

This program uses vehicle rest positions and impact damage to calculate impact velocities, the velocity change during the crash, acceleration pulse and predicted damage. The goal is to reconstruct the accident crash dynamics in sufficient detail that inputs needed by our standards makers are available. The use of the SMAC program may permit us to get, at a reasonable cost, an adequate representative sample of crashes once our national accident sampling program is up and running. However, it should be pointed out that the crash recorder is clearly a very valuable tool in developing necessary refinements to the SMAC program. Ideally, and this is a moderate size "if", IF the crash recorder and the SMAC work hand in hand well enough, we can succeed in reducing considerably the numbers of crash recorders required down stream. Precise calibration of the SMAC program will enable us to use Level 2 data for crash dynamics at a reasonable cost. Currently the cost per case, using the SMAC program is $150.

In the short time available to me I have had to gloss over lightly much of the work related to crash data. Three members of my staff are here to provide detailed back-up and to join in any subsequent discussion of these points. They are Don Mela, Dr. Charles Kahane and Dr. Charles Moffatt. Before finishing these very brief remarks, I want to repeat a point I made earlier. We need to consider all relevant aspects of the data systems in a comprehensive fashion before making decisions on any of them separately. While we may not be able to devote that amount of detail to all aspects of the data systems in this conference I think that at least the major aspects should be considered before coming to any conclusions or decisions.

## OBJECTIVES OF ACCIDENT DATA SYSTEMS

CLASSIFY AND COUNT ACCIDENTS

MEASURE ACCIDENT CONSEQUENCES

DESCRIBE CRASH INJURY MECHANISMS

# CRASH DATA SYSTEMS CRITERIA

## QUALITY

- Representativeness
- Accuracy
- Recency
- Responsiveness to Ad Hoc Demands

## QUANTITY

- Precision
- Costs

## BREADTH

## COST-EFFICIENCY

DATA BREADTH

| | HUMAN | VEHICLE | ENVIRONMENT |
|---|---|---|---|
| EXPOSURE | | | |
| CAUSAL FACTORS | | | |
| CRASH EVENTS | | | |
| CONSEQUENCES | | | |

## CURRENT DATA COLLECTION SYSTEMS

A. STATE/LOCAL RECORDS SYSTEMS
   1. FATAL ACCIDENT REPORTING SYSTEM (FARS)
   2. NATIONAL ACCIDENT REPORTING SYSTEM (NAI

B. INVESTIGATIVE FILES - SPONSORED STUDIES
   1. MULTIDISCIPLINARY ACCIDENT INVESTIGATION (MDAI)
   2. RESTRAINT SYSTEMS STUDY
   3. BI-LEVEL STUDIES

---

## NEW AREAS

1. NATIONAL ACCIDENT SAMPLING SYSTEM

2. MEASUREMENT OF ACCIDENT DYNAMICS

APPENDIX I

# A DISCUSSION OF DATA GATHERING SYSTEMS

Edwin A. Kidd
CALSPAN CORPORATION

January 16 & 17, 1975

## INTRODUCTION

A data sampling plan that provides an accident data file that is representative of the national population is important. A system for data processing, storage and retrieval to allow early determination of trends in accident, injury and fatality frequencies is essential. But the most critical problems are those concerned with the collection of consistent, coherent data on individual accident sequences with a volume far exceeding that now available.

For too long, those concerned with accident studies of the effects of safety standards already in force have had to make do with either too small samples of reasonably good data or relatively large samples of data whose content is inadequate for the purpose. In the first category is the data bank (and "bank" is too grandiose a term) that has resulted from the individual federal teams of multidisciplinary, professional investigators. These teams can serve useful purposes in special studies, in discovery of problems that would otherwise go undetected and, particularly, in the area of accident causation. By their very nature, they cannot provide a sufficiently large data sample relevant to the implementation of standards aimed at injury and fatality reduction without excessive expenditure of funds.

In the second category are the presently available state data banks of relatively low content data obtained through the use of routine police and driver reporting. These data have been valuable in demographic studies, in the broad-look definition of trends and in statements concerning the magnitude of the overall problem, primarily in fatality frequency. In most cases, such data is totally inadequate in content and precision and, despite the relatively large numbers available at relatively low cost, cannot adequately define injury and fatality reduction resulting from standards implementation.

There is a third category of data collection systems that has evolved over the past few years that lies between the very detailed team approach and the routine police reporting as established independently by the states. The potential exists with this multi-level approach at selected centers around the country (present examples include Calspan in New York, HSRI in Michigan, and HSRC in North Carolina) for a combined data bank that would be a major step toward the attainment of a greatly increased sample size with, and most important, accurate individual accident data with the content required for the purpose.

proper utilization of the potential of these data centers can be realized only if investigator and accident reconstruction aids are implemented that will allow the police to obtain the necessary information with orders of magnitude improvement in accuracy. Local and state police already have the charter to investigate accidents. There are no unsurmountable problems in providing them with the new tools that have been developed for collecting the data that would be the basis for a national data bank sufficient for NHTSA needs in surveillance and effectiveness studies.

## Data Requirements

The list of specific data elements in each accident that are deemed to be essential can hardly ever be complete for the serious analyst. However, the routine and continuous collection of accident data can be tedious, time consuming and costly. Every effort must be made to keep the data requirements to a sufficient set commensurate with the need.

Such sets have been defined a number of times for various ongoing studies. The one presented in Figure 1 is stated in somewhat general terms as it is required, in this instance, primarily for the comparison of data gathering techniques.

| DESCRIPTION OF | INDICATION OF DETAIL |
| --- | --- |
| GENERAL ACCIDENT INFORMATION | SINGLE OR MULTIVEHICLE, RURAL OR URBAN, HIGHWAY CATEGORY, PROPERTY DAMAGE ONLY OR INJURY, OBJECT STRUCK, OVERALL SCENE DESCRIPTION, ROAD SURFACE, AMBIENT CONDITIONS |
| SPECIFIC VEHICLE INFORMATION | MAKE, MODEL, VIN, IMPACT DIRECTION AND DEFORMATION (VDI OR IMPROVED EQUIVALENT), AVAILABLE RESTRAINT SYSTEM, LOADED WEIGHT, TIRES, INOPERATIVE SYSTEMS PRIOR AND AFTER IMPACT |
| OCCUPANT INFORMATION | SEX, AGE, HEIGHT AND WEIGHT, INJURY (MEDICAL REPORT), SEATED LOCATION, USE OF RESTRAINT SYSTEM |
| DRIVER INFORMATION | DETAILS AS IN OCCUPANT ABOVE PLUS DRIVING EXPERIENCE, TRAINING, CONVICTIONS, PHYSIOLOGICAL CONDITION, PSYCHOLOGICAL INDICATIONS, ACTIONS PRIOR TO AND DURING ACCIDENT SEQUENCE |
| IMPACT ENVIRONMENT | SPEED AT IMPACT, RESULTANT SPEED CHANGE TIME HISTORY, COMPARTMENT DECELERATION |

Figure 1 ACCIDENT DATA REQUIREMENTS

Even this rather simplified listing appears formidable. However, to some degree each of these elements or approximations thereof are being obtained, by one means or another, by some of the present ongoing programs. There is no element in the outline presented in Figure 1 that is not germane to existing standards. If we settle for a system that provides accurate information less than this and/or only for a quantity of a few hundred or even a few thousand cases, NHTSA cannot do the job it has been directed to do.

Definition of the total number of accident cases required annually for an adequate national data bank can be made if (1) the questions to be asked of the system can be identified both for the present and future; (2) the accuracy with which the particular data elements can be measured is known or can be appropriately approximated; and (3) the statistical analysis techniques to be employed can be agreed upon. This is not meant to imply that such analyses and decisions should not be made. However, there are no statistical procedures that can adequately overcome the past and current inaccuracies with which such extremely important data elements as impact speed and speed change have been reported if they have been reported at all.

There will need to be a parallel effort of statistical analyses to indicate what questions can be addressed with acceptable statistical significance as a function of particular sample sizes along with the determination of the funds that can be made available. The financial impact of standards on the consumer has been and will be considerable – billions of dollars annually. Figure 2 presents an average cost per car for the FMVSS to date based upon idividual autombile manufacturer's data. It seems prudent to schedule funding for the primary surveillance effort -- accident data collection -- commensurate with the far reaching decisions that depend upon such data.

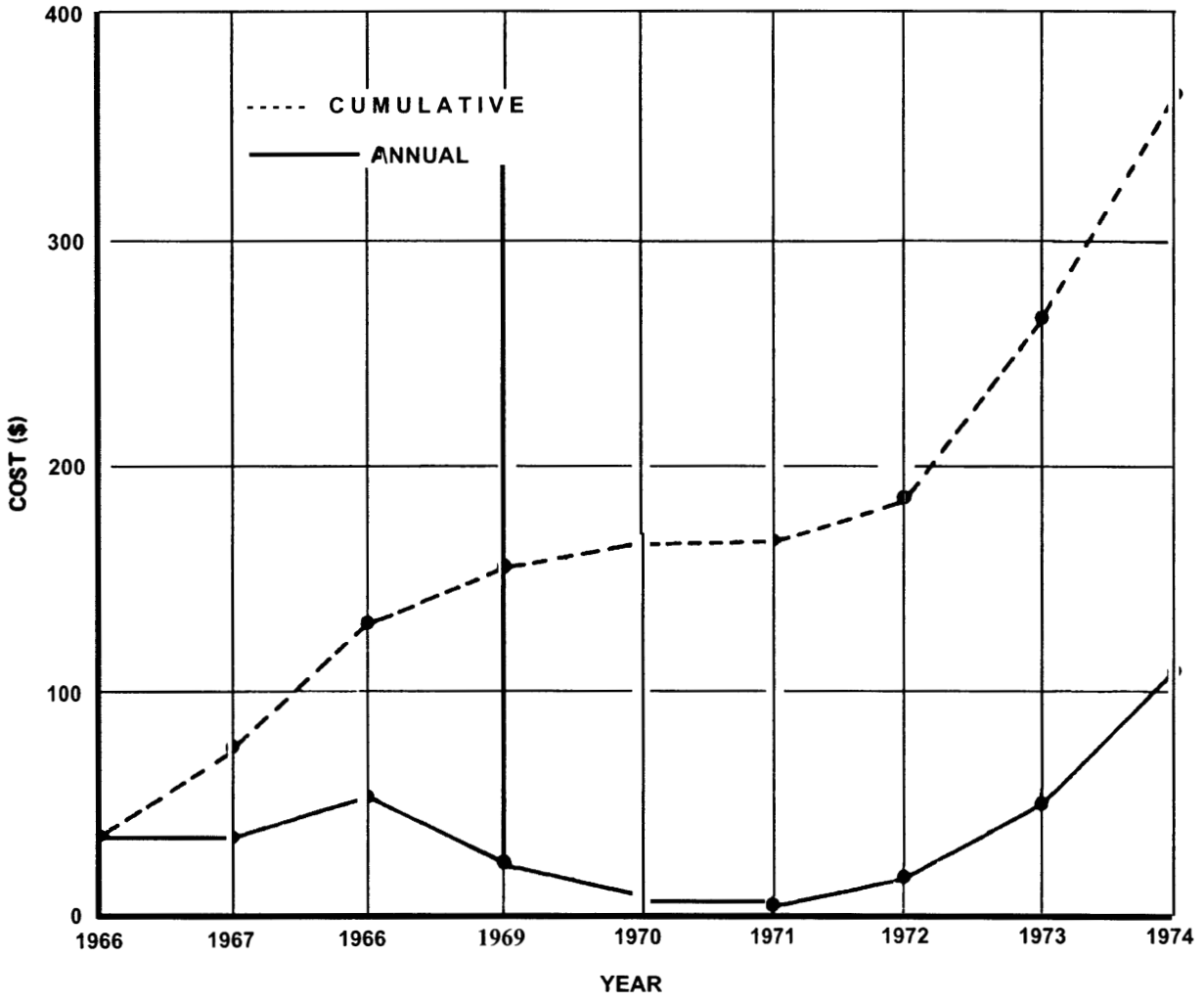



Figure 2 AVERAGE COST IMPACT OF FMVSS ON AUTOMOBILE PRICE

## Data Gathering Techniques

In order to obtain the data required on each accident, every accident analyst would gladly utilize whatever data gathering techniques are available. Ideally, crash recorder information, police and driver reports, intensive investigation team reports and on-scene reconstructions of the accident through computer aids to investigators would all be gratefully accepted by every serious analyst for each accident. In fact, no analyst would refuse any available high speed photographic coverage (in color, of course) .

Obviously, it is neither practical nor essential that all of these systems be provided for the achievement of the basic national data bank. It has already been stated that the intensive investigation teams may have other purposes but cannot provide the data in the quantity required. It has also been noted that existing state data, comprised of merged police report, vehicle registration, and driver licensing files do not provide the content required for the evaluation of safety standards.

## Crash Recorders

Crash recorders can only provide a portion of the desired information as a supplement to continuous accident investigations. At best, a recorder can provide only the information outlined in Figure 1 under "Impact Environment" plus driver control actions during the accident sequence and an identification of the vehicle in which it is installed. Despite the fact that the information a crash recorder is designed to obtain is the impact environment, and that this is the data now totally lacking or sadly inaccurate, a detailed accident investigation would still be required to provide the essential general accident, specific vehicle (including the other vehicle) and occupant and driver information. Thus, the overall cost of an accident investigation would include essentially the present costs plus those associated with the provision of crash recorders.

The numbers game must be considered as well in the consideration of crash recorder Installations. The actual number of accidents that would be available for analysis would be a marked attenuation from the total number of crash recorder installations (Figure 3). Assume that recorders were installed in 100,000 automobiles. No more than 1 in 4 of these automobiles would be involved in any sort of accident annually. This reduces the number of accident cases with crash recorder information to no more than 25,000. If it is further assumed that the accidents of principal interest are those of more than minor severity, for example, tow-away accidents (approximately 12.5% of all accidents), then only 3125 accidents would be available. If we examined only the highest volume model of the major American manufacturer (approximately 25%) the number of accidents available would be approximately 781. Further division of these accidents into accident type, direction of impact, etc., would further diminish the numbers. This severe attenuation would be greatly increased for car make and models other than the one with the greatest penetration of the market.

It is recognized that the crash recorder is designed to provide crucially important information on impact environment that has not been otherwise available, at least in quantities with acceptable accuracy, However, there is now available another method, as discussed below, for obtaining this information with accuracies that appear excellent. Both methods should be compared in staged crash tests and considered for some possible joint use as mutually reinforcing data sources. However, the computer aided system, with its outputs of a detailed scene description and an accurate reconstruction of the accident, offers the most promise, as a fundamental element of a continuing data gathering system.

| | |
|---|---|
| ANNUAL CRASH RECORDER INSTALLATIONS | 100% |
| VEHICLES INVOLVED IN ANY ACCIDENT | 25% |
| VEHICLES INVOLVED IN SEVERITIES > TOW AWAY | 3.125% |
| MAJOR MANUFACTURERS HIGHEST VOLUME MODEL | .078% |

Figure 3 CRASH RECORDER EQUIPPED AUTOMOBILES IN ACCIDENTS

Computer Aided Investigation and Reconstruction

With support from NHTSA, **this** on-scene accident investigation and reconstruction system has been developed and demonstrated (Figure 4), An automated range-finder transit with associated computer hardware and readout (Figure 5) provides a drawing of the accident scene and supplemental accident information as required (Figure 6). These physical evidence data are transmitted via a radio-telephone link to a centrally located computer which returns a reconstruction of the accident (Figure 7).

In actual reconstructions of staged accidents, this investigator's tool has faithfully reproduced the accident sequence with impact speed and speed change reconstructions of 2-3% accuracy. With this system, police investigation teams can generate high quality accident data in the course of performing their normal police functions. Yet the system has been found, during field trials by police personnel, to actually ease the tasks of scene measurement and reporting. Thus, both the users of accident data and the police can benefit from adoption of this system.

The economics of adopting the system would be extremely attractive from the viewpoint of elimination of labor costs in the generation and reporting of accident data for research purposes. The end product is already in digital format for statistical analysis.

The nature of the output from the van also lends itself directly to a central data bank or regional data banks receiving reconstructed accidents and supplementary data over existing telephone lines. This continuous updating of current data is particularly attractive. At present, the best a state can do, those few that can supply merged accident tapes, is provide a year's data six-eight months after year's end. A dedicated data collection center, such as presently sponsored by NHTSA, can provide computer tape updates of collected, augmented police reported data every three months with a three month delay.

Figure 4 ACCIDENT VAN

Figure 5 VAN INTERIOR

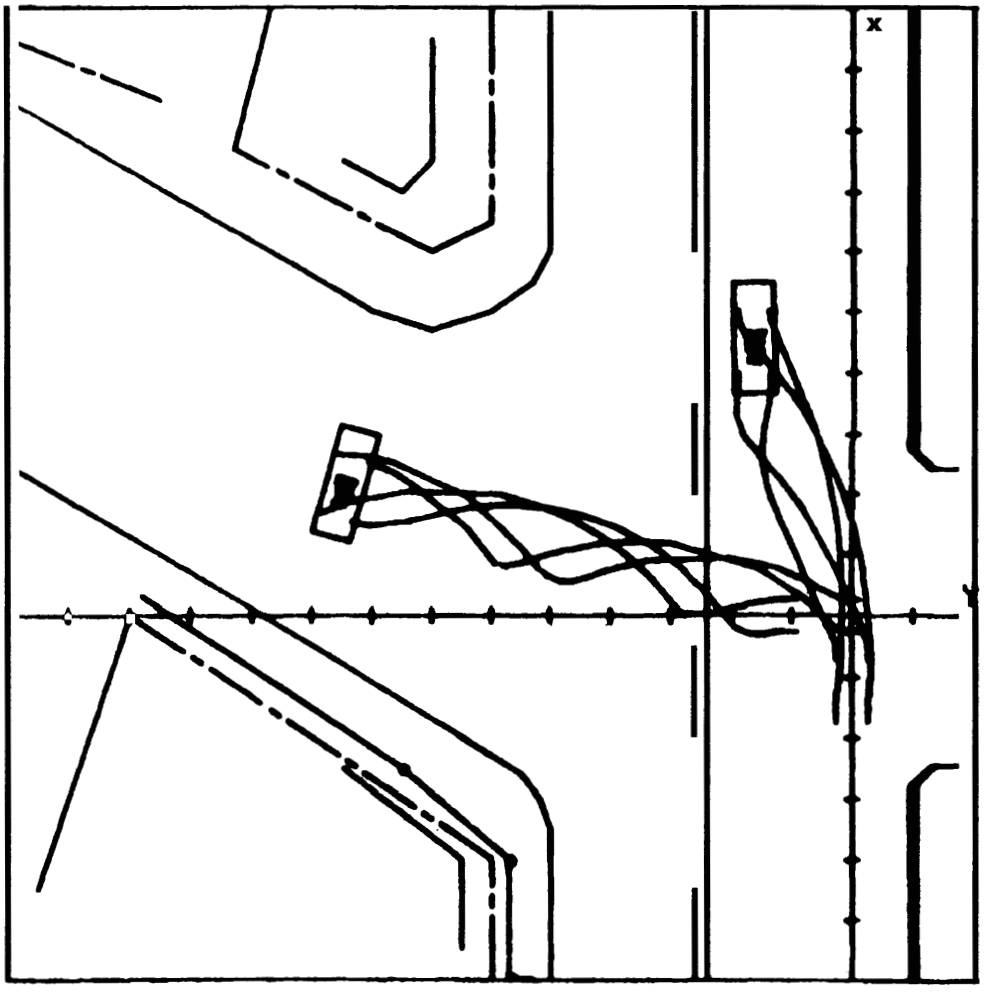
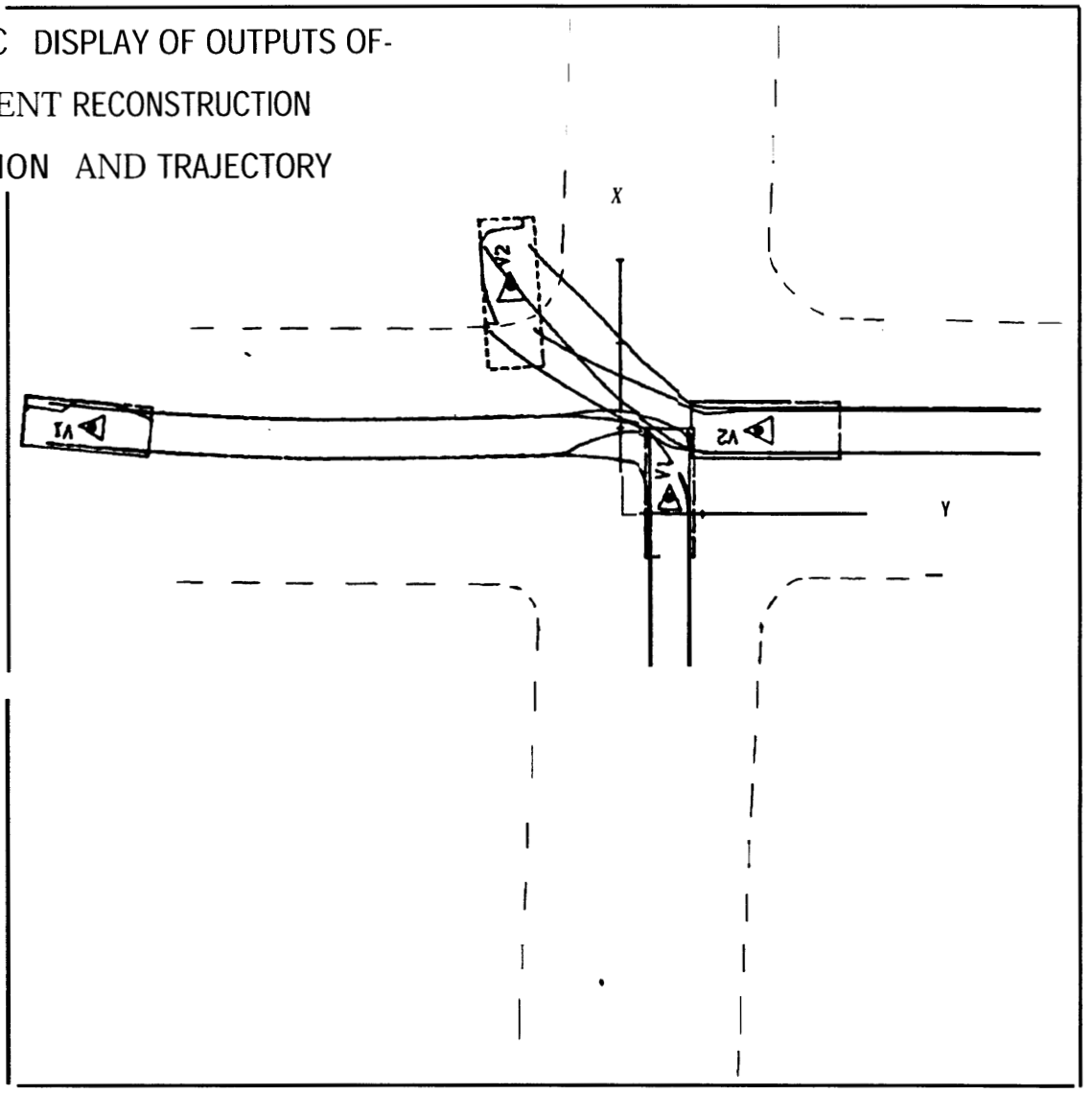



FIGURE 6 SCENE AND ACCIDENT RECONSTRUCTION

GRAPHIC DISPLAY OF OUTPUTS OF-

ACCIDENT RECONSTRUCTION

COLLISION AND TRAJECTORY




AXIS INTERVALS ARE 10. FEET

| | RECONSRUCTED POSITIONS AND VELOCITIES AT IMPACT | | | | | | DISPLAYED FINAL POSITIONS | | | | VEHICLE | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C.G. POSITION | | HEADING | | | | C.G. POSITION | | HEADING | | DAMAGE | |
| | XC1 | YC1 | PSI1 | FWD | LATERAL | ANGULAR | XC1F | YC1F | PSI1F | REMARKS | INDICES | ΔV |
| | FT. | FT. | DEG. | MPH | MPH | DEG/SEC | FT. | FT. | DEG. | | | MPH |
| VEHICLE # 1 | 1.9 | 6.0 | -0.0 | 24.0 | 0.0 | 0.0 | 10.3 | -69.5 | -84.6 | IN MOTION AT 3.5 SEC AFTER INITIAL CONTACT | 02RYEW4<br>04RZEW3 | 20.8<br>10.2 |
| VEHICLE # 2 | 10.0 | 16.8 | -90.0 | 36.8 | 0.0 | 0.0 | 26.8 | -13.2 | -3.4 | VEHICLE AT REST | 11LFEW2<br>09LPEW3 | 14.9<br>6.9 |

FIGURE 7 SAMPLE ACCIDENT RECONSTRUCTION

## Data Collection System

The mobile accident van will be ready for general police use after some additional field trials. The accident reconstruction computer model (SMAC), as incorporated in the van reconstruction software and hardware, has been distributed and is in extensive use; additional validation for a variety of accident situations is planned. The total system works and works very well. It is appropriate to consider how it might be incorporated in a complete data collection system that would provide greatly improved and accurate accident data with the quantity and content required for surveillance of standard's effects on accident consequences.

strategically located data centers have been suggested for the collection of regional data samples of the multilevel type. An appropriate distribution of accident vans for use by police investigators within each of these regions would provide continuous data into regional data banks and/or to a single data bank.

There are two primary options for van configurations. Police investigators can be equipped with either a Scene Van or a Reconstruction Van. The Scene Van would provide a description of the physical evidence and supplementary accident data (Figure 6). Hard copies of this information in appropriate format would be available as the police report. In addition, all data would go on tape cassettes to be forwarded to the particular data center for reconstruction of the accident by the SMAC model. The reconstructed accident information would supply the center's data bank. Appropriate retrieval and analysis programs would provide immediate analyses as required.

Reconstruction Vans, with either self-contained reconstruction capability (more comprehensive on-board computing equipment) or radio link to a regional time-share computer or computers would by-pass the step of accident reconstructions at the data center. Also, a more desirable feedback to the accident investigator at the scene would be available. (A successfully reconstructed accident is the best check of the completeness and accuracy of the scene data.) Each reconstructed accident data set would be stored (short term) within the van and transmitted, when convenient, to the data center at available terminal locations (already present at police agencies).

A rough approximation of the cost for two assumed data collection systems is given in Figure 8. These are given to provide an approximate range of system costs for collection of 100,000 cases annually. The cost of a radio link reconstruction van system would be somewhat less than the self-contained van with a resulting overall cost close to that of the scene van system. Final selection among these alternatives should consider, in addition to basic costs, operational factors including the advantages of program updating and modifications with either the Scene Van or the radio-link Reconstruction Van and the overall data improvement that would result from the Reconstruction Van.

Regardless of the system selected, costs per case of less than $100 are estimated. This appears to be quite a bargain. The system would provide 100,000 cases per year for whatever investigation criteria is desired, e.g., tow away cases. Costs per case are essentially independent of data sample size. The assumed rate of cases per van per year is conservative, considering that police agencies operate 24 hours per day.

| | ANNUAL COST - $1,000 | |
|---|---|---|
| | SCENE VANS | RECONSTRUCTION VANS |
| 500 VANS | $2,500* | $3,750* |
| 5 RECONSTRUCTION UNITS (1DEDICATED MINI COMPUTER AT EACH OF 5 CENTERS) | 0* | – |
| SERVICE, MAINTENANCE | 500 | 750 |
| ACCIDENT RECONSTRUCTION LABOR (100,000) | 200 | – |
| TOTAL | $3,210 | $4,500 |
| COST PER CASE 100,000 CASES, 200/YEAR/VAN) | $32 | $45 |

*ASSUMES 10 YEAR USEFUL LIFE WITH LINEAR DEPRECIATION

Figure 8 APPROXIMATE, FIRST-O[illegible] COSTS ACC DENT VAN DATA COLLECTION SYSTEM

Obviously missing from the estimates of Figure 8 are labor costs for data acquisition at the scene. It is assumed that police would do the investigations. With police use of vans for their own investigative purposes with improved efficiencies over the present, acceptance of the vans should be readily realized.

Based upon our experience to date, we believe the usefulness of this mobile system to the police themselves can be demonstrated and there is no concern that they cannot properly operate the equipment. Implementation of a total collection system employing this scene data gathering capability will provide NHTSA with the information needed.

APPENDIX J

# MASS ACCIDENT DATA ACQUISITION AND WHY IT'S NEEDED

John Versace
FORD MOTOR COMPANY

January 16, 1975

# MASS ACCIDENT DATA ACQUISITION AND WHY IT'S NEEDED

OTA Automobile Collision Data Workshop
John Versace
Ford Motor Company
January 16, 1975*

Accident data have been collected for a long time, and we have learned a great deal from them. These data aid in establishing safety need and proper priority of effort. Government, industry, and the public can benefit from more knowledge regarding the real world of traffic. However, times change and designs change, and we believe the present rate of gathering accident information on current designs and events is not adequate. Large amounts of data, carefully collected so as to assure representativeness, are needed. In addition, special kinds of data, more accurate than numerous, perhaps, are also needed to fill in some significant research lacks.

## Approaches to Data Collection

There are three basic approaches to data programs, with some variations. First, the researcher might incisively phrase the particular questions that are going to be asked of the data, and he would design a data collection program to answer those questions. A point of particular significance in this approach is that the data collection program is then part of an integrated research project. For example, both the MVMA and NHTSA have, during the past year, been conducting a study to measure the accident performance of the 1974 interlock type of restraint in comparison to the 1973 system. The number of items of data collected in each case were deliberately kept few so that investigative resources could be allocated toward getting as many cases as possible -- instead of much data on fewer cases.

*With additions, January 22, 1975.

The second type of approach would be to run the study like a controlled experiment, in which the hardware to be evaluated would be assigned to members of the public in such a way that there would be both broad representativeness of use and freedom from bias, those not receiving the device being the control group. This approach is seldom practical, although manufacturers sometimes are able to equip certain cars with experimental features prior to their full market introduction in order to develop field experience with them. Again, the data collection is integrated into the research project.

The third approach to data collection -- and the one I believe we are concerned with here -- is to create a data file which is a microcosm, in all its particulars, of the real world. This approach is independent of any particular research project; its purpose is for the data file to "become" the real world insofar as any researcher is concerned. Different researchers will dip into that data file to answer questions which may arise as issues emerge, issues perhaps unforeseen by those who devised the data collection scheme. Such a method requires highly detailed recording of data on an enormous number of variables. This allows for variables previously disregarded to now be investigated, and also allows the researcher to control confounding effects by selecting for comparison only those cases in which the effect of the extraneous variables cancels out. The most desirable kind of data collection approach, providing sufficient resources can be brought to it, is this third type. If resources are not sufficient, then probably the first type of approach -- in which the data program is tailored for the specific questions to be asked of it -- would be most appropriate.

Uses for Data

Among the uses for accident data -- and each use has its own requirement on scope and precision -- are: (i) evaluating the safety performance of

past and current safety designs, and most importantly, verifying that required countermeasures have, in fact, been effective; (ii) determining if particular safety problems are of such magnitude that countermeasures are needed and supporting the specifics of rulemaking; and (iii) supplying normative information about accident occurrence so the future effect of countermeasures not yet designed or produced can be anticipated and a wise policy regarding them be instituted.

In regard to this last point -- anticipating future performance -- let's consider an example. It is easy to conclude that if the 30-mph crash test requirement contributes to reducing death and injury, then surely an increase to 40, 50, perhaps even 60 mph would be proportionately better. But there is very little information available that would unequivocally support such a conclusion. Because there obviously are no cars on the road meeting such advanced requirements, we cannot test this conclusion by comparing their casualty rate to cars meeting only the 30 mph criterion -- assuming we had accident data collection and analysis procedures adequate to the task. Because there are no such cars, resort must be made to calculation.

Two things are needed to make such calculations: real inputs of population exposure -- drawn from accident data -- and theoretical system models. Validity of the models will of course be an important matter to consider.

Need for Population Exposure Data

Being able to determine whether, or in what way, to increase the test requirements of crash performance standards, or to inaugurate any rule, depends upon our being able to predict the probable effect of such actions in the future. particularly lacking as an input to any calculation of future effects, is an accurate estimate of the dynamic environment to which people are exposed.

The particular form of the exposure variable most useful for calculating the magnitude of need and in estimating the future effects is the probability distribution of collision speed (with all types of likely obstacles.)

Ordinary accident investigation data can be useful in estimating crash speeds, given some care in adjusting for the mechanical nature of the struck object. However, derived speed estimates from accident reports quickly lose reliability as impacts other than head-on are considered. The ogival cumulative distribution of barrier-equivalent speed has been a prominent part of most analyses aimed at estimating population exposure and hence need, and in calculating the probable effectiveness of different restraints. A single shape and location of this curve has not been accepted among all its users. The absence of this one item of information on occupant exposure can make what should be a factual matter rather a matter of contentious advocacy. It is our belief that a crash recorder supplement to a general accident data program has the potential to assist in clarifying this particular area of need.

Accuracy of Crash Severity Data

For a successful program of crash severity determination, there must also be the right protocol for defining an accident so that the resulting distribution of measurements is not biased upwards by deliberately selecting only "interesting" cases -- an unfortunate characteristic of most data sets available today. If the speed distribution is incorrectly displaced upscale, or inflated due to errors of measurement, there will appear to be many more high speed crashes than really occur; the result will be to lean toward excessively high crash requirements, with resulting cost-effectiveness being less than it appears. While precision of measurement of crash speed is important in estimating the speed distribution, it is even more important that there be no bias in the data collected.

It may be useful here to distinguish between the accuracy of measurement and the accuracy of estimation, in the statistical sense. The former refers to the degree of correctness in any one reading, and the average measurement error is an index of this quality. Accuracy of estimation, for data analysis purposes, refers to the relative absence of bias in the sample of data: i.e., that the sample values fairly reflect the population from which they were drawn: that the sample distribution can be accepted as an estimate of the population distribution because there are no funnies in it which warp it, or skew it, or displace it except for the action of random influences.

Different data purposes place different requirements on measurement accuracy. Crash recorder data presumably are more accurate than other indices of collision severity, such as the measured vehicle deformation or the Vehicle Damage Index (VDI). Whether such accuracy is required depends on the type of study. For many purposes, plan view photographs of the case and struck vehicle would be a significant improvement over VDI, as they would allow for an energy-derived calculation of severity.

When comparing injury outcome between accident cases with, as compared to without, a side guard beam, for example, we would want to control for collision severity because the degree of injury is correlated with collision severity. The control could be effected either mathematically or by partitioning the sample of cases in groups of equal collision severity. Controlling on collision severity will do two things: increase the efficiency of the comparison and eliminate the bias that results from fortuitous concentration of milder collision cases among one or another of the groups under comparison.

Because the degree of injury depends on many factors other than impact severity --such as restraint use, occupant age, and adventitious

posture -- the correlation of injury with the collision severity control variable is necessarily going to be less than perfect. As a result, increases in the precision of measurement of collision severity will not proportionally improve the efficiency of making the comparison when using it as a control variable. So, it is not so important to have high precision when doing routine accident comparison studies. The crash recorder has a different utility, and its evaluation should be based on other considerations.

Crash Recorder Use

A crash recorder will have utility for at least three types of studies. The first, as already mentioned above, is to provide correct normative information about such things as -- and particularly for correctly establishing -- the occupant exposure in terms of the probability distribution of collision speeds. To make such a determination requires a research project to be defined with this as its objective; the project could be based on the crash recorder as a particular tool of unusual usefulness. The research project could terminate when the determination has been made. Since the accuracy provided by the crash recorder is not essential for the kind of data-adjusting purposes described in the paragraph above -- i.e., in order to provide a control variable for accident case comparisons -- it would not be needed as a permanent part of a national data collection program. It should be viewed primarily as a research tool used for fairly particular purposes in a particular research program, more than an instrument for general accident investigation.

Another use for the crash recorder would be in research programs for establishing human tolerance to impact and to aid in establishing dynamic specifications for impact test devices. Thus, crash recorder data could be used as inputs in the programming of experimental crash tests or computer-simulated tests. These studies would determine the design characteristics needed in the test devices (e.g., crash dummies) so they would yield test

readings comparable to those experienced by actual accident victims. This kind of research requires data that are in dynamic physical form -- not rating scale indexes or qualitative descriptions. This usage of crash recorders would be contained within a research program designed to that end, and except for considerations of administrative efficiency, not be an intrinsic part of the national accident data collection system.

Still another useful purpose for the crash recorder would be to calibrate or to improve the more subjectively determined indexes which are now commonly used in accident investigation. Again, once that calibration has been effected, there would be no on-going necessity for the crash recorder.

Other Data Needs

There are two other areas of safety evaluation to which there has been inadequate attention. The first is to measure the overlapping and interactive effect of different safety requirements: e.g., strength of door fixtures and occupant restraints. Some safety evaluations, carried out in different studies, can count the same persons as being saved more than once by different means in each study, so that the total of the saved casualties might even exceed the population at risk. Our own studies have had this problem.

But even more significant is the almost total lack of information regarding the safety benefit in the 100-series federal standards. The whole concept of accident causation and avoidance needs to be clarified: to date it has been expressed more figuratively than in quantitative terms which will relate to vehicle design. Lack of good ideas in this area suggests that a conceptual breakthrough must be made before we are able to properly attribute that part of causation/reduction to the vehicle and its design, separate from the mediating influence of the driver and of the roadway, and so cost-

effective countermeasures can be imposed at the right place in the system for each aspect of accident causation, and in such a way -- and this is crucial -- that the specific effect can be evaluated, both prospectively and retrospectively, in accident data.

Procedures for establishing the safety effectiveness of both the current and proposed 100-series standards should be a major research challenge to the government and industry in the years to come. Current government accident avoidance research emphasis is to experimentally compare different vehicles on arbitrary control tasks. But programs of a different type are also needed, programs that will define measures of accident avoidance performance and then from that establish minimum criterion levels for performance, but the kind of performance that can be validated by accident statistics in the long run. For example, the effectiveness of existing braking and handling capability has not been definitely established in a real world context, much less the need for any changes. This is admittedly a difficult area in which to do research; there are very difficult conceptual problems. It is here, especially, that an interdisciplinary approach is needed.

Need for Greater Quantity of Data

Over the years, the Safety Administration has done an admirable job of developing in-depth studies (referred to as multidisciplinary accident investigations) of limited numbers of accidents, providing some information on how effectively certain designs may be functioning in specific instances. On the other hand, these special studies have not adequately revealed from a national viewpoint safety effectiveness on a representative basis. Thus, the accident teams which are employed for these in-depth studies can usually give a reasonably accurate description of any one accident -- and sometimes its causes or at least the causes of the injuries -- but they are not satisfying

our current pressing need for a comprehensive estimate of the nationwide accident picture.

A detailed and highly precise description of any one accident cannot by itself reveal where the overall priorities lic. There are three reasons why accident data must be collected in great quantity: First, there is considerable variability in the injury resulting from accidents that are, on the surface, similar; second, some accident features are quite infrequent and thus comparisons are often based on so little data they are unreliable; and third, we have to account for so many factors which can affect the outcome of each accident.

The first of these reasons -- variability in injury among similar accidents -- is seen when some people can get out of a total wreck and walk away with only minor injuries while in other crashes people sometimes die even though the car is so little damaged it can be driven away. A great number of crashes must be examined so that the entire range of injuries in any one type of crash can be accounted for.

secondly, certain events are relatively rare because most accidents are of comparatively low intensity and the injuries are of correspondingly low grade. It has been common to combine the counts of severely injured cases with the counts of fatalities in order to get a large enough total count to allow reliable comparisons to be made. Furthermore, some factors of interest -- such as restraints -- have had a relatively low rate of usage so not many cases have been available for investigation. It was only until B. J. Campbell, at North Carolina, was able to examine a few hundred thousand cases that he could find enough applicable ones to reliably detect the profound effect of the lap belt on the fatality rate -- as distinguished from its effect on the rate of severe injury or the rate of combined severe-plus-fatality. The base

fatality rate is quite a bit less than one percent; he found an overall 70 percent reduction in that rate in the lap-belted cases.

The third reason for needing a lot of data is the presence of numerous variables which affect the accident. The art of doing research and arriving at findings and conclusions about any aspect of accident or injury prevention is still fairly experimental. It is experimental because we do not have unequivocal, established scientific methods to cope with the present accident data. The reason for this is most of it fails to satisfy the basic requirements of analysis: that comparisons be made on an "all else equal" basis. By "all else equal" I mean that conclusions about the effectiveness of, say, the side guard beam must be made on data from crashes involving the same kind of vehicles in the same kind of trajectory with the same kind of people at risk, etc. However, given the diversity of vehicle models, it takes a lot of accident chasing to find enough crashes of the same type, of the same severity, and with the same type of vehicles and drivers, etc. -- that is, in which all else is equal. Mathematical adjustment of the data can take care of some confounding of variables in the data, but to be confident a considerable degree of representativeness in the original data is still needed.

Not the least consideration for achieving the proper representativeness of data is that there should be standardized definitions and protocols used by all the investigating agencies. Since a future investigator will query the data file as a microcosm of the universe of accidents, it would be most disagreeable that cases which are essentially similar were described in the same file differently only because the data were collected by different agencies using their own interpretations.

Need for Scientific Sampling

Not only is an increased quantity of data required but the sampling of the accident universe must be by sophisticated protocol. The last of the three reasons given above implies the need for a disciplined approach to the data, to avoid ending up with data which are biased in the factors underlying them. That requires a scientific approach to data collection, not just pouring more dollars into it and cranking Up the administrative machine to get a bigger program going but doing it in the same old way. Data gathering programs must be designed by the same people as will design the analyses that will be applied to the data. No less expertise than the Census Bureau applies, or the Gallup Poll, will suffice. Fortunately, NHTSA has been bringing in very competent people of late, people who know that a data collection scheme must be designed from the start with the method of analysis of the resulting data a key determiner of how the data should be gathered.

It is the Government Who Should Collect Data

Mass accident data acquisition, processing, analysis, and broad scale distribution requires great effort and much resource. Only the federal government has the necessary resource and easy access to the agencies which can supply information. Furthermore, it seems that it is the responsibility of the federal government to assemble data which will allow an accurate public review of the real dimensions of the crash and injury problem on our highways.

We appreciate the difficulty of developing and implementing a large scale, comprehensive plan for the acquisition of detailed data on motor vehicle related injuries and fatalities. We are aware that the Safety Administration has over the past several years developed and implemented a portion of such a plan which is related to fatalities. This effort has resulted in what is known

as the "Fatality Analysis File." We believe that data from most of the 50 states is going into that file and are hopeful that all interested parties will have access to that file in order that we all may comprehend the true and detailed dimensions of the fatality problem in the United States.

The Safety Administration has also requested funds for a large scale field survey of automobile accidents in which crash recorders would be employed. The data from this program is equally important to that from the Fatality Analysis File and would provide an accurate determination of the crash speeds at which the several levels of injury and fatality occur and can be employed as a basis for defining the performance levels needed in crashworthiness standards. We support a crash recorder program.

Certain fundamental questions cannot be answered without first having an adequate base of public data: What do we really know about the need for increased performance -- increased performance on the types of test criteria in the rules -- based on what is happening out there on the highway? What will be the effect on injury at lower speed levels when systems designed for a high speed compliance test are used? What are the proper speed levels to target for? While accident data are important, they are of course insufficient in themselves; other questions must still be considered: Can we mass produce these cars to provide such protection at reasonable cost? Should we approach an increased performance level in one massive jump or would we be better served to work toward it incrementally? What lead times are required to achieve these goals? These are obvious questions that should be considered before such rules are proposed.

In summary, we believe it is necesary to greatly expand accident data collection, in a well-disciplined scientifically devised program. Crash recorders cannot supplant an accident investigation program. Crash recorders will

be most useful in research projects whose ends specifically require the dynamics information which only such a tool can provide rather than in general data collection programs. There is a great challenge to undertake new studies of need in the accident avoidance area; indeed, new concepts, of pragmatic utility and based on what is actually happening on the roads, are needed in order to get a grasp on the whole issue of vehicle control and its relation to accidents. It is the government which has the responsibility and the resources for carrying out such programs.

APPENDIX K

# POSITION STATEMENT ON AN EX¼AMDED. LOW-COST NATIONAL ACCIDENT DATA COLLECTION PROGRAM

J.R. Cromack
B.J. Campbell
L. Patrick
B. O'Neill

February 7, 1975

# POSITION STATEMENT
# ON
# AN EXPANDED, LOW-COST NATIONAL ACCIDENT DATA COLLECTION PROGRAM

February 7, 1975

J. Robert cromack~ Southwest Research Institute:
B. J. Campbell. Highway Safety Research Center,
University of North Carolina; Lawrence Patrick,
Wayne State University; Brian O'Neill, Insurance
Institute for Highway Safety.

present real-world accident data have some deficiencies and limitations for both researchers and policymakers. Despite these limitations, much progress has been made on the basis of these data and useful information will continue to be obtained from these sources. However, much can and should be done to improve real-world accident data.

One major contribution would be the development of a large scale accident data base, possibly modeled on the data base developed at the Highway Safety Research Center of the University of North Carolina. This would require the upgrading of police accident reporting in a number of states and combining the data into a single base that could be assessed both by researchers and policymakers. Ideally, real-world accident data in such a base should include a measure, or measures, of both crash and injury severity.

At the present time the only available measure of crash severity is obtainable from the vehicle deformation or crush appropriately defined in relation to the manner of damage. Crash severities derived from vehicle deformation or crush can, however, only be compared among vehicles of the same make and model. It is possible that future research will enable the grouping of similar types and styles of vehicles with respect to crash

severities so derived, but at present there are no strong objective data to support such comparisons. Additional controlled laboratory type experimentation is needed to verify crash severity measures obtained from vehicle deformation or crash.

Meanwhile there are additional descriptors of real-world accidents that could be valuable to both researchers and policymakers. Crash recorders could provide such additional data. It seems likely that sophisticated recorders will continue to be too expensive to be deployed in the very large numbers needed to substantially augment present real-world data. Serious efforts should be devoted towards the development and large scale deployment of very inexpensive crash recorders that are designed to record a small number of Parameters that can be related to the severity of the crash.

The present measures of injury severity obtained from police accident reports are far from satisfactory and considerable efforts should also be devoted to upgrading these measures. Ideally, injuries should be classified either by the Abbreviated Injury Scale and its derivatives such as the Injury Severity Score or other appropriate injury scales.

A better understanding Of the nature and effect of traffic accidents can result from an expanded low Cost, well planned National Accident Data Collection program. The increased availability of data so derived will provide a higher confidence in the results derived from analysis of these data. It should be a major goal of such an effort to investigate the correlation between injury and damage, a topic presently not addressed due to inadequate data, but one that promises Significant Clarification to the problem of injury causation.

SOUTHWEST RESEARCH INSTITUTE

8500 CULEBRA ROAD • POST OFFICE DRAWER 28510 . SAN ANTONIO, TEXAS 78284

February 7, 1975

Mr. Howard P. Gates, Jr.
Economics and Science Planning
1200 - 18th Street, NW
Washington, D. C. 20036

Dear Mr. Gates:

Enclosed is the approximate concensus of the persons working on the assigned Issue No. 2. In the interest of time, I am sending you this document without final approval from each of the members. They will, however, receive copies of this letter and should they object too strenuously to any of the final changes or corrections, I feel certain you will hear from them.

In all fairness to them, I must state that I added the last paragraph based on my own convictions. It probably represents (at least in general) their views but this is the major divergence from the last draft position statement that was circulated. Incidentally, Larry Patrick did not have an opportunity to comment on the position statement after making several original contributions at our meeting on January 17.

None of the participants indicated an intention to take a position on Federal funding or inducements. Again, it was a pleasure to work with you and the other individuals at the workshop. I look forward to future meetings.

Sincerely,

Bob Cromack

J. Robert Cromack, Manager
Vehicle Safety Section
Department of Special Projects
Automotive Research Division

JRC:mr
Enclosure
cc: Lawrence Patrick
B. J. Campbell
Brian O'Neill

SAN ANTONIO, HOUSTON, CORPUS CHRISTI, TEXAS, AND WASHINGTON, D.

## POSITION STATEMENT - ISSUE 2

Present real-world accident data have some deficiencies and limitations for both researchers and policymakers. Despite these limitations, much progress has been made on the basis of these data and useful information will continue to be obtained from these sources. However, much can and should be done to improve real-world accident data.

One major contribution would be the development of a large scale accident data base, possibly modeled on the data base developed at the Highway Safety Research Center olf the University of North Carolina. This would require the upgrading of police accident reporting in a number of states and combining the data into a single base that could be assessed both by researchers and policymakers. Ideally, real-world accident data in such a base should include a measure, or measures, of both crash and injury severity.

At the present time the only available measure of crash severity is obtainable from the vehicle deformation or crush appropriately defined in relation to the manner of damage. Crash severities derived from vehicle deformation or crush can, however, only be compared among vehicles of the same make and model. It is possible that future research will enable the grouping of similar types and styles of vehicles with respect to crash severities so derived, but at present there are no strong objective data to support such comparisons. Additional controlled laboratory type experimentation is needed to verify crash severity measures obtained from vehicle deformation or crush.

Meanwhile there are additional descriptors of real-world accidents that could be valuable to both researchers and policymakers. Crash recorders could provide such additional data. It seems likely that sophisticated recorders will continue to be too expensive to be deployed in the very large numbers needed to substantially augment present real-world data. Serious efforts should be devoted towards the development and large scale deployment of very inexpensive crash recorders that are designed to record a small number of parameters that can be related to the severity of the crash.

The present measures of injury severity obtained from police accident reports are far from satisfactory and considerable efforts should als o be devoted to upgrading these measures. Ideally, injuries should be classified either by the Abbretiated Injury Scale and its derivatives such as the Injury Severity Score or other appropriate injury scales.

A better understanding of the nature and effect of traffic accidents can result from an expanded low cost, well planned National Accident Data Collection program. The increased availability of data so derived will provide a higher confidence in the results derived from analysis of these data. It should be a major goal of such an effort to investigate the correlation between injury and damage, a topic presently not addressed due to inadequate data, but one that promises significant clarification to the problem of injury causation.

APPENDIX L

RESOLUTION
OF
THE NATIONAL MOTOR VEHICLE SAFETY ADVISORY COUNCIL

November 14, 1974

A Resolution of the

**National Motor Vehicle Safety Advisory Council**
**(A body advisory to the Secretary of Transportation)**

November 14, 1974

**Whereas, analysis of the cost/benefits of revising standard 208 to require passive restraints has produced limited field evidence of the life saving value of passive restraints, including air bags; and**

Whereas, the analysis indicates that mathematical projection and tests with dummies do not predict with sufficient accuracy the potential value of these restraints in actual use; and

**Whereas,** there is likelihood that indignation over installation of passive restraints may eventually result in public pressure for the removal 'of such restraints after huge investment is made in them, as in the case of the seat belt interlock; and

Whereas, the mandated addition of yet another costly feature to new automobiles would be a financial hardship to the American consumer who must depend on automobiles for transportation; therefore be it

RESOLVED, that this Council recommends that the Secretary make a concerted effort to come to an agreement with industry on a plan that would result in increased passive restraint usage on the road and defer a call for rulemaking with respect to passive restraints until such time as further actual experience with them on the highways proves that they will reduce deaths and injuries.

APPENDIX M

# THE NEED FOR STANDARDIZATION IN REPORTING COLLISION DAMAGE AND INJURY IN TRAFFIC ACCIDENTS

J. Robert Cromack

January 16 & 17, 1975

# THE NEED FOR STANDARDIZATION IN REPORTING COLLISION DAMAGE AND INJURY IN TRAFFIC ACCIDENTS

PRESENTED BY J. ROBERT CROMACK AT THE COLLISION DATA WORKSHOP CONFERENCE, ROSSLYN (ARLINGTON) VIRGINIA, JANUARY 16 AND 17, 1975.

THE IMPORTANCE OF COLLECTING ACCURATE AND MEANINGFUL ACCIDENT DATA CAN BE MEASURED ONLY IN TERMS OF THE NEED FOR OR BENEFITS DERIVED FROM WHAT WE LEARN FROM ANALYSES OF THESE DATA. IF THE DATA ARE SELECTED FROM AN ATYPICAL SAMPLE OF THE GENERAL POPULATION, THEN THE CONCULSIONS WE DRAW FROM STUDYING THEM ARE APPLICABLE ONLY TO THE SAMPLE. SIMILARLY, IF THE DATA ACQUIRED DO NOT ENTIRELY SATISFLY THE NEEDS OF THE STUDY, THEN ONCE AGAIN WE MAY BE UNABLE TO DRAW APPROPRIATE CONCLUSIONS. FOR EXAMPLE, QUESTIONS CONCERNING INJURY CAUSATION CANNOT BE ANSWERED WHEN INJURY REPORTS ARE STRATIFIED SIMPLY ACCORDING TO THE CATEGORIES "PROBABLY NOT INJURED," "PROBABLY INJURED," OR "KILLED."

CONCLUSIONS EVOLVED FROM SUCH STUDIES MAY LEAD US TO MAKE INAPPROPRIATE RECOMMENDATIONS OR FORMULATE INVALID REGULATIONS. UNFORTUNATELY, EXCEPT FOR CERTAIN SPECIALIZED STUDIES IN WHICH DATA WERE SPECIFICALLY GATHERED FOR THE PURPOSE, MOST DATA DERIVED FROM STATE AND LOCAL POLICE RECORDS NOW COMPILED IN THE MAJOR ACCIDENT DATA FILES, HAVE MANY OF THESE DEFICIENCIES OF INCOMPLETENESS, INACCURACY, AND LACK OF COORDINATION OF ELEMENTS OF INTEREST.

HOW, THEN CAN WE AVOID THESE SNARES AND DELUSIONS. HOW, FOR EXAMPLE, CAN CRASH INDUCED INJURIES IN THE FIELD BE RELIABLY COMPARED WITH LABORATORY SIMULATED INJURIES TO CADAVERS AND ANIMALS IF THE FIELD INJURY DATA ARE EITHER POORLY DOCUMENTED OR ENTIRELY LACKING. I DON'T MEAN TO IMPLY THAT THERE ARE NOT SOME GOOD DATA COLLECTED IN THE FIELD, BUT UNFORTUNATELY, THERE ARE GENERALLY NOT ENOUGH GOOD CURRENT DATA COLLECTED TO BE STATISTICALLY USEFUL. THE ANSWER, OF COURSE, IS TO COLLECT MORE AND BETTER DATA.

ONE MOST IMPORTANT AREA OF INTEREST TO ALL OF US WHO HAVE BEEN WORKING TO REDUCE INJURY SEVERITY IN ACCIDENTS IS THE RELATIONSHIP OF VEHICLE

IMPACT INTENSITY TO RESULTING INJURY TO OCCUPANTS. RIGHT NOW MASS ACCIDENT DATA ARE COLLECTED ALMOST ENTIRELY BY THE POLICEMAN CALLED TO THE SCENE OF THE ACCIDENT. SUMMARIES OF THESE POLICE DATA TELL US THAT THERE WERE "X" NUMBER OF ACCIDENTS, "Y" NUMBER OF INJURIES, "Z" FATALITIES—THEY SAY THAT: SO MANY ACCIDENTS HAPPENED AT NIGHT, ON A RURAL ROAD, THAT WAS UNDER CONSTRUCTION, AND THAT THE VEHICLE WAS TOWED FROM THE SCENE BY ALVARADO'S WRECKER SERVICE — ALL VERY INTERESTING, BUT WHAT KIND OF A PICTURE DO THESE ELEMENTS OF INTELLIGENCE GIVE US OF THE ACCIDENT SEVERITY. IN FACT, DO THEY PROVIDE ANY MEASURE OR MEASURES OF ACCIDENT SEVERITY.

I BELIEVE THAT TWO VERY IMPORTANT MEASURES OF THE SEVERITY OF AN ACCIDENT ARE: THE EXTENT OF PERSONAL INJURY AND VEHICLE DAMAGE. INJURY AND DAMAGE CAN BE QUANTIFIED AND WHEN THESE QUANTIFICATIONS ARE COMBINED SPECIFICALLY WITH OTHER ROUTINE INFORMATION NORMALLY COLLECTED AT THE POLICE INVESTIGATION LEVEL, A VAST ARRAY OF USEFUL INFORMATION CAN BE EVOLVED.

WHILE VARIOUS SCALES, INDICES AND METHODS HAVE BEEN USED TO QUANTIFY OCCUPANT INJURY AND VEHICLE DAMAGE, THE TWO MOST VERSATILE AND, IN MY OPINION, VALUABLE, ARE THE ABBREVIATED INJURY SCALE (1) (FOR INJURY QUANTIFICATION) AND THE VEHICLE DEFORMATION INDEX (2) (FOR VEHICLE DAMAGE SPECIFICATION).

THE ABBREVIATED INJURY SCALE, OR AIS AS IT IS OFTEN REFERRED TO, IS A METHOD DEVELOPED BY THE AMERCIAN MEDICAL ASSOCIATION TO DESCRIBE PERSONAL INJURY. NUMERICAL INJURY CODES ARE APPLIED TO SPECIFIC TRAUMA IN SPECIFIC ANATOMICAL REGIONS; FOR EXAMPLE, A COMPOUND FRACTURE OF THE LEFT HUMERUS MIGHT RATE AN AIS LEVEL 3. THE INCREASING NUMBERS FROM ZERO TO SIX INDICATE INCREASING SEVERITY OF THE SPECIFIC LESION. THESE CODES WERE DEVELOPED BY PHYSICIANS WHO CONSIDERED SUCH FACTORS AS: ENERGY DISSIPATION AS A CAUSE OF THE TRAUMA, THREAT TO LIFE RESULTING FROM THE TRAUMA, POSSIBILITY OF PERMANENT IMPAIRMENT, THE TREATMENT PERIOD REQUIRED FOR HEALING, AND THE INCIDENCE OF SUCH INJURIES OCCURRING IN THE ROUTINE TREATMENT OF TRAUMA

FROM ROUTINE TRAFFIC ACCIDENTS AS BASED ON THE EXPERIENCE OF THE ATTENDING PHYSICIANS. THE AIS CODES ARE ASSIGNED ONLY BY MEDICAL AND PARA-MEDICAL PERSONNEL.

VEHICLE DAMAGE IS QUANTIFIED USING THE VEHICLE DEFORMATION INDEX, CALLED THE VDI, DEVELOPED BY THE SOCIETY OF AUTOMOTIVE ENGINEERS. THE VDI IS A COMPREHENSIVE DAMAGE DESCRIPTION TECHNIQUE IN WHICH A SIX COMPONENT, SEVEN CHARACTER ALPHANUMERIC CODE IS USED TO DESCRIBE: THE GENERAL DIRECTION OF FORCE, GENERAL AREA OF DEFORMATION, SPECIFIC HORIZONTAL AND VERTICAL AREAS OF DAMAGE, THE TYPE OF DAMAGE DISTRIBUTION, AND THE EXTENT OF DAMAGE. A VDI CODE OF 12FCEN2 INDICATES THAT THE VEHICLE RAN HEAD-ON INTO A NARROW OBJECT, SUCH AS POLE, CAUSING CENTRAL DAMAGE TO THE FRONT OF THE VEHICLE AND THE CONTACT DAMAGE EXTENDED BACK APPROXIMATELY ONE-THIRD THE LENGTH OF THE HOOD.

USE OF THE VEHICLE DEFORMATION INDEX BY A DIVERSE GROUP OF INVESTIGATORS WITH ONLY MINIMAL TRAINING, WAS SHOWN TO BE FEASIBLE IN A STUDY OF 520 ACCIDENT CASES COLLECTED BY 33 TEAMS REPRESENTING NINE EUROPEAN AND NORTH AMERICAN NATIONS IN THE PILOT STUDY ON ROAD SAFETY[3,4]. IN THIS CONTROLLED STUDY, ACCIDENT INVESTIGATORS CORRECTLY APPLIED THE VDI TO DAMAGED VEHICLES IN APPROXIMATELY THREE-FOURTHS OF THE CASES. LESS THAN 5% OF THE CASES CONTAINED INVALID INFORMATION. THE AUTHORS OF THE REPORT CONCLUDED THAT THE VDI COULD BE APPLIED EFFECTIVELY AT THE POLICE LEVEL 90 TO 95% OF THE TIME IN REGULAR USE.[5]

THE DAMAGE AND INJURY SCALES WERE STUDIED FURTHER IN A SPECIAL PROGRAM IN BEXAR COUNTY, TEXAS, DURING THE SIX-MONTH PERIOD ENDING MAY 30, 1973. THE AUTOMATED DATA FILE FROM THIS STUDY CONTAINS INFORMATION ON NEARLY 5500 ACCIDENTS INVOLVING OVER 10,000 VEHICLES AND 15,000 OCCUPANTS. THE CORRELATION BETWEEN INJURY SEVERITY AND VEHICLE DAMAGE WAS STUDIED AND GROSS PATTERNS WERE EVIDENT WHEN THE TWO VARIABLES WERE COMPARED ON SELECTED BASIS.[6,7]

WITH REASONABLE ASSURANCE, INJURY SEVERITY COULD BE PREDICTED ON THE BASIS OF OBSERVED DAMAGE EXTENT. I HASTEN TO ADD, HOWEVER, THAT THIS STATEMENT MUST BE ACCEPTED IN VIEW OF MANY OTHER QUALIFYING FACTORS. GROSS STATISTICS WERE REPORTED BECAUSE WE SIMPLY DID NOT HAVE A DATA FILE OF SUFFICIENT SIZE TO PRODUCE STATISTICALLY SIGNIFICANT RESULTS IN ANY DEGREE OF REFINEMENT.

A RECOMMENDATION OF THIS STUDY, WHICH I NOW PROPOSE FOR CONSIDERATION OF THIS GROUP IS: USING NOW OBTAINABLE LARGER QUANTITIES OF ACCIDENT DATA, CONDUCT FURTHER STUDIES AIMED AT REFINING THE INJURY SEVERITY PREDICTION POTENTIAL OF THE VDI COLLISION DAMAGE CLASSIFICATION TECHNIQUE USED IN CONJUNCTION WITH THE ABBREVIATED INJURY SCALE; PROMULGATE THE VEHICLE DEFORMATION INDEX AS THE STANDARD COLLISION DAMAGE CLASSIFICATION TOOL FOR USE BY POLICE IN REPORTING VEHICLE DAMAGE IN TRAFFIC ACCIDENTS; INTRODUCE PROGRAMS WHICH WOULD PERMIT UTILIZATION OF THE ABBREVIATED INJURY SCALE (ASSIGNED ONLY BY MEDICAL AND PARA-MEDICAL PERSONNEL) IN CONNECTION WITH POLICE REPORTS IN MAJOR TRAUMA TREATMENT CENTERS.

APPENDIX N

# REGULATORY RULEMAKING BASED ON LESS THAN TOTAL INFORMATION

David Morganstein
CENTER FOR AUTO SAFETY
and
L.A. Goldnuntz
ECONOMICS & SCIENCE PLANNING, INC.

February 21, 1975
(date of receipt)

# REGULATORY RULEMAKING

# BASED ON LESS THAN TOTAL INFORMATION

David Morganstein, Center for Auto Safety,
and L. A. Goldmuntz, Economics & Science Planning, Inc.

Received February 21, 1975

Estimation of the costs and benefits expected from regulatory programs is complicated by a lack of precise information. Several areas where a lack of knowledge exist are: the methods to be used by those regulated to meet the requirements; the efficacy of the methods chosen; the details which enter into the pricing effort of changes brought about by the regulation; alterations in the initial conditions which may occur over time, causing unpredictable variations in costs or benefits; the effectiveness of the regulation in achieving the desired benefits; and the impacts the regulation might have in other areas.

One subject not frequently addressed is the variation of the process to be regulated. If a population characteristic is time-varying, the potential benefits may be similarly varying. In such a situation, the possible conflicting conclusions that might be arrived at must be considered. There are well known tools, such as decision theory, which may provide a better conclusion than some undefined subjective process. Thus, the cost, the need or the value of additional data collection can be evaluated in light of its potential for clarifying the issues.

Nevertheless, governmental expectations are sufficiently high and the public demand sufficiently intense that programs may proceed even though complete information is unavailable or unattainable. After programs have been in place for some period, improvements

may be realized more slowly than initial expectations. Ensuing discussions are polarized around industry and the regulatory agency: Is industry using unnecessarily expensive methods, and not choosing methods most likely to meet the intent of the regulation? Or on the other hand, are bureaucrats acting to enlarge their domain or justify their existence as a regulator? Or is there a lack of communication between industry , the regulators and the public so that there is little understanding of the issues and therefore little progress in resolving them?

Advocates may reference controlled laboratory experiments to estimate the efficacy of a regulation. They argue that the learning process will improve the methods used to meet the intent of the regulation and lower costs. Cynics question the extent to which laboratory experiments represent the real world. When cynics argue that the introduction of a new technology has a price tag which will ultimately be paid by the public, the advocates counter that the withholding of such technology has its own price tag. Clearly, there are societal costs to be borne without the protection of the regulation, with inadequate regulation or with excessive regulation. These issues have no general answers but require analysis case by case at each stage of the development of the regulation.

Analyses of the complex issues can best be carried out by a number of independent professional sources working independently. These efforts should then be compared, and the analyses and reasons for proceeding or not proceeding with a suggested program should be subject to public scrutiny. The consumer is potentially victimized when information is in the hands of any one monolithic organization, be it a regulatory agency or an industry. The consumer may also tend to be victimized by oversimplified sensationalized commentary by either side to the debate.

The consumer has to rely on the different perspectives within society to accomplish the various analyses that expose the issues. We believe this pluralism can then lead to modifications of various points of view and perhaps lead to an eventual crystallization of the issues in a form that can be more readily understood by the public. At this point, it is essentially a public or political decision as to whether to proceed or not to proceed with any given regulatory program. The public interest is served best by having the issues fully explored from many points of view by many independent sources in estimating the potential costs and benefits of proposed regulatory programs.

APPENDIX O

AUTOMOBILE COLLISION DATA WORKSHOP:
AGENDA
SCHEDULED PRESENTATIONS
SALIENT RESIDuAL ISSUES

January 16 & 17, 1975

TENTATIVE AGENDA

AUTOMOBILE COLLISION DATA WORKSHOP

Part I. Data Requirements.

(a) Collision data needed for the design of crashworthy passenger cars including the restraint system, and to permit compliance testing; kinds of information, their relative importance, and precision required.

(b) Collision data needed for rational regulatory rulemaking; kinds and amounts of information, priorities, precision.

(c) Adequacy of the existing collision data base and the utility of data being gathered by current methods.

(d) Statistical requirements: rate at which data should be gathered to be timely in the environment of a temporally-varying car-design population; the data file size to assure statistical significance when divided into cells of interest; time to accrue the required data file as a function of sampling rate; statistical adequacy of current and proposed programs.

(e) Dollar-equivalent benefits of adequate data; costs of not having data or using incorrect data.

Part II. Data Gathering Techniques and Programs.

(f) Crash recorders: capabilities, costs and limitations of alternative designs and programs.

(g) Accident reporting: extent, accuracy, costs and limitations; potential and cost of improving reporting accuracy.

(h) utility, cost and limitations of computer crash simulation.

(i) Derivation of crash data statistics through correlation of laboratory crashes with real world experience; clinical investigations; adequacy, accuracy, cost and limitations of these approaches.

Part III. Public, Legal and Legislative Reactions.

(j) The potential impact of crash recorders on tort claim settlement.

(k) The reaction of public interest groups to alternate collision-data-gathering programs.

(l) The legislative history of collision data gathering proposals and programs.

## SCHEDULED PRESENTATIONS

## AUTOMOBILE COLLISION DATA WORKSHOP

January 16, 1975

### DATA REQUIREMENTS

"Mass Accident Data Acquisition and Why It's Needed",
John Versace, Ford Motor Company

"Inadequacy of Accident Data to Conduct Meaningful Research",
Robert Cromack, Southwest Research Institute

"Need for Better Crash Data",
Brian O'Neill, Insurance Institute for Highway Safety

"Collision Data Required to Improve and Evaluate Safety",
Lawrence Patrick, Wayne State University

"How Data Fits Into the Rulemaking Process",
James Hofferberth, National Highway Traffic Safety Administration.

"Adequacy and Limitations of Current Data",
Marie Eldridge, National Highway Traffic Safety Admin.

### DATA GATHERING TECHNIQUES AND PROGRAMS

"A Discussion of Data Gathering Systems",
Edwin Kidd, Calspan Corporation

"How to Make Crash Recorders Support Other Data Collection Programs"
B. J. Campbell, Highway Safety Research Center, U. of N. C.

"Crash Recorders: A Solution Seeking A Problem?"
James O'Day, Highway Safety Research Institute, U. of Mich.

"NHTSA Crash Recorders" ,
Lynn Bradford, National Highway Traffic Safety Administration

"Automotive Tape Recorder"
Charles Conlon, AVCO Systems Division

"All Solid State Triaxial Accelerometer for Crash Testing",
Louis Roberts, Transportation Systems Center

**Economics & Science Planning**

1200 18th Street, N.W. Washington, D.C. 20036 Telephone: 202-223-8444

**January 20, 1975**

AUTOMOBILE COLLISION DATA WORKSHOP
January 16-17, 1975

A number of major issues surfaced at the January 16, 1975 Automobile Collision Data Workshop. The following people have agreed to write brief position papers on these issues and to forward them to Economics & Science Planning, Inc., before February 1, 1975:

ISSUE 1

Estimate the potential societal cost of not having better accident data than available from current resources.

From the point of view of the automobile manufacturer: (Working separately)

- John Versace, Ford Motor Co.
- Richard Wilson, General Motors Corp.

From the point of view of the regulator:

- James Hofferberth, NHTSA

ISSUE 2

What are the advantages of an expanded low cost national accident data collection program that might provide 600,000 to a million reports per year? How would such a data program be organized? Are there any models for such a data program? What Federal funding or inducements would be appropriate to achieve it?

(Working together)

- Brian O'Neill, Insurance Institute for Highway Safety
- Lawrence Patrick, Wayne State University
- B. J. Campbell, Highway Safety Research Center
- Robert Cromack, Southwest Research Institute

Cable . . . ESPINC
Telex . . . 248482

ISSUES TO BE CONSIDERED BY THE OFFICE OF TECHNOLOGY ASSESSMENT IN ITS RESPONSE TO THE HOUSE APPROPRIATIONS COMMITTEE ARE THE FOLLOWING:

1. How much has NHTSA spent in each of the past three years to gather accident data? Is that data sufficient, or is further data on the characteristics of automobile collisions necessary for effective NHTSA standards-setting? If the existing data base is inadequate; in what ways is it inadequate?

2. An evaluation of the type of data being produced by existing crash recorders and an explanation of how this data is being used by NHTSA should be conducted.

3* If the data base is inadequate, how might an adequate data base be obtained and what are the consequences associated with obtaining the data in different ways (including the possibility of not obtaining the necessary data)? The cost effectiveness of the crash recorder and the crash impact approaches proposed by NHTSA should be examined.

4. Secondary consequences of implementing these or other programs should be identified and evaluated. Examples of these secondary consequences include legal questions associated with the existence of actual physical data from an accident and the potential value (to driver training programs) of a knowledge base concerning how drivers actually respond in accident situations. For each type of approach investigated, the implementation costs to the Federal Government, industry and consumers should be identified.

## ISSUE 3

Define the role of crash recorders in capturing field data needed to evaluate and calibrate accident investigators reports, crash tests, and crash simulation.

* Gene Mannella, NHTSA
* James O'Day, Highway Safety Research Institute, University of Michigan
* Edwin Kidd, Calspan Corporation

## ISSUE 4

what is the statistical rationale for the number of recorders proposed for procurement and installation by NHTSA? Is the number appropriate to the calibration uses described in 3 above? injury and fatality prevention rulemaking? damageability rulemaking or assessment?

* Gene Mannella, NHTSA
* Don Mela, NHTSA

## ISSUE 5

Reliable data is sometimes unavailable to the extent desired when a regulatory action may seem to some to be desirable. What general policy guidelines if any can be developed to guide regulatory actions in an environment of imperfect data.

* David Morganstein, Center for Auto Safety
* Lawrence Goldmuntz, Economics and Science Planning

# APPENDIX P

LETTER FROM JOHN GARRETT
CALSPAN CORPORATION
AND
ATTACHMENT

March 12, 1975

Calspan

March 12, 1975

Air Mail

Dr. Lawrence Goldmuntz
Economics and Science Planning
1200 18th Street, N.W.
Washington, D.C. 20036

Dear Dr. Goldmuntz:

Following our telephone conversation about two weeks ago, I gathered some material on our use of police photography for estimation of vehicle damage severity and/or speed, as I had agreed. The material is attached to this letter. To provide background, and some additional detail, I have summarized relevant information below.

We first became concerned with the problem of assessing accident severity in our Automotive Crash Injury Research (ACIR) program in the early to mid-1950's. At that time, we developed an accident Severity Index (Attachment A) based on damage to the vehicle. The police provided interior and exterior photographs of the accident vehicle but the ratings were made by a small staff of trained Calspan (then Cornell) personnel. This procedure tended to minimize the inter-coder variability that would have resulted if thousands of police had rated the accidents. Also, it was not necessary to train police to code, but only to take the proper photographs. Thus, training costs were kept low.

Accuracy of ratings were further assured through the use of fairly extensive computer edit procedures. "Illegal" (impossible) codes resulted in a case being returned for checking. Consistency checks also were used, i.e., a case that was rated minor could not have severe overall damage to the car elsewhere or any damage to basic structure such as the chassis. Low probability events that were inconsistent with the severity also required a recheck of the case. Thus, a fatality in a case where the severity rating for the vehicle was minor, warranted a check. Some corrections were made automatically, but many errors required a recheck.

The reliability of rating procedures also was checked periodically by ACIR to ensure that rater variability was kept to a minimum. A copy of one report on this subject (Attachment B) is enclosed.

Bob Campbell later developed the TAD scale which is used by police in North Carolina and several other states. Here, all ratings are made by police in the field. Bob's studies have shown that they do rather well, but I think that I would prefer the additional control which our system provides.

The Collision Deformation Code (CDC) developed by G.M. generally succeeded the earlier systems for use by many researchers and the in-depth teams. In some ways this always seemed odd to me since the in-depth teams had measurements of the actual vehicle damage which were more accurate than the CDC. This scale clearly is too complicated for police use in the field. However, we have compared CDC ratings obtained by our personnel from police photographs with those obtained by an experienced invetigator rating the CDC from actual inspection and measurement of the vehicle. The results were quite good (Attachment C, pages 37-56) and we would have confidence in ratings provided by such a system. Again, ratings were made by a small staff of Calspan personnel with appropriate checks to maintain accuracy.

We later summarized available data from Calspan crash tests in a first attempt to develop an aid for estimating speed from vehicle damage (Attachment D). The amount of useful data was limited and the approach was dropped when additional inputs were not forthcoming.

Development of the SMAC program by Ray McHenry permitted accurate estimates of impact speeds, but requires such information as vehicle damage, point of impact and vehicle rest positions. Use of the Calvan simplifies the collection procedure for police and ensures accuracy. Ray is now working on a simplified version of the START program for SMAC which, it appears, may provide reasonably accurate speed estimates. A brief description appears in Attachment E.

Data collection cost was another point that we discussed. The cost of our most recent program to collect police photographs (last year) was approximately $5,000 for 1,200 cases. Costs include only purchase and processing of film. We have purchased relatively inexpensive Instamatic cameras ($20-25) for police use, with good results. Generally, one camera per car is needed.

In our discussion, you also mentioned the possible use of templates for measuring the vehicle damage photographed. We explored this, but it is quite difficult to do without an overhead shot of the vehicle or the use of photogrammetry. If we go that far, then I believe that the Calvan would be competitive in terms of cost and would provide far better data.

Dr. Lawrence Goldmuntz
March 12, 1975

This has become a rather lengthy letter with many attachments, but since I agree that the use of police photographs can provide good vehicle damage/speed data, I have tried to provide what useful information I can. It may still be sketchy for your purposes, however. If so, I will be pleased to provide any additional information that we have available,

Sincerely,

John

**John W. Garrett, Head**
**Accident Research Branch**
**Transportation Safety Department**

JWG:jem
Attachments

# ACIR INDEX OF ACCIDENT SEVERITY AND POTENTIAL SURVIVABILITY

## Introduction

The method of rating accident severity described here was developed for use in the Automotive Crash Injury Research program (ACIR) of Cornell Aeronautical Laboratory, Inc. Thus far, 31 states have participated in the program. In this study, accident data are reported by participating state police on a report form designed by ACIR. Police also provide photographs of the vehicle interior and exterior. Medical data are provided by the physicians who attend the injured victims.

The description of the rating system which follows was prepared for use by ACIR case analysts. In assessing vehicle damage, photographs of the car and pertinent information from the accident report form are used. The report form provides information concerning damage to various structural elements of the car, such as the chassis frame, engine or mounts, firewall, floor, etc., which are not always visible in photographs. This information is essential if an accurate appraisal of vehicle damage is to be made.

Although the ACIR accident data have been collected by thousands of police officers throughout the country, rating of accident severity and survivability has been performed by a few ACIR case analysts. Thus, variability in rating can be minimized and closer control maintained over the process.

## Discussion of Rating Method

The rating of accident severity and survivability represents an effort to classify accident-involved cars not only in terms of damage but in terms of the potential survival of occupants. Each car in an accident is evaluated and rated individually since severity and potential survivability often differ even for cars involved in the same accident. Data on which the evaluation is based are obtained from a series of interior and exterior photographs of the car, and from the accident report form which describes the accident and provides additional information on certain structural components: engine, engine mounts, chassis frame, firewall, front wheels, floor, etc. Accident severity and survivability are rated only when adequate photographs and sufficient accident information are available. When adequate photographs or accident data are not available, cases are classed as NAC (not able to classify).

The classification of accident severity and survivability requires an assessment of the type and amount of car damage, type of components affected, and the influence of this damage on potential survival of car occu-

pants. Accident severity and survivability are rated semi -independently although in fact they are inextricably related. Broadly speaking, accident severity is classified in terms of the type, extent and area (side, rear, etc. ) of the car damaged, whereas survivability is classified in terms of occupant environment, i.e., whether there is collapse or invasion of the compartment. Accident severity and survivability are not mutually exclusive categories, as is shown in the gross relationship between c a r damage, accident severity, and survivability.

In classifying accident severity a six-point scale (below) ranging from minor to extreme is used. In descriptive terms, damage ranges from denting and scratching of surface metal to complete disintegration or crushing of the car. Thus, the accident severity rating rises progressively as damage increases and more of the structural elements of the car are affected.

Accident Severity and Survivability Scale

| Car Damage | Accident Severity | Survivability y |
|---|---|---|
| Sheet Metal Damage | | |
| No damage to basic structure; no invasion of compartment. | Minor | Survivable |
| | Mode rate | Survivable |
| Structural elements progressively involved; compartment may, or may not, be invaded. | Moderately Severe | Survivable, Questionable or Partial |
| | Severe | Survivable, Questionable, or Partial |
| | Extremely Severe" | Survivable, Questionable, Partial, or Non -Survivable |
| Complete Destruction | Extreme | Non-Survivable |

When an accident is rated minor ***or*** mode rate in ***severity*** it is conside red survivable. Moderately severe, severe or extreme 1 y severe acci - dents may also be survivable, or survivability may be rated as questionable or partial. Extremely severe accidents may also be classified as non-survivable. Extreme accidents are always regarded as non- survivable be - cause they involve almost complete destruction of the car. A more detailed description of both accident severity and survivability is provided in the sections which follow.

Accident Severity

o Minor

Damage is most often confined to the sheet metal surface of the car although bumpers may be slightly dented, headlights or taillights broken, radiator grill bent or broken, ornamental molding torn free. When forces are applied to sheet metal, damage may be de scribed in such terms as "small dent", "slight deformation", scratches ", etc. Such damage is con side red minor whether a small or large area of the car is affected. Minor severity accidents never involve structural components of the car.

l Mode rate

Damage most often involves sheet metal, but such structures as bumpers, bumper guards, or radiator grill may be damaged. Sheet metal or grill damage may be described as "slight buckling", "pushed in' "crumpled", or "torn". For stronger components -- such as a steel burn - per -- descriptive phrases such as "large dent", "twisted", or "bent" might be used. In accidents of mode rate severity, structural components of the car are undamaged.

l Moderately Severe

Damage involves forces sufficiently great so that stronger structural elements as well as sheet metal are affected. Usually sheet metal begins to collapse and, depending on the area of impact, comer posts, center posts, or chassis frame may be deformed.

l Severe

Damage in this category always involves collapse or marked displacement of structural elements, as well as c rushing or telescoping of sheet metal. This grade of accident severity often involves penetration of compartment are as 'either as a result of direct impact, or as a result of displacement of other parts of the car due to impact or overturn.

- Extremely Severe

Damage to the impacted area in these accidents is very extensive. Structural elements and sheet metal in the affected areas are gene rally crushed . There is considerable telescoping of the impacted area, and there is usually some invasion or collapse of the compartment.

- Extreme

This category is reserved for accidents so severe that the automobile involved is almost completely demolished, and often is scarcely recognizable as an automobile. Damage may be de scribed as almost complete disintegration or crushing of the entire car. Photographs of extreme dam - age are not provide d in the figure illustrating accident severity because all damage beyond that illustrated for extremely severe is considered extreme.

Survivability

The concept of survivability is based on the assumption that survival is dependent on the compartment area remaining essentially intact. In rating survivability, it is recognized that other forms of protection -- interior redesign, padding, lap belt and harness, or even other devices as yet not available -- may be required in order to fully capitalize on the potential survivability afforded by the compartment. Without a reasonably intact environment, however, there is no assurance that occupants could survive even with other protective devices. The criteria used in determining survivability, there fore, are the degree of compartment collapse and its influence on the normal seated position areas, i. e. , whether there would be sufficient space for Survival if all seats had been occupied by persons seated in a normal, upright position, and all occupants had remained in their seats. In brief, whether the area surrounding each seat in the car could still hold an upright occupant.

Data concerning the actual fate of automobile occupants indicate that many occupants die in accidents that are relatively mild and, conversely, some occupants survive even when the car is demolished. Although all cars in the ACIR study contain at least one occupant, in classifying survivability the presence or absence of occupants, as well as the fate of those occupants actually present in the car, is ignored. In effect, the car is rated without considering the number of occupants or whether they lived or died. Thus, occupants may survive a non- survivable accident, or may die in a survivable accident.

A "survivable " rating signifies that the compartment (occupant are a) was essentially intact and that there was no c rushing or invasion of the compartment. As the compartment area collapses or is progressively invade d by surrounding structure, survivability may be classified as survivable, questionable, partial, or non- survivable. Survivability categories and the appropriate accident severity categories are described below.

- Survivable

When there is little or no invasion of the compartment area, survivability for all occupant areas is normally assumed. Minor and moderte accident severities, by definition, must be considered survivable. Moderately severe, severe, and extremely severe accidents may be survivable if there is little invasion of the compartment. An extreme accident (again, by definition) cannot be considered survivable. (Rated survivable: Front photographs - minor, moderate, moderately severe; Side - minor, moderate, moderately severe, severe; Rollover - minor, moderate, moderately severe.)

- Questionable Survivability

When the area surrounding one or more seated positions is somewhat compressed, but there is some doubt as to whether one or more normally seated persons could survive, survivability is considered questionable. This classification may be used only with moderately severe, severe, and extremely severe accidents. (Rated questionable survivability: Side photograph - severe; Rear - extremely severe; Rollover - severe. )

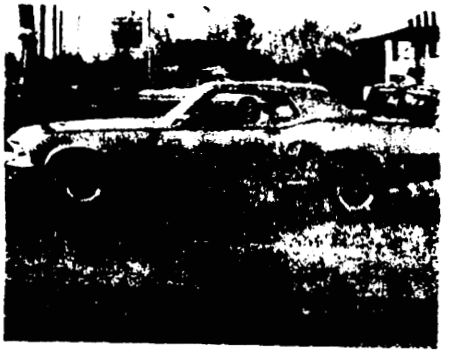
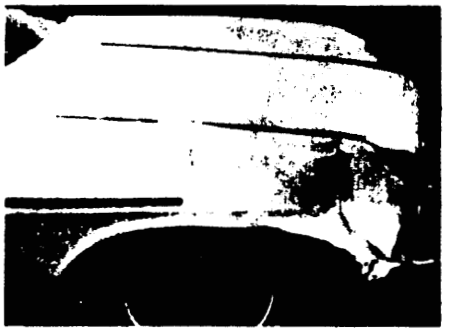
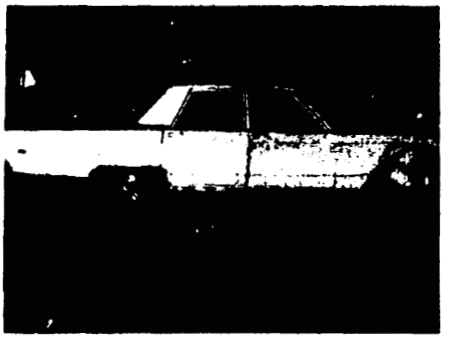
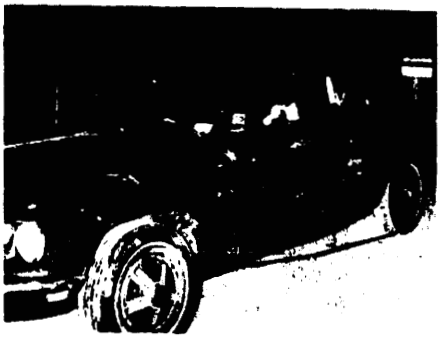
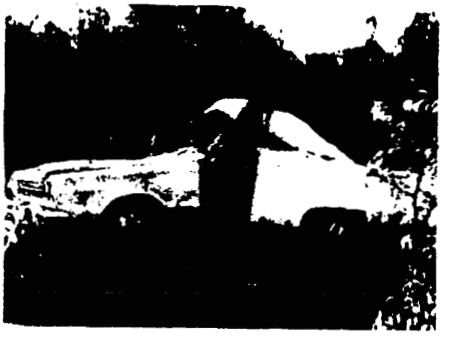
- Partially Survivable

This category is used when one or more, but not all) seated posi - tions are compressed to such a degree that it is considered non-survivable for a normally se ate d person. This classification may be used only with mode rate 1 y severe, severe, and extremely severe accidents. (Rated partially survivable: Side photograph - extremely Severe.)

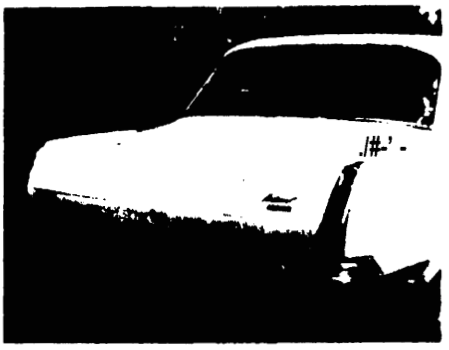
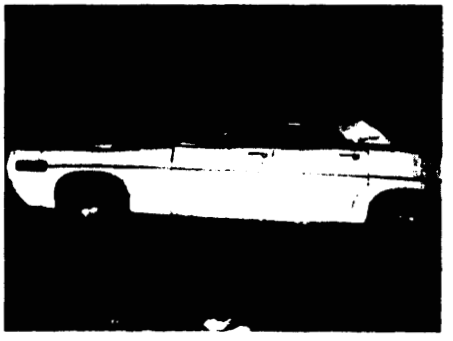
- Non-Survivable

When the entire compartment is compressed Or invaded to such an extent that there is insufficient room for an occupant seated upright in all the normal seating areas, the accident is considered non- survivable. Extremely severe accidents may be classified as non-survivable, and extreme accidents must be so classified. (Rated non-survivable: Front photograph - extremely severe; Rollover - extremely severe. )

| Accident Severity and Survivability | Column 69 |
|---|---|
| **Classification** | **Code** |
| Survivable | |
| Minor | 1 |
| Moderate | 2 |
| Moderately severe | 3 |
| Sever e | 4 |
| Extremely severe | 5 |
| Non- survivable | |
| Extremely severe | E |
| Extreme | F |
| Partially survivable | |
| Moderately severe | L |
| Sever e | M |
| Extremely severe | N |
| Questionable survivability | |
| Moderately severe | T |
| Severe | u |
| Extremely severe | v |
| Not Able to Classify | X |

| | FRONT | SIDE |
|---|---|---|
| MINOR | |  |
| MODERATE |  | |
| MODERATELY SEVERE |  |  |
| SEVERE | | |
| EXTREMELY SEVERE | |  |

# SEVERITY

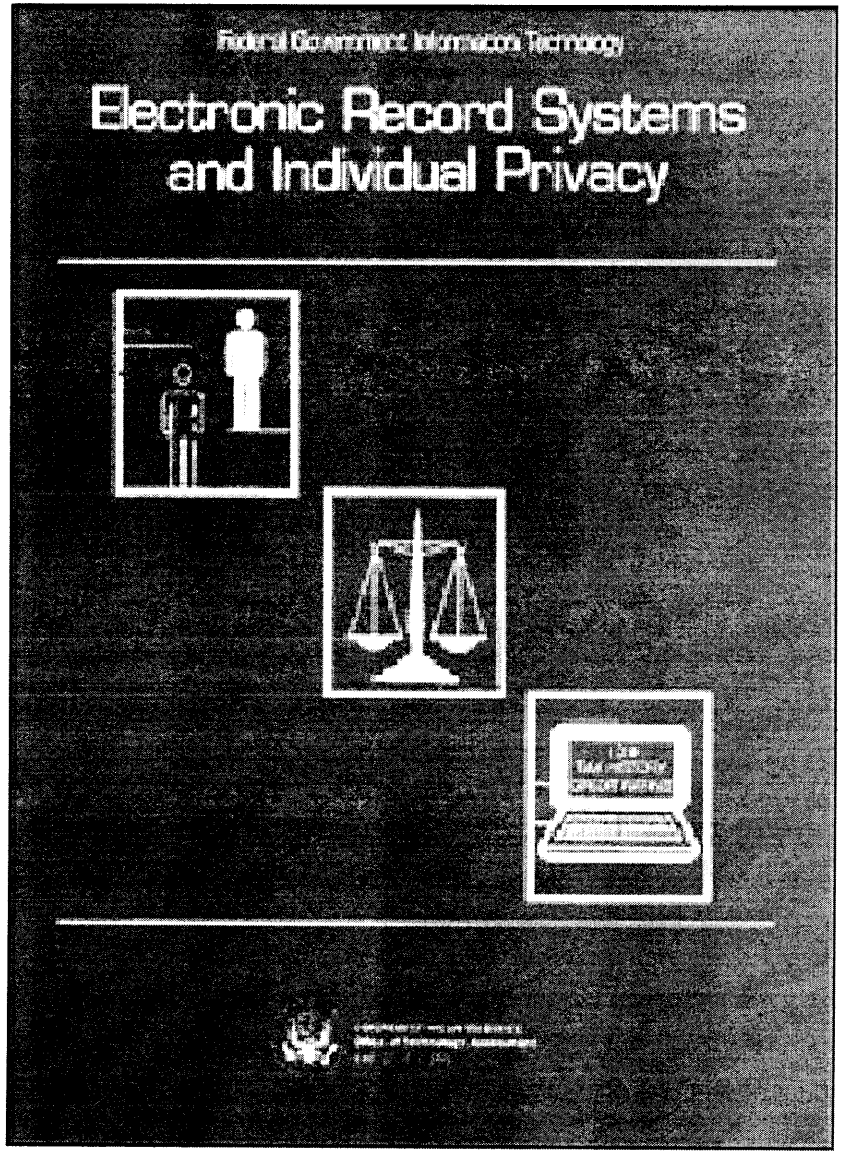
| REAR | ROLLOVER | |
|---|---|---|
|  | | MINOR |
| | | MODERATE |
| | | MODERATELY SEVERE |
| |  | SEVERE |
| | | EXTREMELY SEVERE |

ACCIDENT RESEARCH BRANCH
TRANSPORTATION RESEARCH
CORNELL AERONAUTICAL LABORATORY

*Electronic Record Systems and Individual Privacy*

June 1986

NTIS order #PB87-100335



Recommended Citation:

U.S. Congress, Office of Technology Assessment, *Federal Government Information Technology: Electronic Record Systems and Individual Privacy, OTA-CIT-296* (Washington, DC: U.S. Government Printing Office, June *1986).*

Library of Congress Catalog Card Number 86-600524

For sale by the Superintendent of Documents
U.S. Government Printing Office, Washington, DC 20402

# Foreword

Public policy on the protection of personal information collected, maintained, or disseminated by the Federal Government has been based on a balancing of the privacy of individual citizens versus management efficiency and law enforcement. New technological applications—such as the computerized matching of two or more sets of records, extensive electronic networking of diverse computerized record systems, and preparation of computer-based profiles on specific types of individuals—are challenging the existing statutory framework for balancing these interests.

This report addresses four major areas: 1) technological developments relevant to government record systems; 2) current and prospective Federal agency use of electronic record systems; 3) the interaction of technology and public law relevant to protecting privacy; and 4) possible policy actions that warrant congressional attention, including amendment of existing laws such as the Privacy Act of 1974 and establishment of new mechanisms such as a Data Protection Board or Privacy Protection Commission.

Prepared at the request of the Senate Committee on Governmental Affairs and the House Committee on the Judiciary, Subcommittee on Courts, Civil Liberties, and the Administration of Justice, this report is the third component of the OTA assessment of "Federal Government Information Technology: Congressional Oversight and Civil Liberties. " The first component, *Electronic Surveillance and* Civil *Liberties,* was published in October 1985, and the second, *Management, Security, and Congressional Oversight,* was published in February 1986.

In preparing this report on electronic record systems and privacy, OTA has drawn on working papers developed by OTA staff and contractors, the comments of participants at an OTA workshop on this topic, and the results of an OTA survey that was completed by over 140 agency components. Drafts of this report were reviewed by the OTA project advisory panel, officials from the U.S. Office of Management and Budget and the General Services Administration; U.S. Departments of Justice, State, Defense, and Health and Human Services, among other Federal agencies; and a broad spectrum of interested individuals from the governmental, academic, private industry, and civil liberty communities.

OTA appreciates the participation of the advisory panelists, workshop participants, external reviewers, Federal agency officials, and others who helped bring this report to fruition. The report itself, however, is solely the responsibility of OTA, not of those who so ably advised and assisted us in its preparation.

JOHN H. GIBBONS
*Director*

## Electronic Record Systems and Individual Privacy Advisory Panel

Theodore J. Lowi, *Chairman*
Professor of Political Science, Cornell University

Arthur G. Anderson
IBM Corp. (Ret.)

Jerry J. Berman
Legislative Counsel
American Civil Liberties Union

R.H. Bogumil
Past President
IEEE Society on Social Implications of Technology

James W. Carey
Dean, College of Communications
University of Illinois

Melvin Day
Vice President
Research Publications

Joseph W. Duncan
Corporate Economist
The Dun & Bradstreet Corp.

William H. Dutton
Associate Professor of Communications and Public Administration
Annenberg School of Communications
University of Southern California

David H. Flaherty
Professor of History and Law
University of Western Ontario

Carl Hammer
Sperry Corp. (Ret.)

Starr Roxanne Hiltz
Professor of Sociology
Upsala College

John C. Lautsch
Chairman, Computer Law Division
American Bar Association

Edward F. Madigan
Office of State Finance
State of Oklahoma

Marilyn Gell Mason
Director
Atlanta Public Library

Joe Skinner
Corporate Vice President
Electronic Data Systems Corp.

Terril J. Steichen
President
New Perspectives Group, Ltd.

George B. Trubow
Director, Center for Information Technology and Privacy Law
The John Marshall Law School

Susan Welch
Professor and Chairperson
Department of Political Science
University of Nebraska

Alan F. Westin
Professor of Public Law and Government
Columbia University

Langdon Winner
Associate Professor of Political Science
Rensselaer Polytechnic Institute

Congressional Agency Participants

Robert L. Chartrand
Senior Specialist
Congressional Research Service

Robert D. Harris
Deputy Assistant Director for Budget Analysis
Congressional Budget Office

Kenneth W. Hunter
Senior Associate Director for Program Information
U.S. General Accounting Office

NOTE: OTA appreciates and is grateful for the valuable assistance and thoughtful critiques provided by the advisory panel members. The panel does not, however, necessarily approve, disapprove, or endorse this report. OTA assumes full responsibility for the report and the accuracy of its contents.

# OTA Electronic Record Systems and Individual Privacy Project Staff

John Andelin, *Assistant Director, OTA*
*Science, Information, and Natural Resources Division*

Fred W. Weingarten, *Manager*
*Communication and Information Technologies Program*

Project Staff

Fred B. Wood, *Project Director*

Jean E. Smith, *Assistant Project Director*

Priscilla M. Regan, *Principal Author and Analyst*

Jim Dray, *Research Analyst*

Jennifer Nelson, *Research Assistant*

Administrative Staff

Elizabeth A. Emanuel, *Administrative Assistant*

Shirley Gayheart,* *Secretary*

Audrey Newman, *Secretary*

Renee Lloyd, *Secretary*

Patricia Keville, *Clerical Assistant*

Contractors

William Dutton and Robert Meadow
The University of Southern California

David H. Flaherty
The University of Western Ontario

Karen B. Levitan, Patricia D. Barth, and Diane Griffin Shook
The KBL Group, Inc.

Robert Ellis Smith
The Privacy Journal

*Deceased, Dec. 11, 1985.

## OTA Electronic Record Systems and Individual Privacy Workshop

Robert Belair
Attorney
Kirkpatrick & Lockhart

William Cavaney
Executive Secretary
Defense Privacy Board
U.S. Department of Defense

Louis D. Enoff
Acting Deputy Commissioner for Programs and Policy
Social Security Administration

Robert Freeman
Committee on Open Government
Department of State
State of New York

John Gish
Vice President
W.R. Grace & Co.

John Grace
Privacy Commissioner of Canada

Richard P. Kusserow
Inspector General
U.S. Department of Health and Human Services

Robert Meadow
Professor, Annenberg School of Communications
University of Southern California

Philip Natcharin
Director of Program Integrity
New York State Department of Social Services

Robert Oakley
Assistant Inspector General for Auditing
Veterans Administration

Alan Rodgers
Massachusetts Law Reform Commission

James Rule
Professor, Department of Sociology
State University of New York at Stony Brook

Andy Savitz
Assistant Secretary of Administration and Finance
State of Massachusetts

Robert Ellis Smith
Editor
The Privacy Journal

Jane Tebbutt
Office of Inspector General
U.S. Department of Health and Human Services

Tom Tiffany
Acting Director of Legislative Affairs
Internal Revenue Service

Rob Veeder
Desk Officer
Office of Information and Regulatory Affairs
Office of Management and Budget

# Contents

*Chapter* *Page*

1. Summary . . . . . 3
2. Electronic Record Systems and the Privacy Act:An Introduction . . . . . 11
3. Computer Matching to Detect Fraud, Waste, and Abuse . . . . . 37
4. Computer-Assisted Front-End Verification . . . . . 67
5. Computer Profiling . . . . . 87
6. Policy Implications . . . . . 99

*Appendix*

A. Update on Computerized Criminal History Record Systems . . . . . 129
B. OTA Federal Agency Data Request . . . . . 135
C. List of Contractor Reports . . . . . 145
D. Other Reviewers and Contributors . . . . . 146
E. Summary of Final Rules for Income and Eligibility Verification Required Under the Deficit Reduction Act of 1984 . . . . . 147
F. Privacy and Data Protection Policy in Selected Foreign Countries. . . . . 150

Chapter 1

# Summary

Chapter 1

# Summary

## INTRODUCTION

All governments collect and use personal information in order to govern. Democratic governments moderate this need with the requirements to be open to the people and accountable to the legislature, as well as to protect the privacy of individuals. Advances in information technology have greatly facilitated the collection and uses of personal information by the Federal Government, but also have made it more difficult to oversee agency practices and to protect the rights of individuals.

In 1974, Congress passed the Privacy Act to address the tension between the individual's interest in personal information and the Federal Government's collection and use of that information. The Privacy Act codified principles of fair information use that specified requirements agencies were to meet in handling personal information, as well as rights for individuals who were the subjects of that information. To ensure agency compliance with these principles, the act enabled individuals to bring civil and criminal suits if information was willfully and intentionally handled in violation of the act. In addition, the Office of Management and Budget (OMB) was assigned responsibility for overseeing agency implementation of the act.

At the time the Privacy Act was debated and enacted, there were technological limitations on the use of individual records by Federal agencies. The vast majority of record systems in Federal agencies were manual. Computers were used only to store and retrieve, not to manipulate or exchange information. It was theoretically possible to match personal information from different files, to manually verify information provided on government application forms, and to prepare a profile of a subset of individuals of interest to an agency. However, the number of records involved made such applications impractical.

In the 12 years since the Privacy Act was passed, at least two generations of information technology have become available to Federal agencies. Advances in computer and data communication technology enable agencies to collect, use, store, exchange, and manipulate individual records in electronic form. Microcomputers are now widely used in the Federal Government, vastly increasing the potential points of access to personal record systems and the creation of new systems. Computer matching and computer-assisted front-end verification are becoming routine for many Federal benefit programs, and use of computer profiling for Federal investigations is expanding. These technological advances enable agencies to manipulate and exchange entire record systems, as well as individual records, in a way not envisioned in 1974. Moreover, the widespread use of computerized databases, electronic record searches and matches, and computer networking is leading rapidly to the creation of a de *facto* national databasel containing personal information on most Americans. And use of the social security number as a de *facto* electronic national identifier facilitates the development of this database.

These technological advances have opened up many new possibilities for improving the efficiency of government recordkeeping; the detection and prevention of fraud, waste, and abuse; and law enforcement investigations. At the same time, the opportunities for inappropriate, unauthorized, or illegal access to and use of personal information have expanded. Be cause of the expanded access to and use of personal information in decisions about individuals, the completeness, accuracy, and relevance of information is even more important. Additionally, the expanded access and use make

[1]The term de *facto* national database is used to distinguish it from a national database that was created by' law, i.e. ! a de *jure* national database.

it nearly impossible for individuals to learn about, let alone seek redress for, misuse of their records. Even within agencies, it is often not known what applications of personal information are being used. Nor do OMB or relevant congressional committees know whether personal information is being used in conformity with the Privacy Act.

Overall, OTA has concluded that Federal use of new electronic technologies in processing personal information has eroded the protections of the Privacy Act of 1974. Many of the electronic record applications being used by Federal agencies, e.g., computer profiling and front-end verification, are not explicitly covered by the act or by subsequent OMB guidelines. The rights and remedies available to the individual, as well as agency responsibilities for handling personal information, are not clear. Even where applications are covered by the Privacy Act or related OMB guidelines, there is little oversight to ensure agency compliance. More importantly, neither Congress nor the executive branch is providing a forum in which the conflicts-between privacy interests and management or law enforcement interests—generated by Federal use of new applications of information technology can be debated and resolved. Absent such a forum, agencies have little incentive to consider privacy concerns when deciding to establish or expand the use of personal record systems.

## POLICY PROBLEMS

OTA'S analysis of Federal agency use of electronic record systems, specifically for computer matching, front-end verification, and computer profiling, revealed a number of common policy problems.

First, new applications of personal information have undermined the goal of the Privacy Act that individuals be able to control information about themselves. As a general principle, the Privacy Act prohibits the use of information for a purpose other than that for which it was collected without the consent of the individual. New computer and telecommunication applications for processing personal information facilitate the use of information for secondary purposes, e.g., use of Federal employee personnel information to locate student loan defaulters, or use of Federal tax information to evaluate a Medicaid claim.

The expanded use and exchange of personal information have also made it more difficult for individuals to access and amend information about themselves, as provided for in the Privacy Act. In effect, the Privacy Act gave the individual a great deal of responsibility for ensuring that personal information was not misused or incorrect. Technological advances have increased the disparity between this responsibility and the ability of the individual to monitor Federal agency practices. For example, individuals may not be aware that information about them is being used in a computer match or computer profile, unless they monitor the *Federal Register* or questions about them arise as a result of the application. In computer-assisted front-end verification, individuals may be notified on an application form that information they provide will be verified from outside sources, but are unlikely to be told which sources will be contacted.

Additionally, new computer and telecommunication capabilities enable agencies to exchange and manipulate not only discrete records, but entire record systems. At the time the Privacy Act was debated, this capability did not exist. The individual rights and remedies of the act are based on the assumption that agencies were using discrete records. Exchanges and manipulations of entire record systems make it more difficult for an individual to be aware of uses of his or her record, as those uses are generally not of immediate interest to the individual.

Second, there is serious question as to the efficacy of the current institutional arrangements for oversight of Federal agency compliance with

the Privacy Act and related OMB guidelines. Under the Privacy Act, Federal agencies are required to comply with certain standards and procedures in handling personal information—e.g., that the collection, maintenance, use, or dissemination of any record of identifiable personal information should be for a necessary and lawful purpose; that the information should be current, relevant, and accurate; and that adequate safeguards should be taken to prevent misuse of information.

OMB is assigned responsibility for oversight of agency implementation of the Privacy Act. Prior studies by the Privacy Protection Study Commission (1977), the U.S. General Accounting Office (1978), and the House Committee on Government Operations (1975 and 1983) have all found significant deficiencies in OMB'S oversight of Privacy Act implementation. For example, under the Privacy Act, information collected for one purpose should not be used for another purpose without the permission of the individual; however, a major exemption to this requirement is if the information is for a "routine use' '—one that is compatible with the purpose for which it was collected. Neither Congress nor OMB has offered guidance on what is an appropriate routine use; hence this has become a catchall exemption permitting a variety of exchanges of Federal agency information.

Looking more specifically, OTA found that OMB is not effectively monitoring such basic areas as: the quality of Privacy Act records; the protection of Privacy Act records in systems currently or potentially accessible by microcomputers; the cost-effectiveness of computer matching and other record applications; and the level of agency resources devoted to Privacy Act implementation. OTA also found that neither OMB nor any other agency or office in the Federal Government is currently collecting or maintaining this information on a regular basis. Given the almost total lack of information concerning the activities of Federal agencies with respect to personal information, OTA conducted its own one-time survey of major Federal agencies and found that:

- the quality (completeness and accuracy) of most Privacy Act record systems is unknown even to the agencies themselves; few (about 13 percent) of the record systems are audited for record quality, and the limited evidence available suggests that quality varies widely;
- even though the Federal inventory of microcomputers has increased from a few thousand in 1980 to over 100,000 in 1985, very few agencies (about 8 percent) have revised privacy guidelines with respect to microcomputers;
- few agencies reported doing cost-benefit analyses either before (3 out of 37) or after (4 out of 37) computer matches; authoritative, credible evidence of the cost-effectiveness of computer matching is still lacking; and
- in most Federal agencies, the number of staff assigned to Privacy Act implementation is limited; of 100 agency components responding to this question, *33* reported less than 1 person per agency assigned to privacy and 34 reported 1 person.

Additionally, OTA found that there is little or no governmentwide information on, or OMB oversight of: 1) the scope and magnitude of computer matching, front-end verification, and computer profiling activities; 2) the quality and appropriateness of the personal information that is being used in these applications; and 3) the results and cost-effectiveness of these applications.

Third, neither Congress nor the executive branch is providing a forum in which the privacy, management efficiency, and law enforcement implications of Federal electronic record system applications can be fully debated and resolved. The efficiency of government programs and investigations is improved by more complete and accurate information about individuals. The societal interest in protecting individual privacy is benefited by standards and protections for the use of personal information. Public policy needs to recognize and address the tension between these two interests.

Since 1974, the primary policy attention with respect to Federal agency administration has shifted away from privacy-related concerns. Interests in management, efficiency, and budget have dominated the executive and legislative agenda in the late 1970s and early 1980s. Congress has authorized information exchanges among agencies in a number of laws, e.g., the Debt Collection Act of 1982 and the Deficit Reduction Act of 1984. In these instances, congressional debates included only minimal consideration of the privacy implications of these exchanges.

A number of executive bodies have been established to make recommendations for improving the management of the Federal Government, e.g., the President's Council on Integrity and Efficiency, the President Council on Management Improvement, and the Grace Commission. All have endorsed the increased use of applications such as computer matching, front-end verification, and computer profiling in order to detect fraud, waste, and abuse in government programs. However, these bodies have given little explicit consideration to privacy interests. Some executive guidelines remind agencies to consider privacy interests in implementing new programs, but these are not followed up to ensure agency compliance.

In general, decisions to use applications such as computer matching, front-end verification, and computer profiling are being made by program officials as part of their effort to detect fraud, waste, and abuse. Given the emphasis being placed on Federal management and efficiency, agencies have little incentive to consider privacy concerns when deciding to establish or expand the use of personal record systems. As a result, ethical decisions about the appropriateness of using certain categories of personal information, such as financial, health, or lifestyle, are often made without the knowledge of or oversight by appropriate agency officials (e.g., Privacy Act officers or inspectors general), OMB, Congress, or the affected individuals.

Fourth, within the Federal Government, the broader social, economic, and political context of information policy, which includes privacy-related issues, is not being considered. The complexity of Federal Government relations—within executive agencies, between the executive and legislature, between the Federal Government and State governments, and between the Federal Government and the private sector—is mirrored in interconnecting webs of information exchanges. This complexity and interconnectedness is reflected in myriad laws and regulations, most of which have been enacted in a piecemeal fashion without consideration of other information policies.

Some of these policies may be perceived as being somewhat inconsistent with others, e.g., the privacy of personal information and public access to government information. Some laws and regulations may only partially address a problem, e.g., Federal privacy legislation does not include policy for the private sector or for the flow of information across national borders. In other instances, issues that are inherently related and interdependent, such as privacy and security, are debated and legislated in separate forums with only passing attention to their relationship.

Additionally, the Federal Government information systems, as well as its information policy, are dependent on technological and economic developments. Federal funding for research and development and Federal financial and market regulations will have significant implications for information technologies and markets. Yet, under the present policymaking system, there is no assurance that these implications will be considered. Likewise, the international information policy environment, as well as international technological and economic developments, affects domestic information policy; again, these factors are not systematically considered in the existing policy arenas.

# POLICY ACTIONS

OTA identified a range of policy actions for congressional consideration:

1. Congress could do nothing at this time, monitor Federal use of information technology, and leave policymaking to case law and administrative discretion. This would lead to continued uncertainty regarding individual rights and remedies, as well as agency responsibilities. Additionally, lack of congressional action will, in effect, represent an endorsement of the creation of a de *facto* national database and an endorsement of the use of the social security number as a *de facto* national identifier.
2. Congress could consider a number of problem-specific actions. For example:
   - establish control over Federal agency use of computer matching, front-end verification, and computer profiling, including agency decisions to use these applications, the process for use and verification of personal information, and the rights of individuals;
   - implement more controls and protections for sensitive categories of personal information, such as medical and insurance;
   - establish controls to protect the privacy, confidentiality, and security of personal information within the microcomputer environment of the Federal Government, and provide for appropriate enforcement mechanisms;
   - c review agency compliance with existing policy on the quality of data/records containing personal information, and, if necessary, legislate more specific guidelines and controls for accuracy and completeness;
   - review issues concerning use of the social security number as a de *facto* national identifier and, if necessary, restrict its use or legislate anew universal identification number; or
   - review policy with regard to access to the Internal Revenue Service's information by Federal and State agencies, and policy with regard to the Internal Revenue Service's access to databases maintained by Federal and State agencies, as well as the private sector. If necessary, legislate a policy that more clearly delineates the circumstances under which such accesses are permitted.
3. Congress could initiate a number of institutional adjustments, e.g., strengthen the oversight role of OMB, increase the Privacy Act staff in agencies, or improve congressional organization and procedures for consideration of information privacy issues. These institutional adjustments could be made individually or in concert. Additionally or separately, Congress could initiate a major institutional change, such as establishing a Data Protection or Privacy Board or Commission.
4. Congress could provide for systematic study of the broader social, economic, and political context of information policy, of which information privacy is a part.

# ABOUT THE REPORT

Chapters 2 through 6 of this report provide technical and policy analyses relevant to electronic record systems privacy, and to proposed legislation such as: the "Data Protection Act of 1985" that would establish a Data Protection Board as an independent agency of the executive branch; possible amendments to the Privacy Act and Paperwork Reduction Act; and management improvement legislation.

Appendix A to this report updates trends and issues relevant to the privacy of information in computerized criminal history record systems, the subject of a prior OTA study. Ap-

pendix B describes the methodology of and respondents to the OTA survey (known officially as the OTA Federal Agency Data Request). Appendix C lists the OTA contractor papers relevant to this report. Appendix D lists the outside reviewers and contributors. Appendix E summarizes the Deficit Reduction Act regulations on front-end verification. Appendix F describes the privacy and data protection policies in selected countries.

Other components of this OTA assessment include the October 1985 OTA report on *Electronic Surveillance and Civil Liberties* that discusses issues and options relevant to electronic communications privacy, and the February 1986 OTA report on *Management, Security, and Congressional Oversight* that discusses, among other things, management, technical, and legal issues and options relevant to protecting the security (and, hence, privacy) of computer systems.

Chapter 2

# Electronic Record Systems and the Privacy Act: An Introduction

# Contents


| | Page |
|---|---|
| summary | 11 |
| Introduction | 12 |
| Background | 13 |
| Privacy | 13 |
| History of the Privacy Act | 14 |
| Implementation of the Privacy Act | 16 |
| Requirement | 17 |
| Requirement | 18 |
| Requirement | 19 |
| Requirement | 20 |
| Requirement | 21 |
| Requirement | 21 |
| Findings | 22 |
| Finding 1 | 22 |
| Finding 2 | 25 |
| Finding 3 | 26 |
| Finding 4 | 29 |


## Tables


| *Table No.* | *Page* |
|---|---|
| 1. Statutes Providing Protection for Information Privacy | 15 |
| Z. Privacy Act Record Systems Reported by Federal Agencies | 23 |
| 3. Computerized and Manual Privacy Record Systems | 23 |
| 4. Seriousness of Breaches of Confidentiality | 29 |
| 5. Support for Potential Federal Lawson Information Abuse | 31 |


## Figures


| *Figure No.* | *Page* |
|---|---|
| I. Beliefs That Computers Are an Actual Threat to Personal Privacy in This Country | 27 |
| 2. Change in Percent of Public Believing That Files Are Kept on Themselves | 28 |
| 3. Percent of Public That Believes Each Agency "Shares" Information About Individuals With Others | 30 |

Chapter 2

# Electronic Record Systems and the Privacy Act: An Introduction

## SUMMARY

Although privacy is a value that has always been regarded as fundamental, its meaning is often unclear. Privacy includes concerns about autonomy, individuality, personal space, solitude, intimacy, anonymity, and a host of other related concerns. There have been many attempts to give meaning to the term for policy purposes. In 1890, Samuel Warren and Louis Brandeis defined it as "the right to be let alone. " In 1967, Alan Westin defined it as "the claim of individuals, groups, or institutions to determine for themselves when, how and to what extent information about them is communicated to others. " This latter definition served as the basis for the Privacy Act of 1974 (Public Law 93-579).

The Privacy Act was enacted by Congress to provide legal protection for and safeguards on the use of personally identifiable information maintained in Federal Government record systems. The Privacy Act established a framework of rights for individuals whose personal information is recorded, and the responsibilities of Federal agencies that collect and maintain such information in Privacy Act record systems.

When the Privacy Act was debated and enacted, Federal agency record systems were still based largely on paper documents. In 1986, many Federal agency record systems are based largely on electronic record-keeping. Computers and telecommunications are used to process detailed information on millions of citizens. No longer is personal information merely stored in and retrieved from file cabinets; now large volumes of such information are collected, retrieved, disclosed, disseminated, manipulated, and disposed of by computers. Moreover, direct on-line linkages now make it possible to compare individual information with a host of public and private agencies. Computer tapes, software, and networking also make it possible to compare personal information stored in different record systems.

The Privacy Act, with the goal of providing the means by which individuals could control information about themselves, balanced the interests of Federal agencies in collecting and using personal information against the interests of individuals in controlling access to and use of that information. Technology has now altered that balance in favor of the agencies. Computers and telecommunication capabilities have expanded the opportunities for Federal agencies to use and manipulate personal information. For example, there has been a substantial increase in the matching of information stored in different databases as a way of detecting fraud, waste, and abuse, as will be discussed in chapter 3. Likewise, computers are increasingly being used to certify the accuracy and completeness of individual information before an individual receives a benefit, service, or employment, as will be discussed in chapter 4 on front-end verification. These technological capabilities appear to have outpaced the ability of individuals to protect their interests by using the mechanisms available under the Privacy Act.

In addition to technological threats to Privacy Act protections, several studies of the act's effectiveness have been critical of both agency implementation and Office of Management and Budget (OMB) oversight, and have questioned the individual's ability to use the remedies in a meaningful way. The technological changes have aggravated these problems, and have created some new ones as well.

OTA reached four general conclusions about individual privacy and electronic record sys-

terns that cut across all areas of information technology application:

1. Advances in information technology are having two major, and somewhat opposing, effects on the electronic record-keeping activities of Federal agencies. They are facilitating electronic record-keeping by Federal agencies, enabling them to process and manipulate more information with great speed. At the same time, the growth in the scale of computerization, the increase in computer networking and other direct linkages, the electronic searches of computerized files, and the proliferation of microcomputers are threatening Privacy Act protections.
2. Federal agencies have invested only limited time and resources in Privacy Act matters. Few staff are assigned to Privacy Act implementation, few agencies have developed agency-specific guidelines or updated guidelines in response to technological changes, and few have conducted record quality audits.
3. Privacy continues to be a significant and enduring value held by the American public. General concern over personal privacy has increased among Americans over the last decade, as documented by several public opinion surveys over the past 6 years. About one-half of the American public believes that computers are a threat to privacy, and that adequate safeguards to protect information about people are lacking. There is increasing public support for additional government action to protect privacy.
4. The courts have not developed clear and consistent constitutional principles of information privacy, but have recognized some legitimate expectations of privacy in personal communications.

An OTA survey of the use of information technology by Federal agencies revealed that:

- components within 12 cabinet-level departments and 13 independent agencies reported 539 Privacy Act record systems with 3.5 billion records. Forty-two percent of the systems were fully computerized, 18 percent were partially computerized, and 40 percent were manual. Of the large Privacy Act record systems (i.e., over 500,000 persons), 57 percent were fully computerized, 21 percent were partially computerized, and 22 percent were manual; [1]
- agencies responding reported an increase from a few thousand microcomputers in 1980 to about 100,000 in 1985;
- only about 8 percent of Federal agencies that responded have revised or updated their Privacy Act guidelines with respect to microcomputers; and
- only about 12 percent of agencies reported that they have conducted record quality audits.

[1]Agencies were asked to report only their 10 largest Privacy Act record systems. Twelve of thirteen cabinet departments responded (only the Department of Housing and Urban Development did not), as did 20 selected independent agencies. However, some major personal information collectors within cabinet departments (e.g., the Internal Revenue Service within the Department of the Treasury and the Departments of the Army and Navy within the Department of Defense) did not respond.

# INTRODUCTION

The Federal Privacy Act of 1974 was enacted by Congress to provide legal protection for and safeguards on the use of personally identifiable information maintained in Federal Government record systems. The Privacy Act established a framework of rights for individuals and responsibilities for Federal agencies that collect and maintain personally identifiable information. This framework incorporates a number of "fair information principles" including, primarily, that there should be no secret record systems, individuals should be able to see

and correct their records, and information collected for one purpose should not be used for another.

At the time the Privacy Act was debated, Federal agency record systems were still based largely on paper documents, with some agencies using large mainframe computers for the storage and retrieval of information in very large record systems. By 1986, Federal agencies have become electronic environments with computers and telecommunications being used to process detailed information on millions of citizens. Agencies now use computers, often microcomputers, to collect, disclose, disseminate, manipulate, and dispose of personal information. Direct on-line linkages between computerized databases make it possible to almost instantaneously compare information. Additionally, computer tapes and computer software make it possible to compare entire record systems.

The Privacy Act, with the goal of providing the means by which individuals could control personal information, balanced the interests of Federal agencies in collecting and using personal information against the interests of individuals in that information. Computer and telecommunication capabilities have expanded the interests of Federal agencies in personal information and enhanced their ability to process it. These capabilities have also overshadowed the ability of individuals to use the mechanisms available in the Privacy Act because, in general, it is more difficult for them to follow what occurs during the information-handling process.

The use of computers and telecommunications for processing personal information also offers opportunities for protecting that information. Techniques such as passwords, encryption, and audit trails are available to protect the confidentiality and security of information in an electronic environment. Although their use may provide more protection for the individual, these techniques do not necessarily give the individual control over the stages of information processing, as provided for in the Privacy Act.

# BACKGROUND

## Privacy

Privacy is a value that continues to be highly esteemed in American society, yet its meaning, especially for policy purposes, is often unclear. Privacy is a broad value, representing concerns about autonomy, individuality, personal space, solitude, intimacy, anonymity, and a host of other related concerns. There have been many attempts to define a "right to privacy. " In a seminal article, Warren and Brandeis[2] defined it as "the right to be let alone. " They found the primary source for a general right to privacy in the common law protection for intellectual and artistic property, and argued that:

> ,.. the principle which protects personal writings and all other personal productions, not against theft and physical appropriation, but against publication in any form, is in reality not the principle of private property, but that of an inviolate personality.

Subsequent legal debates have been structured by two points raised by Warren and Brandeis. The first is whether privacy is an independent value whose legal protection can be justified separately from other related interests, such as peace of mind, reputation, and intangible property. The second is controversy over their definition of the "right to privacy" as the "right to be let alone. " Such a definition is so broad and vague that the qualifications necessary to make such a definition practical in society negate the right itself.

Second only to the Warren and Brandeis article in influence on the development of legal thinking regarding protection of privacy in the United States is Dean Presser's 1960 *Califor-*

[2] "(The Right to Privacy, " *Harvard* Law Revriew, 1890.

*nia Law Review* article, "Privacy. " His primary finding is that:

> At the present time the right of privacy, in one form or another is declared to exist by the overwhelming majority of the American courts.[3]

Presser analyzed four distinct torts—intrusion, disclosure, false light, and appropriation—that could be isolated in State common law decisions and that represented four different types of privacy invasions. Each of these torts depends on physical invasion or requires publicity, and hence offers little protection for privacy of personal information. Although Presser's analysis has received wide acceptance as a way of categorizing tort law relating to privacy, most legal scholars doubt that these traditional privacy protections in common law can, or should, be extended to cover more general privacy concerns.

In the mid-1960s, concern with the "privacy" of computerized personal information held by credit agencies and the government rekindled interest in defining a right to privacy. Edward Shils viewed privacy of personal information as:

> . . . a matter of the possession and flow of information, . . Privacy in one of its aspects may therefore be defined as the existence of a boundary through which information does not flow from the persons who possess it to others.[4]

Alan Westin conceived of privacy as "an instrument for achieving individual goals of self-realization, and defined it as "the claim of individuals, groups or institutions to determine for themselves when, how and to what extent information about them is communicated to others.

The "right to privacy' as "the right to control information about oneself" has served as the definition for policy purposes in the United States. Various statutes have been designed to give individuals the means to control information about themselves. Such means include primarily the right to know and the right to challenge and correct. Organizations are also expected to follow "Principles of Fair Information Use, "[6] which establish standards and regulations for collection and use of personal information. See table 1 for a list of statutes providing protection for information privacy.

## History of the Privacy Act

In the mid-1960s, Congress and certain executive agencies began to study the privacy implications of records maintained by Federal agencies. The congressional concern with privacy and individual records was precipitated by the 1965 Social Science Research Council proposal that the Bureau of the Budget establish a National Data Center to provide basic statistical information originating in all Federal agencies.

In 1966, the Senate Committee on the Judiciary, Subcommittee on Administrative Practice and Procedure[7] and the House Committee on Government Operations, Special Subcommittee on Invasion of Privacy,[8] held hearings on the proposals for a National Data Center. Both committees were unconvinced of the need for such a center or of its ability to keep data confidential. In 1967 and 1968, the House and Senate again held hearings on the proposal for a National Data Center, and remained unconvinced that such a center could adequately pro tect the privacy of individual records. The committees and various witnesses feared that once such a center was established, its limited role would not be maintained. There was also great

[3]William L. Presser, "Privacy," *California Law Review*, vol. *48, 1980, Pp. 383, 386.*

[4]Edward Shils, "Privacy: Its Constitution and Vicissitudes, " *Law and Contemporary Problems, vol. 31, 1966,* pp. 281, 282.

[5]Alan Westin, *Privacy and Freedom (New* York: Atheneum, 1967), p. 39.

[6]A "Code of Fair Information Practice" was first developed in: U.S. Department of Heath, Education, and Welfare, *Records, Computers and the Rights of Citizens* (Washington, DC: U.S. Government Printing Office, 1973).

[7]See U.S. Congress, Senate Committee on the Judiciary, Subcommittee on Administrative Practice and Procedure, *invasions of Privacy* (Government Agencies), Hearings, 89th Cong., February 1965, June 1966 (Washington, DC: U.S. Government Printing Office, 1965-67).

[8]See U.S. Congress, House Committee on Government Operations, Special Subcommittee on Invasion of Privacy, *The Computer and Invasion of Privacy,* Hearings, 89th Cong., 2d sess., July 25, 27, 28, 1966 (Washington, DC: U.S. Government Printing Office, 1966).

**Table 2-1 .—Statutes Providing Protection for Information Privacy**

Fair *Credit Reporting Act of* 1970 **(Public Law 91-508.15 U S.C 1681) requires credit Investigation and reporting agencies to make their records available to the subject, provides procedures for correcting Information, and permits disclosure only to authorized customers**

**Crime** *Control Act of* 1973 **(Public Law 93-83) requires that State criminal justice Information systems, developed with Federal funds, be protected by measures to insure the privacy and security of information**

*Family Educational Rights and Privacy Act of* 1974 **(Public Law 93-380 20 U.S.C. 1232(g)) requires schools and colleges to grant students or their parents access to student records and procedures to challenge and correct Information, and limits disclosure to third part [es**

**Privacy** *Act of* 1974 **(Public Law 93-579, 5 U S C 552(a)) places restrictions on Federal agencies' collection, use, and disclosure of per sonally identifiable Information, and gives individuals rights of access to and correction of such Information**

**Tax** *Reform Act of* **1976 (26 U S C 6103) protects confidentialty of tax Information by restricting disclosure of tax Information for nontax purposes The list of exceptions has grown since 1976**

**Right to** *Financial Privacy Act of* **1978 (Public** Law **95.630, 12 U S C 3401) provides bank customers with some privacy regarding their records held by banks and other financial Institutions, and provides procedures whereby Federal agencies can gain access to such records**

**Privacy** *Protection for Rape Victims Act of 1978* **(Public Law 95-540) amends the Federal Rules of Evidence to protect the privacy of rape victims**

**Protection of Pupil Rights** *of* **1978 (20 U S C 1232(h)) gives parents the right to Inspect educational materials used in research or ex perimentation projects, and restricts educators from requiring in - trusive psychiatric or psychological testing**

*Privacy Protection Act of* **1980 (Public Law 96.440, 42 U S C 2000(a)(a)) prohibits government agents from conducting unannounced searches of press offices and files if no one in the office is suspected of committing a crime**

*Electronic Funds Transfer Act of* **1978 (Public Law 95-630) provides that any Institution providing EFT or other bank services must notify its customers about third-party access to customer accounts**

*Intelligence Identifies Protection Act of 1982* **(Public Law 97-200) prohibits the unauthorized disclosure of Information Identifying certain U S. Intelligence officers, agents, Informants, and sources**

**Debt Collection** *Act of* **1982 (Public Law 97-365) establishes due process steps (not Ice, reply. etc ) that Federal agencies must follow before they can release bad debt information to credit bureaus.**

**Cable** *Communications Policy Act of* **1984 (Public Law 98-549) requires the cable service to inform the subscriber of the nature of per sonally identifiable Information collected and the nature of the use of such information, the disclosures that may be made of such information the period during which such information Will be maintained, and the times during which an individual may access such information Also places restrictions on the cable services' collection and disclosures of such Information**

*Confidentiality provisions* **are Included in several statutes, including: the Census Act (13 U S C 9214), the Social Security Act (42 U S C 408(h)), and the Child Abuse Information Act (42 U.S.C. 5103( b(2)(e)))**

NOTE All statutes embody the same scheme of individual rights and fair information practices

SOURCES Robert Aldrich Privacy Protection Law in the *United* **States** (NTIA **Report 82/98.** May 1982 Sarah P Collins Citizens Control over Rec *orals Held by Third Parties* ( CRS Report No 78 255, Dec 8 1978 and the Office of Technology Assessment

reluctance to condone the centralization of both personal information and responsibility for that information within an *executive agency.* Although the committees agreed that the existing situation was inefficient, they believed that such decentralized inefficiency was amenable to congressional oversight, whereas centralized efficiency would be more difficult to check. The proposal for a National Data Center was therefore rejected.

In 1970, the Senate Judiciary Committee, Subcommittee on Constitutional Rights, chaired by Senator Sam Ervin, Jr., began a 4-year study of Federal Government databanks containing personal information and held related oversight hearings.[8] These hearings and the survey of agencies conducted by the Ervin Subcommittee laid the groundwork for the Privacy Act of 1974.

In 1972, Alan Westin and Michael Baker, with the support of the Russell Sage Foundation and the National Academy of Sciences, released a report, *Databanks in a Free Society,* in which they concluded that computerization of records was not the villain it had often been portrayed to be. Their policy recommendations applied to both computerized and manual systems and included:

1. a "Citizen's Guide to Files";
2. rules for confidentiality y and data sharing;
3. limitations on unnecessary data collection;
4. technological safeguards;
5. restricted use of the social security number; and
6. the creation of information trust agencies to manage sensitive data.[9]

[8]See U.S. Congress, Senate Committee on the Judiciary, Subcommittee on Constitutional Rights, *Federal Data Banks, Computers and the Bill of Rights,* Hearings, 92d Cong., 1st sess., Feb. 24-25 and Mar. 2, 3, 4, 9, 10, 11, 15, and 17, 1971, parts 1 and 11 (Washington. DC: U.S. Government Printing Office, 1971).

[9]Alan F, Westin and Michael A. Baker, *Databanks* in *a Free Society (New* York: Quadrangle The New York Times Book Co., 1972).

In 1973, the Secretary of Health, Education, and Welfare's Advisory Committee on Automated Personal Data Systems released its report, *Records, Computers and the Rights of Citizens,* in which it discussed three changes resulting from the use of computerized record-keeping:

1. an increase in organizational data processing capacity;
2. more access to personal data; and
3. the creation of a class of technical record-keepers.

It recommended the enactment of a Federal "Code of Fair Information Practice" that would apply to both computerized and manual systems. This code served as the model for the Privacy Act, as well as for the Council of Europe's 1974 "Resolution on the Protection of the Privacy of Individuals vis~a-vis Electronic Data Banks in the Private Sector. "[11] The major principles of the code include:

- There must be no personal data record-keeping system whose very existence is secret.
- There must be a way for an individual to find out what information about him or her is in a record and how it is used.
- There must be a way for an individual to prevent information about him or her that was obtained for one purpose from being used or made available for other purposes without his or her consent.
- There must be a way for an individual to correct or amend a record of identifiable information about him or her.
- Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuse of the data.[12]

In 1974, in the wake of Watergate, hearings on numerous privacy bills were held in both the Senate and the House.[13] In the subcommittee hearings, there was little disagreement on the need for individual rights with respect to personal information held by Federal agencies. Discussions centered instead on the logistics of enabling individuals to use these rights, and the specific fair information practices that agencies were to follow. The Senate version also provided for a permanent Federal Privacy Board with regulatory powers, while the House version provided no such oversight mechanism. As a compromise, the Privacy Protection Study Commission was created, and oversight responsibilities were given to the Office of Management and Budget.

In 1977, the Privacy Protection Study Commission released its comprehensive report, *Personal Privacy in an Information Society,* which analyzed the policy implications of personal record-keeping in a number of areas including credit, insurance, employment, medical care, investigative reporting, education, and State and local government.[14] The report made numerous policy recommendations, very few of which have been realized in statutory law.

## Implementation of the Privacy Act

A number of studies have evaluated the implementation and effectiveness of the Privacy Act. Most notable are analyses done by the House Committee on Government Operations, the Privacy Protection Study Commission, and the General Accounting Office. All conclude

[11] Reprinted in *Privacy and Protection of Personal Information in Europe,* Staff Report of the Senate Committee on Government Operations (Washington, DC: U.S. Government Printing Office, March 1975).

[12] U.S. Department of Health, Education, and Welfare, *Records, Computers and the Rights of Citizens* (Washington, DC: U.S. Government Printing Office, 1973).

[13] %ee U.S. Congress, Senate Committee on Government Operations, Ad Hoc Subcommittee on Privacy and Information Systems, and Committee on the Judiciary, Subcommittee on Constitutional Rights, *Privacy-The Collection, Use and Computerization of Personal Data,* Joint Hearings, 93d Cong., 2d sess., June 18-20, 1974 (Washington, DC: U.S. Government Printing Office, 1974).

[14] Privacy Protection Study Commission, *Personal Privacy in an Information Society* (Washington DC: U.S. Government Printing Office, 1977) with five appendices: *Privacy Law in the State; The Citizen* as *Taxpayer; Employment Records; The Privacy Act of 1974: An Assessment; and Technology and Privacy.*

*See* U.S. Congress, House Committee on Government Operations, Government Information and Individual Rights Subcommittee, *Implementation of the Privacy Act of 1974: Databanks (1975);* Privacy Protection Study Commission, *The*

that the act has been disappointing in providing protection for individuals from misuse of personal information by Federal agencies. For example, the Privacy Protection Study Commission reached three general conclusions:

1. the Privacy Act represents a large step forward, but it has not resulted in the general benefits to the public that either its legislative history or the prevailing opinion as to its accomplishments would lead one to expect;
2. agency compliance with the act is difficult to assess because of the ambiguity of some of the act requirements, but, on balance, it appears to be neither deplorable nor exemplary; and
3. the act ignores or only marginally addresses some personal-data record-keeping policy issues of major importance now and for the future. 'G

in his opening statement before hearings on oversight of the Privacy Act, Representative Glenn English, Chairman of the Subcommittee on Government Information, Justice, and Agriculture of the Committee on Government Operations, remarked that:

> One of my chief concerns is that the bureaucracy, with the approval of OMB, has drained much of the substance out of the Act. As a result, the Privacy Act tends to be viewed as strictly a procedural statute, For example, agencies feel free to disclose personal information to anyone as long as the proper notices have been published in the Federal Register. No one seems to consider any more whether the Privacy Act prohibits a particular use of information. 17

All of the studies evaluating the implementation and effectiveness of the Privacy Act cite its major weaknesses to be its reliance on individual initiative; the ambiguity of some of the act's requirements; the casual manner in which OMB has implemented and enforced the act; and OMB guidelines issued subsequent to the act that seem to contradict the purpose of the act. These studies report that the act has been used less than anticipated. This *is* attributed to the investment of time and money an individual must make, and to the finding that agencies have not made it easy to use the Privacy Act.

The purpose of the Privacy Act is "to provide certain safeguards for an individual against an invasion of privacy" [Public Law 93-579, sec. 2(b)]. To this end, the act stipulates that Federal agencies meet six major requirements. Each of these requirements, and agency experience to date in meeting each requirement, is discussed below.

### Requirement 1

Permit an individual to determine what records pertaining to him are collected, maintained, used, or disseminated by such agencies.

To this end, agencies are to publish in the *Federal Register an* annual notice of the existence and character of all systems of records containing personal information, and a notice of any new systems of records or new uses of the information in an existing system. The purpose of this was to ensure that there were no secret systems of records by giving the individual notice of agency record-keeping practices. However, most agree that the *Federal Register* is not the ideal vehicle for such notice as it is not easily accessible to most people. In "The President's Annual Report on the Agencies' Implementation of the Privacy Act of 1974" for calendar years 1982 and 1983, OMB identified the effectiveness of the public notice process as one area for further study, noting that:

> The problem may lie in the method used to disseminate this kind of information. While the *Federal Register* stands as the official organ of the government, it is a publication with limited circulation read by few ordinary citizens.'*

*Privacy Act of 1974: An Assessment (1977);* General Accounting Office, *Agencies Implementation of and Compliance With the Privacy Act Can Be Improved (1978);* and House Committee on Government Operations, Government Information, Justice, and Agriculture Subcommittee, *Oversight of the Privacy Act of 1974 (1983).*

"Privacy Protection Study Commission, app. 4, op. cit., p. 77.

"House Committee on Government Operations, 1983, op. cit., p. 5.

'R" The President's Annual Report on the Agencies' Implementation of the Privacy Act of 1974, " CY 1982-1983 (issued Dec. 4, 1985), p. 118.

In 1983, OMB, on the basis of the Congressional Reports Elimination Act of 1982 (Public Law 97-375), eliminated the requirement that agencies republish all of their system notices each year in the *Federal Register.* The reason offered for this decision was lack of public and congressional interest. OMB viewed agency republication as a duplication of the *Federal Register's* annual compilation of Privacy Act notices. OMB recently estimated that the elimination of this requirement, including its administrative expenses, had saved the government over $1 million.[19]

Additionally, the Privacy Act requires agencies to inform individuals, on an application form or on a separate form that individuals can retain, of the following information: 1) the authority that authorizes the solicitation of the information and whether disclosure of such information is mandatory or voluntary; 2) the principal purpose or purposes for which the information is intended to be used; 3) the routine uses that may be made of the information; and 4) the effects of not providing all or any part of the requested information [see Public Law 93-579, sec. 3(e)(3)]. See box A for an example of a Privacy Act notice.

### Requirement 2

Permit an individual to prevent records pertaining to him obtained by such agencies for a particular purpose from being used or made available for another purpose without his consent.

To this end, agencies are to acquire the prior written consent of the individual to whom the record pertains before disclosing a record *unless* one of *twelve* exceptions is met [see Public Law 93-579, sec. 3(b)]. Included in this list are the releases of information to: 1) those officers and employees of the agency that maintains the record who have a need for the record in the performance of their duties; 2) the Bureau of the Census for census-related activities; 3) the National Archives of the United States for historical preservation; 4) a government agency for a civil or criminal law enforcement activity; 5) either House of Congress; and 6) the Comptroller General. The Debt Collection Act of 1982 added an exception for agency disclosure of bad debt information to credit bureaus.

Additionally, an agency may disclose a record without the consent of the individual if the disclosure would be for a "routine use, " defined as "the use of such record for a purpose which is compatible with the purpose for which it was collected" [Public Law 93-579, sec. 3(a)(7)]. If an agency intends to disclose personal information for a "routine use, " then it must publish a notice in the *Federal Register.* This exemption has proved to be quite controversial. In the 1983 Oversight of the Privacy Act Hearings, James Davidson, former counsel to the Senate Subcommittee on Intergovernmental Relations of the Committee on Government Operations, stated that the "routine use" exemption was:

> . . . designed to require that the agencies examine the data, see if the use that the other agency was going to put it to was compatible with the reason for which it was collected, then issue notice so the public and other agencies and OMB could comment on the propriety of the exchange.[20]

Davidson went on to note that this has not been the way that agencies have used the routine use exemption; rather, if agencies had been routinely exchanging information over the years, they have assumed that the routine use exemption allows them to continue.

There have been a number of legislative proposals to amend the "routine use' definition. The Privacy Protection Study Commission recommended that, in addition to the requirement that the use of a record be "compatible with the purposes for which it was collected, " the use also be "consistent with the conditions or reasonable expectations of use and disclosure under which the information in the record was provided, collected, or obtained. "[21] In the 1982

[19] 1 bid., p. 10.

[20] "House Committee on Government Operations, 1983, op. cit., p. 51.

[21] "Privacy Protection Study Commission, app. 4, op. cit., p. 120.

**Box A.—U.S. Department of Education Application for Federal Student Aid, 1986=87 School Year**

## INFORMATION ON THE PRIVACY ACT AND USE OF YOUR SOCIAL SECURITY NUMBER

The Privacy Act of 1974 says that each Federal agency that asks for your social security number or other information must tell you the following:

1. Its legal right to ask for the information and wheiher [he law says you must give it;
2. what purpose the agency has in asking for it and how it will be used; and
3. what could happen if you do not give it.

Our legal right to require that you provide us with your social security number for the Pell Grant and Guaranteed Student Loan programs is based on Section 7 (a) (2) of the Privacy Act Of 1974.

You must give us your social security number to apply for a Pen Grant or a Guaranteed Student Loan. We need the number on this form to be sure we know who you are, to process your application, and to keep track of your record. We also use your social security number in the Pen Grant Program in recording information about your college attendance and progress, in making payments to you directly in case your college does not, and in making sure that you have received your money. If you do not give us your social security number, you will not get a Pen Grant or a Guaranteed Student Loan.

We also ask you to voluntarily give us your social security number if you are using this form only to apply for financial aid under the College Work-study, National Direct Student Loan, and Supplemental Educational Opportunity Grant programs. We use your social security number in processing your application. If you do not give us your social security number, you may still receive financial aid under these three programs.

our legal right to ask for all information except your social security number is based on sections of the law that authorizes the Pell Grant, Supplemental Educational Opportunity Grant, College Work-Study, National Direct Student Loan, and Guaranteed Student Loan programs. These sections include sections 411, 4138, 443, 48, 425, 428, and 482 of the Higher Education Act of 1965, as amended.

If you are applying for Federal student aid under all five programs, you must fill in everything except questions 4-3 and 4-4 on either form, Step 12 on Form 1, and question 1-7 on Form 2. But if you are not applying for a pen Grant or a Sup plemental Educational Opportunity Grant. YOU can also skiP question 4-2 on either form. If you are using Form 1 and *you* are not applying for a Pen Grant or a Guaranteed Student Loan, you can skip questions 5-1 through 5-3 (as well as questions 4-3 and 4-4 and Step 12). Finally, if you are only applying for a Pen Grant and you are using Form 1, you can skip 7-2, 7-3, and 6-3 as well as questions 4-3 and 4-4 and Step 12. if you skip question 4-4, we will count your answer as .. No" for that question.

We ask for the information on the form so that we can figure your '"student aid index" and "expected family contribution." The student aid index is used to help figure out how much of a Pen Grant you will get, if any. The student aid index or the expected family contribution may also be used to figure out how much other Federal financial aid you will get. if any, While you are not required to respond, no Pell Grant may be awarded unless this information is provided and filed as required under 20 U.S.C. 1070a; 34 CFR 690.11.

We will send your name, address, social security number, date of birth, student aid indices, student status, year in college, and State of legal residence to the college that you list in question 4-3 (or its representative), even if you check '"No" in question 44. This Information will also go to the State scholarship agency in your State of legal residence to help them coordinate State financial aid programs with Federal student aid programs. Also, we may send information to members of Congress if you or your parents ask them to help you with Federal student aid questions. We may also use the information for any purpose which is a ""routine use" listed in Appendix B of 34 CFR 5b.

and 1983 "President's Annual Report on the Agencies' Implementation of the Privacy Act of 1974, " problems with the interpretation and implementation of the "routine use" disclosure were identified as Privacy Act issues for further study. The "Annual Report" stated that it would 'be useful for the Congress to reconsider this problem and provide clearer guidance on routine use disclosures. "[2]

[22]" The President's Annual Report, " 1982-1983, op. cit., p. 121.

### Requirement 3

Permit an individual to gain access to information pertaining to him in Federal agency records, to have a copy made of all or any portion thereof, and to correct or amend such records.

These individual rights are a cornerstone of the act; however, they have not been used as much as anticipated. Reasons offered include:

1. the time an individual must spend in communicating with an agency;

2. the possible difficulty in adequately identifying personal records for which access is requested; and
3. the lack of public awareness of these rights.

The Privacy Protection Study Commission concluded that:

> Agency rules on individual access, and on the exercise of the other rights the Act establishes, appear, in most instances, to be in compliance with the Act's rule-making requirements. Yet, they too are often difficult to comprehend, and because the principal places to find them are in the *Federal Register* and the *Code of Federal Regulations,* it **is** doubtful that many people know they exist, let alone how to locate and interpret them.[23]

An additional reason that this goal has not been realized is that there are seven exemptions to this requirement that are authorized by the Privacy Act itself. In general, these exemptions include those systems of records that include investigatory material compiled for law enforcement purposes or for the purpose of determining suitability, eligibility, or qualifications for Federal civilian employment or promotion, military service, Federal contracts, or access to classified material. Also exempt are those systems of records that are maintained in connection with providing protective services to the President or other individuals, and those that are required by statute to be maintained and used solely as statistical records [Public Law 93-579, sec. 3(k)].

In the 1979 "Annual Report of the President on the Implementation of the Privacy Act of 1974, " the individual access provisions were described as the "most apparently successful provision of the Act. "[24] It was reported that since 1977, agencies had recorded over 2 million requests for access and had complied with over 96 percent of the requests. But, the 1979 Annual Report noted that it was not clear whether the access requests were the "direct result of the Act" because of prior procedures by which employees and clients were given access to their records.[26] In the 1982-83 Annual Report, OMB reported that access requests and requests to amend records had declined for most of the agencies with major record holdings. OMB attributed this to the existence of other agency access policies (for example, for personnel records) that are used rather than filing a Privacy Act request.[2C]

### Requirement 4

Collect, maintain, use, or disseminate any record of identifiable personal information in a manner that assures that such action is for a necessary and lawful purpose, that the information is current and accurate for its intended use, and that adequate safeguards are provided to prevent misuse of such information.

These "Fair Information Principles" are another cornerstone of the act. Yet, the agencies have loosely construed these requirements and have at times ignored them altogether. The Privacy Protection Study Commission concluded that:

> None of these several collection requirements and prohibitions appears to have had a profound impact on agency record-keeping practice, mainly because they are either too broadly worded or have been perceived as nothing more than restatements of longstanding agency policy .27

In testimony before the House Subcommittee on Government Information, Justice, and Agriculture, John Shattuck, then legislative director for the American Civil Liberties Union, reached a similar conclusion, stating that:

> The Code of Fair Information Practices which constitutes the core of the statute is so general and abstract that it has become little more than precatory in practice, and has proved easy to evade.[28]

The vagueness of the principles contributes to agencies' practices. T-he act does not define,

[23]Priva;y Protection Study Commission, **app.** 4, op. cit., p. 84.

[24]"Fifth Annual Report of the President on the Implementation of the Privacy Act of 1974, " Calendar Year 1979 (released August 1980), p. 11.

[25]Ibid.

[26]Ibid., p. 20.

[27]Privacy Protection Study Commission, **app.** 4, op. cit., p. 44.

[28]House Committee on Government Operations, 1983, op. cit., p. 273.

nor does it require agencies to set standards for, such terms as "current" or "necessary." The act also does not develop, nor does it require agencies to develop, procedures to ensure "accurate" information or "adequate safeguards . . . to prevent misuse. "

## Requirement 5

Permit exemptions from the requirements with respect to records provided in this Act only in those cases where there is an important public policy need for such exemption as has been determined by specific statutory authority.

As discussed above, the exemptions for permission to disclose, and for access and correction, are broadly defined. However, overall, agencies exempt only a small percentage of their systems of records. In order to ensure that agencies only exempted systems of records where necessary, the Privacy Act requires that the President report annually on the operation of the exemption provision. In the 1979 Annual Report, OMB concluded that agencies have "implemented this provision in a thoughtful and sparing manner" and that:

- Only 14 percent of total systems have been exempted.
- Agencies have invoked exemptions to completely deny access in only 0.2 percent of cases.
- Agencies routinely screen records in exempt systems and release material not deemed to need protection."

In the 1982-83 Annual Report, OMB reported that, from 1975 to 1983, the number of exempt systems declined by over 16 percent.[30]

## Requirement 6

Be subject to civil suit for any damages which occur as a result of willful or intentional action which violates any individual's rights under this Act.

This requirement is intended to provide individuals the means to enforce agencies to comply with the provisions of the act, if they were not satisfied with the outcome of an administrative appeal. The time and cost involved to bring a suit under the Privacy Act is often prohibitive. In addition, some individuals have used the Freedom of Information Act, rather than the Privacy Act, to gain access to their records, and thus cannot bring suit under the Privacy Act. Where individuals have used the Privacy Act, their civil suits have rarely been successful because of the need to find" willful or intentional" activity, because injunctive relief under the act is unclear, and because the courts have narrowly construed the circumstances under which an individual can recover damages.[31] Richard Ehlke of the Congressional Research Service summarized the situation as follows:

> Despite over seven years of operation, the case law under the Privacy Act is relatively undeveloped. The greater visibility of the Freedom of Information Act, the breadth of many of the Privacy Act exceptions, and the limited remedial scheme of the Act are undoubtedly factors in this development. Much of the litigation has focused on these aspects of the Act–the limitations inherent in the "record" and "system of records" triggers to the Act; the expansive law enforcement exemptions; the exceptions to the consensual disclosure requirement; and the limited remedies available to redress many violations of the Act.[32]

[30] "President's Annual Report, 1979, " op. cit., p. **14.**
[31] "President's Annual Report, 1982 -83," op. cit., p. 19.

[31] See Richard Ehlke, "Litigation Trends Under The Privacy Act, ' June 1983, *Congressional Research Service,* in *Oversight of the Privacy Act of 1974,* op. cit., pp. 437-469.
[32] Ibid., pp. 468-469.

# FINDINGS

OTA has reached four general conclusions about individual privacy and electronic record systems that cut across all areas of application of information technology. Each finding is discussed below.

## Finding 1

Advances in information technology are having two major, and somewhat opposing, effects on the electronic record-keeping activities of Federal agencies.

They are facilitating electronic recordkeeping by Federal agencies, enabling them to process and manipulate more information with great speed. At the same time, the growth in the scale of computerization, the increase in computer networking and other direct linkages, electronic searches of computerized files, and the proliferation of microcomputers are threatening Privacy Act protections.

In the early 1960s, the use of computers to process personal information in Federal agencies was in its beginning stages and Federal agencies were still largely paper environments.[33] At this time, most computing was done on large mainframes by central processing, and only record systems containing a large number of records were stored on computers.

In 1975, the First Annual Report of the President on Implementation of the Privacy Act reported that 73 percent of the personal data systems subject to the act were totally manual, but the remaining 27 percent that were fully or partially computerized contained over 80 percent of the total individual records.[34]

In 1985, the increase in the number of computerized records is significant. In the OTA survey, agencies were asked to report their 10 largest Privacy Act record systems. Components within 12 cabinet-level departments[35] and 13 independent agencies[36] reported a total of 539 Privacy Act record systems containing 3.5 billion records. Of these systems, 42 percent were totally computerized, 18 percent were partially computerized, and 40 percent were wholly manual (see table 2). More importantly, of the large systems of records (i.e., over 500,000 persons), 57 percent were totally computerized, 21 percent were partially computerized, and 22 percent were wholly manual (see table 3).

The qualitative changes that have occurred in the various stages of the information process as a result of computerization are also significant. No longer is information merely stored and retrieved by computer. Now information is routinely collected on computer tapes, used within an agency in computer form, exchanged with and disclosed to regional offices or other agencies in computer form, manipulated and analyzed with sophisticated computer software, and archived on computer tapes.

[33] Before the Privacy Act was passed, two surveys of the degree of computerization of Federal agency record systems were conducted. In 1966, the Senate Judiciary Subcommittee on Administrative Practice and Procedure conducted a survey of "government dossiers" to determine the extent and nature of Federal agencies' collection of personal information. The subcommittee determined that Federal files contained more than 3 billion records on individuals, and that over one-half of these records were retrievable by computers. [See: U.S. Congress, Senate Committee on the Judiciary, Subcommittee on Administrative Practice and Procedure, *Government Dossier* (Committee Print) (Washington, DC: U.S. Government Printing Office, 1967), pp. 7-9.] The Subcommittee on Constitutional Rights, chaired by Senator Sam Ervin, surveyed agencies and found that 86 percent of the 858 databanks with 1.25 billion records on individuals were, at least in part, computerized. The large percentage of computerization found by the Ervin study may be attributed in part to the fact that the study used the phrase "databank containing personal information about individuals." To many, "databank" may imply a computerized system; thus, it is likely that manual systems were underreported in the Ervin survey. (See U.S. Congress, Senate Committee on the Judiciary, Subcommittee on Constitutional Rights, *Federal Data Banks and Constitutional Rights,* 93d Cong., 2d sess., 1974.)

[34] *Federal Personal Data Systems Subject to the Privacy Act of 1974, First Annual Report of the President, Calendar Year 1975, Pp. 4-6.*

[35] Only the Department of Housing and Urban Development did not respond to this question at all. However, some major personal information collectors within cabinet departments (e.g., Internal Revenue Service within the Department of the Treasury and the Departments of the Army and Navy within DOD) did not respond.

[36] Consumer Product Safety Commission, Federal Trade Commission, National Aeronautics and Space Administration, Nuclear Regulatory Commission, Securities and Exchange Commission, Selective Service System, Agency for International Development, Federal Election Commission, Federal Reserve System, Small Business Administration, National Archives and Records Administration, Commission on Civil Rights, and Arms Control and Disarmament Agency.

**Table 2.—Privacy Act Record Systems Reported by Federal Agencies[a]**

| Agency | Fully computerized | | Partially computerized | | Subtotals | | [b]Manual | | Totals | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Number of systems | Number of records | Number of systems | Number of records | Number of systems | Number of records | Number of systems | Number of records | Number of systems | Number of records |
| Agriculture | 22 | 27.0 | 6 | 1.5 | 28 | 28.5 | 14 | 05 | 42 | 290 |
| Commerce | 13 | 8821 | 3 | 04 | 16 | 882.5 | 5 | 14 | 21 | 883.9 |
| DOD | 15 | 500 | 4 | 17 | 19 | 51.7 | 32 | 36 | 51 | 553 |
| Education | 3 | 17 | 1 | 00 | 4 | 17 | 0 | 00 | 4 | 17 |
| Energy | 3 | 04 | 7 | 04 | 10 | 08 | 4 | 03 | 14 | 15 |
| DHHS | 26 | 1,3046 | 16 | 90 | 42 | 1,3136 | 20 | 901 | 62 | 1,4037 |
| Interior | 32 | 45 | 11 | 52 | 43 | 9.7 | 17 | 04 | 60 | 10.1 |
| Justice | 28 | 101 2 | 9 | 2244 | 37 | 325.6 | 31 | 22 | 68 | 3278 |
| Labor | 8 | 16 | 9 | 09 | 17 | 25 | 1 | 00 | 18 | 25 |
| DOT | 36 | 100 | 8 | 30 | 44 | 130 | 17 | 02 | 61 | 132 |
| Treasury | 16 | 488 | 6 | 36.1 | 22 | 849 | 20 | 4603 | 42 | 5452 |
| State | 0 | 00 | 1 | 200 | 1 | 20.0 | 9 | 902 | 10 | 1102 |
| Independent agencies | 27 | 224 | 15 | 10 | 42 | 23.4 | 44 | 51 4 | 86 | 748 |
| Totals | 229 | 2,454.3 | 96 | 3036 | 325 | 2,7579 | 214 | 700.6 | 539 | 3,4589 |

[a]Agencies were asked to report only their 10 largest privacy Act record systems Twelve of thirteen Cabinet departments responded; only the Department of Housing and Urban Development did not as did 13 out of 29 independent agencies see app B at the end of this report for a list, and some major privacy recordholders did not respond e g the Internal Revenue Service of the Department of the Treasury, and the Departments of Army and Navy in the Department of Defense

[b]Millions of records

SOURCE Office Of Technology Assessment

**Table 3.—Computerized and Manual Privacy Record Systems**

| | Large systems[a] | | Medium systems[b] | | Small systems[c] | | Totals | |
|---|---|---|---|---|---|---|---|---|
| | Number | Number of persons | Number | Number of persons | Number | Number of persons | Number | Number of persons |
| 100% computerized | 43 | 1,653,336,199 | 105 | 11,277,938 | 81 | 237,240 | 229 | 1,664,851,377 |
| Parilally computerized | 16 | 285,880,382 | 41 | 3,912,622 | 39 | 213,790 | 96 | 290,006,794 |
| 100% manual | 17 | 695,419,523 | 50 | 5,015.434 | 147 | 327,666 | 214 | 700,762,623 |

[a]Over 500. 000 person:
[b]-o 001 [illegible] 000 persons
[c]Under 10 000 persons

SOURCE Office of Technology Assessment

Another significant change is the direct linkage of computer records via telecommunication systems. This allows for easy disclosure and exchange of information. On-line access can occur, for example, via private or public telephone lines or through local networks within an agency. One factor supporting the transition of Federal information systems to direct linkages is cost–the cost of a typical network interface was $500 in 1982, but is expected to drop to about $50 by 1987.[37] Another factor is the ease and efficiency to an agency official of communicating directly with the computer as information is collected or needed, rather than compiling transactions, batch= processing them on a tape at the end of the day or week, and waiting for a reply.

With such computer networking, the exchanges of information occur rapidly, often leaving no audit trail of who had access to the data or what changes were made. Monitoring the use of agency information becomes much more difficult in this environment. But, at the same time, the environment supports a vast increase in the exchange and manipulation of information, as well as an increase in the number of people having access to the *information.* In 1977, the Privacy Protection Study Commission warned that:

> The real danger is the gradual erosion of individual liberties through the automation, integration, and interconnection of *many small,* separate recordkeeping systems, each of which alone may seem innocuous, even benevolent, and wholly justifiable.[38]

[37]See Michael Killen, "The Microcomputer Connection to Local Networks, " *Data* Communications, December 1982.

[38]Privacy Protection Study Commission, app. 4, op. cit., p. 108.

Another technological development that has implications for Privacy Act protections is efficient electronic searching through computer records. The two most common types of searches are computer matching and computer profiling (or computer screening). In a computer match, two sets of computer files are compared record by record to look for any individuals who appear in both files. In a computer profile or computer screen, a single computer file is searched for selected factors about a specific type of individual. Because of the importance of these electronic searches, each will be discussed in depth in the following chapters.

Another critical factor in the Federal agency technology environment in the mid-1980s is the microcomputer. The microcomputer puts the power of information collection, storage, retrieval, exchange, manipulation, and printing into the hands of discrete individuals. In doing so, it raises privacy, security, productivity, and management issues that had been irrelevant or dormant in other eras of information processing.[39]

Because of the control over information processing that microcomputers give users and because of their relatively low cost, the use of microcomputers has grown dramatically across all sectors of society. The Federal Government has not been immune to this trend. All agencies are experiencing an influx of microcomputers. The OTA survey revealed that the agencies surveyed had a few thousand microcomputers in 1980 and over 100,000 in 1985.

A major impetus in this demand for microcomputers within the Federal Government is the perceived need to increase productivity and efficiency. The broad range of information processing features that a microcomputer offers and the variety of software programs available make microcomputers attractive throughout an agency. For clerical work, microcomputers are used most often for document preparation and data entry .40 At the administrative level, microcomputers are used for accounting, budgeting, and planning. Microcomputers can be used by professionals for data analysis as well as document preparation. For technical users, microcomputers offer control over system design and programming.[41]

Microcomputers complicate the monitoring of the uses of personal information for two reasons. First, they make it easier for individual users to create their own systems of records. This complicates Privacy Act oversight because files created on microcomputers were not considered when the Privacy Act was enacted, and it may be impractical to subject them to the act. The Privacy Act applies to a "record" that is retrieved from a "system of records. " The Privacy Act defines "record" to mean:

> . . . any item, collection, or grouping of information about an individual that is maintained by an agency, including, but not limited to, his education, financial transactions, medical history, and criminal or employment history and that contains his name, or the identifying number, symbol, or other identifying particular assigned to the individual, such as a finger or voice print or a photograph.

The act defines "system of records" to mean:

> . . . a group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual.[42]

If a file created and maintained on a microcomputer meets the criteria for a system of records, i.e., is retrieved by name, identifier, or other identifying particular, then individuals should have the right to access and amend their records. To do so, all microcomputer files containing records that are retrievable by name

[39]The KBL Group, Inc.,"Agency Profiles of Civil Liberties Practices, " OTA contractor report, December 1984, p. 153.

[40]See U.S. Congress, Office of Technology Assessment, *Automation of America's Offices, OTA-CIT-287* (Washington, DC: U.S. Government Printing Office, December 1985) for an in-depth analysis of the effects of microcomputers in the workplace.

[41]National Bureau of Standards, *Microcomputers: Introduction to Features and Uses,* Special Publication 500-110, March 1984, pp. viii-ix.

[42]Privacy Act of 1974 (Public Law 93-579), sec. 3(a)(4)(5).

or other identifier would need to be reported to the Privacy Act Officer and noted in the Federal *Register.*

The second feature of the microcomputer that makes it difficult to monitor the uses of personal information is that a microcomputer serves as a remote terminal to access centralized systems of records. Such shifting of data from mainframes to microcomputers raises critical questions of data integrity and security. For example, when a record is being used by one user, there may be no other access to that information. More importantly, there may be no audit trail of additions and deletions." Additionally, there may be no indication of how current the records are, thus increasing the likelihood that inaccurate data will be disseminated.[44]

At the present time, most microcomputers in Federal agencies are desk-top models. The trend to portable computers—also known as briefcase, lap, or notebook computers—and transportable computers will aggravate the problems of data integrity and security, especially since information will be transported out of government offices into areas that are neither controlled nor secured. Another technological development that will have implications for the processing of personal information is the multiuser microcomputers, or "super microcomputers, which are used primarily for group work situations.

## Finding 2

Federal agencies have invested only limited time and resources in Privacy Act matters. Few staff are assigned to Privacy Act matters, few agencies have developed agency-specific guidelines or updated guidelines in response to technological changes, and few have conducted record quality audits.

The Privacy Act allows agencies much latitude to develop their own arrangements for supervising implementation and compliance with the act. The only requirement the act places on agencies is to:

> . . . establish rules of conduct for persons involved in the design, development, operation, or maintenance of any system of records, or in maintaining any record, and instruct each such person with respect to such rules and the requirements of this section, including any other rules and procedures adopted pursuant to this section and the penalties for noncompliance [Public Law 93-579, sec. 3(e)(9)].

In 1977, the Privacy Protection Study Commission reviewed agency experience and concluded that:

> . . . the 97 Federal agencies that maintain systems of records subject to the Privacy Act of 1974 have all taken different approaches to administration, training, and compliance monitoring. . . agencies or components of agencies that have carefully structured programs for administering the Act appear to be the ones in which the Act's objectives are being best achieved.[45]

Based on responses to the OTA survey of Federal agencies, 67 percent of agencies responding reported one (34 agencies) or less than one (33 agencies) full-time equivalent (FTE) staff assigned to Privacy Act matters. Only seven agencies reported ten or more FTEs assigned to Privacy Act matters, and six of these were located in the Department of Justice. The FBI reported the largest number of FTEs–65—assigned to Privacy Act issues.

The Privacy Act requires agencies to:

> . . . maintain all records which are used by the agency in making any determination about any individual with such accuracy, relevance, timeliness, and completeness as is reasonably necessary to assure fairness to the individual in the determination [Public Law 93-579, sec.3(e)(5)].

OTA asked agencies to specify the procedures they follow to ensure Privacy Act record quality (for example, complete and accurate records). In response, most agencies submitted a copy of their policy directives con-

[44]National Bureau of Standards, op. cit., p. 96.
[44]The KBL Group, Inc., op. cit., p. 162.

[45]Privacy Protection Study Commission, app. 4, op. cit., p. 108.

taining general information and procedures for administering the Privacy Act. Only about 24 percent (30 agencies) have developed agency-specific guidelines or procedures for determining what is "relevant' and 'timely' information within their agency.

The results of the OTA survey also indicated that few agencies had conducted audits of record quality. Of 127 agency respondents, only about 13 percent (16 agencies) indicated that they conducted record quality audits. Of these 16 agencies, none provided copies of the results.[46] With respect to record quality statistics for law enforcement, investigative, and intelligence record systems, only one agency provided statistics (for three systems under its jurisdiction). No statistics were provided for any of the other 82 systems reported.[47]

The OTA survey also asked whether agencies had revised or updated Privacy Act guide lines with respect to microcomputers. Of 119 agency respondents, only 8.4 percent (10 agencies) had done so. One agency noted that microcomputers were not used in connection with the maintenance of Privacy Act information; however, as was noted above, files on microcomputers or accessible through microcomputers may well fall under the Privacy Act "system of records" criteria.

## Finding 3

Privacy continues to be a significant and enduring value held by the American public, as documented by several public opinion surveys over the past 6 years.

About one-half of the American public believes that computers are a threat to society, and that adequate safeguards do not exist to protect information about people. There is increasing public support for additional government action to protect privacy.

This finding is based on a comprehensive review of public opinion surveys that covered issues of technology and civil liberties, with special attention to the question of privacy and information practices.[48] Most studies, although privately sponsored, were designed and conducted by major public opinion research organizations such as Louis Harris & Associates, the Gallup Organization, the Roper Organization, the National Opinion Research Center, and the major news organizations.

A major difficulty in interpreting existing survey research is that most questions have emphasized general concerns about privacy and civil liberties, rather than specific concerns about the implications of particular uses of computing and information technologies, such as computer matching or computer profiling. As a result, much is known about abstract concerns for privacy, but little about levels of support or opposition to emerging technologies and their use by government agencies. An additional problem of survey research is that the meaning of responses is clouded by definitional differences in what constitutes an invasion of privacy, including definitions ranging from personal freedoms, solitude, and freedom from gossipy neighbors to freedom from governmental or employer surveillance. With these caveats in mind, a number of conclusions and trends about public opinion can be made.

General concern over personal privacy has increased among Americans over the last decade. When asked directly whether they are concerned about threats to personal privacy, most Americans will answer in the affirmative. In several Harris surveys[49] the following question was posed:

---

[46] A total of 142 agencies were surveyed; 5 did not respond at all, and 10 others responded that the question was not applicable or the information was not available, for a net total response of 127 agencies.

[47] Again, 142 agencies were surveyed; a total of 85 computerized law enforcement, investigative, or intelligence record systems were identified. Agencies responded as follows: record quality statistics maintained (3 systems); no record quality statistics (63 systems); no response (17 systems); not applicable or information not available (1 system); and classified (1 system).

[48] William H. Dutton and Robert G. Meadow, "Public Perspectives on Government Information Technology: A Review of Survey Research on Privacy, Civil Liberties and the Democratic Process," OTA contractor report, January 1985.

[49] Louis Harris & Associates, Inc., and Dr. Alan F. Westin, *The Dimensions of Privacy: A National Opinion Research Survey of Attitudes Toward Privacy* (conducted for Sentry Insurance), December 1979; and Louis Harris & Associates, Inc., *The Road After 1984: A Nationwide Survey of the Public and Its*

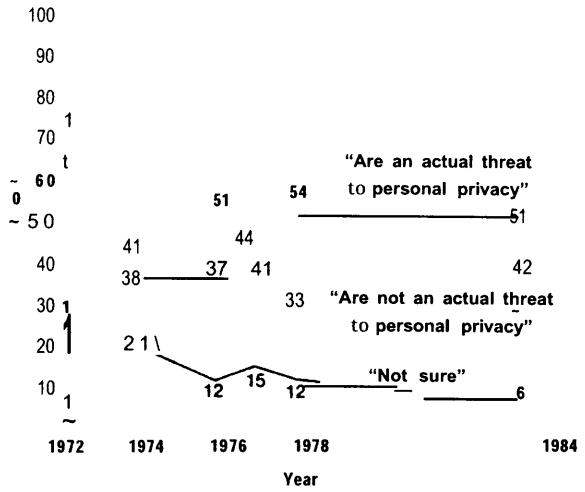
> Now let me ask you about technology and privacy. How concerned are you about threats to your personal privacy in America today? Would you say you are very concerned, somewhat concerned, only a little concerned, or not concerned at all?

In 1983, 48 percent of the public described themselves as "very concerned. " This was double the 25 percent reported in January 1978 and a marked increase from 31 percent in December 1978. In 1983, an additional 29 percent described themselves as "somewhat concerned, " and only 7 percent said they were "not concerned at all, " a significant change from the 28 percent who so described themselves in January 1978. In addition, Americans overwhelmingly disagree (64 percent, compared with 27 percent who agree) with the statement that: "Most people who complain about their privacy are engaged in immoral or illegal conduct. " In other words, privacy is not merely an instrument for avoiding punishment or detection–it is seen as a legitimate value itself.

Most recently, about one-half of the American public believed that computers were a threat to privacy. As figure 1 indicates, the percentage perceiving computers as a threat has increased since 1974. In 1974, 38 percent of the respondents said computers were a threat and 41 percent said they were not. In 1977, 41 percent said computers were a threat and 44 percent said they were not a threat. In December 1978, 54 percent said they were a threat and only 33 percent indicated they were not. However in 1983, the percentage perceiving computers as a threat to privacy decreased slightly, while the percentage believing that computers are not a threat increased by approximately 10 percent. In 1982, Roper reported that 44 percent were very concerned with reports of abuse of personal information that is stored in computers, and 39 percent were very concerned about "reports of embezzlements and rip-offs through the use of a computer. "

*Leaders on the New Technology and Its Consequences for American Life* (conducted for the Southern New England Telephone for presentation at The Eighth International Smithsonian Symposium, December 1983.)

**Figure I.— Beliefs That Computers are an Actual Threat to Personal Privacy in This Country[a]**




[a]Response 10 Do you feel that the present use of computers are an actual threat to personal privacy in this country or not?

**SOURCE. Louis Harris & Associates** Inc *The Road After 1984 A Nationwide Survey of the Public and Its Leaders on the New Technology and Its Consequences for American Life* (conducted for the Southern New England Telephone for **presentation at the Eighth International** Smithsonian **Symposium, December 19S3)**

An increasing percentage of the public does not believe that the privacy of personal information in computers is adequately safeguarded –from 52 percent in 1978 to 60 percent in 1983. Although a majority of the public (60 percent) believes that computers have improved the quality of life,[50] a larger and increasing (68 percent in 1983) percentage of the public believes that the use of computers must be sharply restricted in the future if privacy is to be preserved .51

In general, citizens are concerned with the protections organizations provide for personal information. In 1979, 41 percent agreed and 41 percent disagreed with the statement: "Most organizations that use information about people have enough checks and safeguards against the misuse of personal information. " Government agencies were perceived as intrusive by about one-third of the public, with the Central Intelligence Agency, the Federal Bureau of Investigation, and government welfare agencies

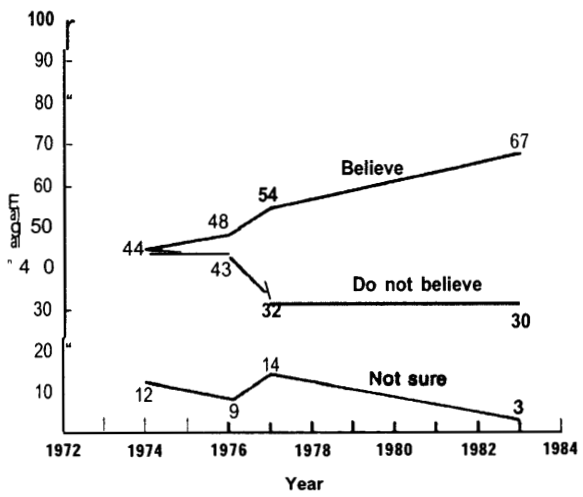
[50] Harris, op. cit., 1979, table 9.2.
[51] Harris, op. cit., 1983, table 3-3.

being mentioned most often as asking for too much personal information. About one-third of the public believe that government agencies should be doing more to maintain the confidentiality of personal information.[52] Most Americans believe that personal information about them is being kept in "some files somewhere for purposes not known" to them. As figure 2 indicates, the percentage of the public believing this to be the case has increased over time, with a high of 67 percent in 1983.

Most Americans, from two-thirds to three-fourths, believe that agencies that release the information they gather to other agencies or individuals are seriously invading personal privacy[53] (see table 4). But, as figure 3 indicates, significant percentages of the public believe that public and private organizations do share information about individuals with others.

[52]Harris, op. cit., 1979, tables 2.2, 2.5, 2.6, 2.8, 2.9, 8.1.
[53]Harris, op. cit., 1983, table 1-6.

**Figure 2.—Change in Percent of Public Believing That Files"Are Kept on Themselves[a]**




[a]Response to "Do you believe that personal (n formation about yourself is being kept In some files somewhere for purposes not known to you, or don't you believe this is so?"

**SOURCE' Louis Harris & Associates,** Inc, *The Road After 1984: A Nationwide Survey of the Public and Its Leaders on the New Technology and Its Consequences for American Life* (conducted for the Southern New England Telephone for presentation at the Eighth International Smithsonian Symposium, December 1983).

The American public does not look favorably upon central files and databanks. Most Americans, 84 percent, believe that master files containing personal information, such as credit and employment histories, organizational affiliations, medical history, voting record, phone calls, buying habits, and travel, could be compiled "fairly easily. " Only 1 percent of the Harris respondents expressed uncertainty over this possibility. Seventy-eight percent believed that if such a master file were put together, it would violate their privacy .54

There is increasing support for additional government action to protect privacy. In 1978, the public was not sure who should be responsible for maintaining privacy. Nearly one-half (49 percent) said it should rest with the people themselves, while 30 percent said the courts, 26 percent Congress, 25 percent the States, 14 percent the President, and 12 percent said employers." Despite confusion over the source of responsibility, two-thirds of the public responded that laws could go a long way to help preserve our privacy. be Sixty-two percent "he public thought it was very important that there bean independent agency to handle complaints about violations of personal privacy by organizations .57 However, 46 percent were opposed to the creation of a National Privacy Protection Agency to protect privacy .68

In surveys conducted by the Roper Center in 1982,[59] large majorities believed that laws were needed to govern how information on individuals can be used by organizations that have computer files, and supported the major principles of the "Code of Fair Information Practices. " In 1982, 85 percent wanted laws to ensure that corrections of information were included in files, 82 percent said that individ-

[54]Ibid., table 1-2.
[55]Harris, op. cit.1979, table 10.11.
[56]Ibid., table 10.3.
[57]Ibid., table 10.5.
[58]Ibid., table 10.4.
[59]The Roper Center, Institute of Social Research, University of Michigan, contains surveys by the major private polling organizations, including Gallup, Harris, Yankelovich, CBS/New York Times, and Roper. OTA commissioned a keyword search at the Roper Center to locate all previous public opinion research studies on any aspect of attitudes toward government information technology.

**Table 4.—Seriousness of Breaches of Confidentiality**

Q.: I'm going to read a few things which might be considered an invasion of privacy, all of which deal with computerized information. Do you feel that (READ EACH ITEM) would be a serious invasion of privacy, or not?

| | | Leaders | | | |
|---|---|---|---|---|---|
| | Total public | Congressmen and top aides | Corporate executives | Media: science editors | Superintendents of schools |
| **Base** | 1,256 | 100 | 100 | 100 | 100 |
| The Internal Revenue Service not keeping individual Federal tax returns confidential: | | | | | |
| Serious | 84% | 98% | 93% | 95% | 89% |
| Not serious | 15 | 2 | 7 | 5 | 11 |
| The FBI not keeping information about individuals confidential: | | | | | |
| Serious | 82 | 95 | 93 | 91 | 86 |
| Not serious | 15 | 4 | 6 | 8 | 14 |
| Banks sharing information about an individual's banking habits and size of bank accounts: | | | | | |
| Serious | 78 | 66 | 60 | 66 | 78 |
| Not serious | 20 | 30 | 38 | 33 | 22 |
| A credit business selling information about an individual credit standing: | | | | | |
| Serious | 77 | 64 | 46 | 73 | 75 |
| Not serious | 22 | 34 | 54 | 25 | 25 |
| The Census Bureau not keeping information about individuals confidential: | | | | | |
| Serious | 73 | 88 | 73 | 82 | 75 |
| Not serious | 25 | 11 | 27 | 18 | 25 |
| Insurance companies sharing information gathered about an individual: | | | | | |
| Serious | 72 | 64 | 63 | 66 | 72 |
| Not serious | 26 | 31 | 35 | 32 | 28 |

SOURCE Lou Is Harris & Associates, Inc , The *Road After* 1984 *A Nationwide Survey of the Public and its Leaders* on the New Technology *and its Consequences for American Life* (conducted for the Southern New England Telephone for presentation at the Eighth International Smith son lan Symposium. December 1983)

uals should be notified of the existence and contents of files containing information about them, 82 percent thought there should be laws to permit people to get copies of any information in files on themselves, and 71 percent thought there should be laws prohibiting most private parties from asking for social security numbers.'" In addition, 72 percent said businesses should have the right to get information only from the person directly, while only 14 percent said databanks were appropriate."

In the 1983 Harris survey (see table 5), strong majorities of the public and majorities of all four leadership groups supported the enactment of new Federal laws to deal with information abuse, including laws that would require that any information from a computer that might be damaging to people or organizations must be double-checked thoroughly before being used, and laws that would regulate what kind of information about an individual could be combined with other information about the same individual. The authors of the Harris analysis observed that:

> Particularly striking is the pervasiveness of support for tough new ground rules governing computers and other information technology. Americans are not willing to endure abuse or misuse of information, and they overwhelmingly support action to do something about it. This support permeates all subgroups in society and represents a mandate for initiatives in public policy.[62]

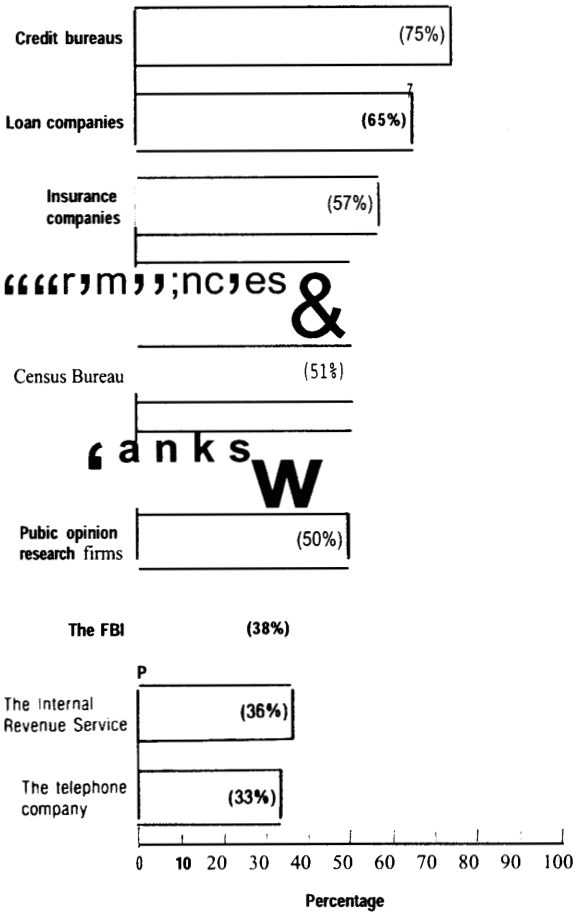
## Finding 4

The Courts have not developed clear and consistent constitutional principles of information privacy, but have recognized some legitimate

"Roper 82.6, June 5-12, 1982.
"Roper 82.8, August 14-21, 1982.

"Harris, op. cit., 1983, P. 41"

**Figure 3.—Percent of Public That Believes Each Agency "Shares" Information About Individuals With Others***




*Response to "Now I'd like to read you a list of organizations which might have a lot of information about individuals. For each, tell me if you think they do have a lot of information but treat it as strictly confidential, have information and probably share it with others, or don't really have information that people ought to be concerned about whether they share it or not."

SOURCE Louis Harris & Associates, Inc., The *Road After 1984: A Nationwide Survey of the Public and Its Leaders on the New Technology and Its Consequences for American Life* (conducted for the Southern New England Telephone for presentation at the Eighth International Smithsonian Symposium, December 1983).

expectations of privacy in personal communications.

Although a "right to privacy" is not mentioned in the Bill of Rights, the Supreme Court has protected various privacy interests. The Court has found sources for a right of privacy in the first, third, fourth, fifth, and ninth amendments. Since the late 1950s, the Supreme Court has upheld a series of privacy interests under the first amendment and due process clause, for example, "associational privacy,"[63] "political privacy,"[64] and the "right to anonymity in public expression."[65] The fourth amendment protection against "unreasonable searches and seizures" also has a privacy component. In Katz v. *United States,* the Court recognized the privacy interests that protected an individual against electronic surveillance. But the Court cautioned that:

> the Fourth Amendment cannot be translated into a general constitutional "right to privacy." That Amendment protects individual privacy against certain kinds of governmental intrusion, but its protections go further and often have nothing to do with privacy at all. Other provisions of the constitution protect personal privacy from other forms of governmental invasion.[66]

The fifth amendment protection against self-incrimination involves a right to privacy against unreasonable surveillance or compulsory disclosure.[67]

Until *Griswold v. Connecticut,* 381 U.S. 479 (1965), any protection of privacy was simply viewed as essential to the protection of other more well-established rights. In *Griswold,* the Court struck down a Connecticut statute that prohibited the prescription or use of contraceptives as an infringement on marital privacy. Justice Douglas, in writing the majority opinion, viewed the case as concerning "a relationship lying within the zone of privacy created by several fundamental constitutional guarantees," i.e., the first, third, fourth, fifth and ninth amendments, each of which creates "zones" or 'penumbras' of privacy. The majority supported the notion of an independent right of privacy inhering in the marriage relationship. Not all agreed with Justice Douglas as to its source; Justices Goldberg, Warren, and Brennan preferred to lodge the right under the ninth amendment.

[63] *NAACP v. Alabama,* 357 U.S. 449 (1958).

[64] *Watkins v. United States,* 354 U.S. 178 (1957), and *Sweezy* v. New *Hampshire,* 354 U.S. 234 (1957).

[65] *Talley* v. *Cab-form-a,* 362 U.S. 60 (1960).

[66] *Katz v. United States,* 389 U.S. 347, 350 (1967).

[67] See *Escobedo* v. *Illinois,* 378 U.S. 478 (1964), *Miranda v. Arizona,* 384 U.S. 436 (1966); and *Schmerber v. California,* 384 U.S. 757 (1966).

**Table 5.—Support for Potential Federal Laws on Information Abuse***

| | Total public | Leaders: Congressmen and top aides | Leaders: Corporate executives | Leaders: Media: science editors | Leaders: Superintendents of schools |
|---|---|---|---|---|---|
| Base | 1,256 | 100 | 100 | 100 | 100 |
| A Federal law that would require that any information from a computer that might be damaging to people or organizations must be double-checked thoroughly before being used: | | | | | |
| Favor | 92% | 85% | 72% | 94% | 94% |
| Oppose | 7 | 12 | 26 | 5 | 5 |
| Federal laws that would make it a criminal offense if the privacy of an individual were violated by an information-collecting business or organization: | | | | | |
| Favor | 83 | 80 | 79 | 94 | 88 |
| Oppose | 14 | 10 | 17 | 5 | 12 |
| A Federal law that would call for the impeachment of any public official who used confidential information to violate the privacy or take away the freedom of an individual or a group of individuals without a proper court order or a court trial: | | | | | |
| Favor | 81 | 69 | 89 | 85 | 91 |
| Oppose | 17 | 26 | 10 | 15 | 8 |
| Federal laws that would require punishment for those in authority responsible for computer mistakes, such as mistakes that hurt people's credit ratings, harm companies, or endanger lives: | | | | | |
| Favor | 71 | 53 | 37 | 69 | 61 |
| Oppose | 25 | 41 | 61 | 25 | 37 |
| Federal laws that could put companies out of business which collected information about individuals and then shared that information in a way that violated the privacy of the individual: | | | | | |
| Favor | 68 | 65 | 78 | 78 | 77 |
| Oppose | 30 | 27 | 20 | 20 | 21 |
| Federal regulations on just what kind of information about an individual could be combined with other information about the same individual: | | | | | |
| Favor | 66 | 77 | 65 | 81 | 87 |
| Oppose | 28 | 18 | 31 | 16 | 13 |

*Response to Would you favor or oppose (READ EACH ITEM)

SOURCE Louis Harris & Associates, Inc., *The Road After 1984 A Nationwide Survey of the Public and Its Leaders on the New Technology and its Consequences for American Life* (conducted for the Southern New England Telephone for presentation at the Eighth International Smithsonian Symposium December 1983)

In *Eisenstadt v. Baird, 405* U.S. *438 (1972),*[68] the Court extended the right to privacy beyond the marriage relationship to lodge in the individual:

> If the right of the individual means anything, it is the right of the *individual,* married or single, to be free from unwarranted governmental intrusion into matters so fundamentally affecting a personas the decision whether to bear or beget a child.

[68] In which the Court struck down a Massachusetts law that made it a felony to prescribe or distribute contraceptives to single persons.

*Roe v. Wade,* 410 U.S. 113 (1973),[69] further extended the right of privacy "to encompass a woman's decision whether or not to terminate her pregnancy." The Court argued that the right of privacy was "founded in the Fourteenth Amendment's concept of personal liberty and restrictions upon state action. The District Court had argued that the source of the right was the ninth amendment reservation of right to the people.

[69] In which the Court struck down the Texas abortion statute.

In the earliest case that raised the issue of the legitimate uses of computerized personal information systems, the Court avoided the central question of whether the Army's maintenance of such a system for domestic surveillance purposes "chilled' the first amendment rights of those whose names were contained in the system.[70] In two cases decided in 1976, the Court did not recognize either a constitutional right to privacy that protected erroneous information in a flyer listing active shoplifters" or one that protected the individual's interests with respect to bank records.[72] In *Paul v. Davis,* the Court specified areas of personal privacy considered "fundamental":

> matters relating to marriage, procreation, contraception, family relationships, and child rearing and education.[73]

Davis' claim of constitutional protection against disclosure of his arrest on a shoplifting charge was 'far afield from this line of decisions" and "we decline to enlarge them in this manner. "7[4] In *United States v. Miller,* the Court rejected Miller's claim that he had a fourth amendment reasonable expectation of privacy in the records kept by banks "because they are merely copies of personal records that were made available to the banks for a limited purpose, " and ruled instead that "checks are not confidential communications but negotiable instruments to be used in commercial transactions. "[75]

In *Whalen v. Roe,* the Court for the first time recognized a right of information privacy, noting that the constitutionally protected "zone of privacy" involved two kinds of interests—"One is the individual interest in avoiding disclosure of personal matters, and another is the interest in independence in making certain kinds of important decisions. "7[6] In this case, a unanimous Court upheld a New York law requiring the State to maintain computerized records of prescriptions for certain drugs, because "the New York program does not, on its face, pose a sufficiently grievous threat to either interest to establish a constitutional violation. "[77] The Court held that as long as the security of a computer is adequate and the information is only passed to appropriate officials, sensitive information may be stored and retrieved without an invasion of a person's right to privacy. In another case in 1977,[78] the Court used a test similar to the one developed in *Whalen,* i.e., balancing the extent of the privacy intrusion against the interests that the intrusion advanced, holding that:

> In sum, appellant has a legitimate expectation of privacy in his personal communications. But the constitutionality of the Act must be viewed in the context of the limited intrusion of the screening process, of appellant's status as a public figure, of this lack of any expectation of privacy in the overwhelming majority of the materials, of the important public interest in preservation of the materials, and of the virtual impossibility of segregating the small quantity of private materials without comprehensive screening.[79]

The court did reaffirm that one element of privacy is "the individual interest in avoiding disclosure of personal matters. "8°

In subsequent lower court cases involving the question of information privacy, the circuit courts have not uniformly followed *Whalen v. Roe.*[81] For example, the Seventh and Ninth Circuit Courts have used autonomy interests rather than informational privacy in-

[70] *Laird v. Tatum* 408 U.S. 1 (1972).

[71] *Paul v. Davis* 424 U.S. 693 (1976).

[72] *United States v. Miller* 425 U.S. 435 (1976).

[73] *Paul v. Davis,* 424 U.S. 693, 713 (1976).

[74] Id. at 713.

[75] *U.S. v. Miller,* 425 U.S. 435, 442 (1976). In response to this decision, Congress passed the Right to Financial Privacy Act of 1978 (PublicLaw 95-630) providing bank customers with some privacy regarding records held by banks and other financial institutions and providing procedures whereby Federal agencies can gain access to such procedures.

[76] *Whalen v. Roe* 429 U.S. 589, 599-600 (1977).

[77] Id. at 600.

[78] *Nixon v. Administrator of General Services, 433* U.S. 425, in which the Court upheld a Federal law that required the national archivists to examine written and recorded information accumulated by the President. Nixon challenged the act's constitutionality on the grounds that it violated his right of privacy.

[79] Id. at 465.

[80] Id. at 457.

[81] See Gary R. Clouse, "The Constitutional Right to Withhold Private Information, " *Northwestern University Law Review,* vol. 77, 1982, p. 536.

terests as the basis for their rulings .82 In *McElrath v. Califano,* the Seventh Circuit Court reiterated that the constitutional right to privacy extends only to those personal rights deemed "fundamental" or "implicit in the concept of ordered liberty, " and that "the claim of the appellants to receive welfare benefits on their own informational terms does not rise to the level of a constitutional guarantee. "[83] In *St. Michael's Convalescent Hospital v. California,* the Ninth Circuit Court ruled that:

> As in Paul v. *Davis,* their [appellants] claim is not based upon any contention that the public disclosure of the cost information will "restrict [their] freedom of action in a sphere contended to be private. " We conclude that no cognizable constitutional right of privacy is implicated here. *4

In 1980, the Third Circuit used *Whalen* to uphold the National Institute for Occupational Safety and Health's request that an employer produce certain medical records of its employees." The Court ruled that:

> The privacy interest asserted in this case falls within the first category referred to in *Whalen v. Roe,* the right not to have an individual's private affairs made public by the government. There can be no question that an employee's medical records, which may contain intimate facts of a personal nature, are well within the ambit of materials entitled to privacy protection.[86]

In a 1981 case involving the compilation and disclosure of juveniles' social histories, the Sixth Circuit explicitly addressed the question of the relationship between *Paul v. Davis* and *Whalen v. Roe,* stating that:

> We do not view the discussion of confidentiality in *Whalen v. Roe* as overruling *Paul v. Davis* and creating a constitutional right to have all government action weighed against the resulting breach of confidentiality. The Supreme Court's discussion makes reference to only two opinions—Griswold v. *Connecticut, supra* in which the court found that several of the amendments have a privacy penumbra, and *Stanley v. Georgia, supra,* a first amendment case—neither of which support the proposition that there is a general right to nondisclosure.[87]

The Sixth Circuit Court went on to state that:

> . . . absent a clear indication from the Supreme Court we will not construe isolated statements in *Whalen* and *Nixon* more broadly than their context allows to recognize a general constitutional right to have disclosure of private information measured against the need for disclosure.[88]

The Supreme Court has not yet accepted a case to clarify the meaning and breadth of *Whalen.*

*Wee: McElrath v. Califano.* 615 F.2d 434 (7th Cir. 1980) which upheld Federal and State regulations that require all family members to disclose their social security numbers as a condition for receiving Aid to Families With Dependent Children benefits; and *St. Michael Convalescent Hospital v. California, 643* F.2d 1369 (9th Cir. 1981) which upheld a California statute requiring that all health care providers who are reimbursed through the Medi-Cal program release their cost information to the public.

[83] *McElrath v. Califano,* 615 F.2d 434,441 (7th Cir. 1980).

[84] *St. Michael Convalescent Hospital v. California, 643* F.2d 1369, 1375 (9th Cir. 1981).

[85] *United States v. Westinghouse, 638* F.2d 570 (3d Cir. 1980).

[86] Id. at 577.

[87] *J.P. v. DeSanti,* 653 F.2d 1080, 1089 (6th Cir. 1981).

[88] Id.

Chapter 3

# Computer Matching To Detect Fraud, Waste, and Abuse

# Contents


| | Page |
|---|---|
| Summary | 37 |
| Introduction | 38 |
| Background | 40 |
| Technology | 40 |
| Policy History | 41 |
| Findings | 43 |
| Finding 1 | 43 |
| Finding 2 | 46 |
| Finding 3 | 50 |
| Finding 4 | 52 |
| Finding 5 | 53 |
| Finding 6 | 55 |
| Finding 7 | 57 |
| Finding 8 | 58 |
| Finding 9 | 59 |
| Finding 10 | 61 |
| Finding 11 | 62 |


Tables


| Table No. | Page |
|---|---|
| 6. Project Match Information Disclosures | 42 |
| 7. Statutes Authorizing Specific Computer Matches | 46 |
| 8. Computer Matches Reported to the PCIE Long-Term Computer Matching Project | 48 |
| 9. Computer Matching Programs Reported toot | 49 |
| IO. Examples of Cost/Budget Analyses | 52 |
| ll. Costs and Benefits of Wage Matching | 52 |
| 12. Estimated Costs and Benefits of Computer Matching in Four Sites | 52 |


Figure


| Figure No. | Page |
|---|---|
| 4. Computer **Matches Conducted** From April 1980 to April 1985 | 49 |

Chapter 3

# Computer Matching To Detect Fraud, Waste, and Abuse

## SUMMARY

Computer matching involves the comparison of two or more sets or systems of computerized records to search for individuals who may be included in more than one file. Matching can be done manually with paper files. But, as a practical matter, time and cost requirements make manual matching prohibitive in cases involving a large number of records. The primary impetus for Federal and State use of computer matching is to detect fraud, waste, and abuse in government welfare and social service programs. However, computer matching has broad applicability to government programs and activities.

Computer matching has the potential to improve the efficiency of government recordkeeping and management of government programs. It is widely used by many States and foreign countries, the private sector, and increasingly by the Federal Government, where the technique is strongly supported by the Office of Management and Budget (OMB) and the inspectors general, among others, and has been endorsed in several public laws.

However, a number of problems have been identified in Federal computer matching activities, including weak oversight, little persuasive evidence or documentation of cost-effectiveness, widely variable record quality, and little consideration of the implications for privacy and civil liberties.

In computer matching, the basic policy conflict is between the efficient management of government programs (including effective law enforcement) and the rights of individuals. The fourth amendment protects "persons, houses, papers, and effects" against unreasonable government searches and seizures. The Privacy Act of 1974 requires that information collected for one purpose not be used for another purpose, unless, among other exemptions, it falls within a "routine use. Under OMB guidelines, personal information used in computer matches can be disclosed under the routine use exemption.

OTA'S assessment of computer matching technology and policy issues found that:

- Although Congress has legislated general and specific restrictions on agency disclosure of personal information, it has also endorsed computer matching and other record linkages in various programmatic areas specified in several public laws. Thus, congressional actions appear to be contradictory.
- It is difficult to determine how much computer matching is being done by Federal agencies, for what purposes, and with what results. However, OTA estimates that in the 5 years from 1980 to 1984, the number of computer matches nearly tripled.
- As yet, nG firm evidence is available to determine the costs and benefits of computer matching and to document claims made by OMB, the inspectors general, and others that computer matching is cost-effective.
- The effectiveness of computer matches used to detect fraud, waste, and abuse can be compromised by inaccurate data.
- There are numerous procedural guidelines for computer matching, but little or no oversight, follow-up, or explicit consideration of privacy implications.
- As presently conducted, computer matching programs may raise several constitutional questions, e.g., whether they violate protection against unreasonable search and seizure, due process, and equal pro-

tection of the laws. But, as presently interpreted by the courts, the constitutional provisions provide few, if any, protections for individuals who are the subjects of matching programs.

- The Privacy Act as presently interpreted by the courts and OMB guidelines offers little protection to individuals who are the subjects of computer matching.
- The courts have been used infrequently as a forum for resolving individual grievances over computer matching, although some organizations have brought lawsuits.
- Computer matches are commonly conducted in most States that have the computer capability. At least four-fifths of the States are known to conduct computer matches, most in response to Federal directives.
- All Western European countries and Canada are using computer matching or record linkages, to an increasing degree, as a technique for detecting fraud, waste, and abuse.
- In designing policy for computer matching, consideration of the following factors is important:
  - which records to make available for computer matches and for what purposes,
  - approval required before a match takes place,
  - notice to individuals,
  - whether to require a cost-benefit analysis,
  - verification of hits, and
  - appropriate action to be taken against an individual who has submitted false information.

In response to the OTA survey of Federal agencies, OTA determined that:

- Forty-three percent of agency components that reported participation in computer matching activities (16 out of 37) said that the matches were required or authorized by legislation.
- Eleven cabinet-level departments and four independent agencies carried out a total of 110 matching programs, with a total of 553 matches conducted from 1980 to April 1985.
- In the 5 years from 1980 to 1984, the number of computer matches nearly tripled.
- For 20 percent of the matches reported, information was available on the number of records matched, number of hits, and percent of hits verified.
- Despite the low percentage of respondents providing information on reported matches, the number of separate records used in the reported matching programs totaled over 2 billion; the total number of records matched was reported to be over 7 billion due to multiple matches of the same records.
- The percentage of hits (i.e., matches between the specific items of interest in two different records) verified to be accurate ranged from 0.1 to 100 percent.
- Sixty-eight percent (25 of 37) of the agencies indicating that they participated in matching programs said that procedures were used to ensure that the subject record files contain accurate information.

# INTRODUCTION

Computer matching involves the electronic comparison of two or more sets or systems of personal records.[1] Matching is used to check for individuals who should not appear in two systems of records, as in the case of Federal employees above a certain salary level and persons receiving food stamps. Matching can also be used to locate individuals who should appear in two systems of records but do not; for example, males registered for the draft and males over the age of 18 with driver's licenses. Although manually comparing the contents of

[1] The Office of Management and Budget (OMB) Guidelines, issued May 11, 1982, define computer matching as "a procedure in which a computer is used to compare two or more automated systems of records or a system of records with a set of non-Federal records to find individuals who are common to more than one system or set."

two record systems is a traditional audit technique, this practice becomes prohibitive when dealing with massive record systems that are not uniformly comparable with other record systems. Computers greatly facilitate such comparisons.

Because of the number of people who may be subject to computer matching and because it can be done without their knowledge, computer matching has raised a number of policy questions. The basic conflict is between the efficient management of government programs and the rights of individuals.

It is well known that government programs are subject to fraud, waste, and abuse. Although the problem is not peculiar to welfare programs, fraud and waste in these programs have been particularly well documented. For example, the General Accounting Office (GAO) reviewed improper payments for fiscal year 1978-79 in 5 of the 58 federally supported welfare programs, and estimated that Federal and State welfare agencies spent about $867 million on erroneous welfare payments because recipients had not properly reported their income and assets.[2]

Since 1977, computer matching has been used extensively by a number of Federal departments and State agencies. Some specific examples of matching include:

1. recipients of Aid to Families With Dependent Children (AFDC) matched with the Social Security Administration's earnings record,
2. the Veterans Administration's rolls matched with the supplemental security income (SS1) benefit rolls,
3. AFDC recipients matched with Federal civilian and military payrolls, and
4. State AFDC rolls matched with other State AFDC rolls.

In general, matching is used to detect unreported income, unreported assets, duplicate benefits, incorrect social security numbers, overpayments, ineligible recipients, incongruous entitlements (SS1 checks mailed to deceased individuals, mothers claiming more children than exist), present addresses of individuals (Parent Locator Service, Student Loan defaulters), and providers billing twice for the same service.

In order to facilitate computer matching, a number of computerized databanks have been created solely for matching purposes. One example is the Medicaid Management Information System that contains information on recipient records, provider data, and claims-processing information.[3] A proposed computerized databank is the Internal Revenue Service (IRS) Debtor Master File that will contain the names of all delinquent Federal borrowers to match against tax returns.[4]

A central policy issue is whether and under what conditions the use of computer matching is appropriate, given the rights of individuals who are the subjects of matching and given the possible long-term societal effects of general electronic searches, as elaborated below.

As discussed in chapter 2, public opinion polls indicate that Americans value their privacy and generally expect that activities in one area of their lives are kept separate from those in other areas. In the 1983 Harris Survey, most Americans (from two-thirds to three fourths) responded that agencies that release the information they gather to other agencies or individuals are seriously invading personal privacy.[5] Two-thirds or more of Americans surveyed believed that the following government information practices would entail a "serious invasion of privacy"—the IRS not keeping individual tax records confidential (84 percent perceived this as a serious invasion); the Fed-

[2]U.S. General Accounting Office, "Legislative and Administrative Changes To Improve Verification of Welfare Recipients Income and Assets Could Save Hundreds of Millions," HRD-82-9, Jan. 14, 1982.

[3]U.S. Department of Health and Human Services. Health Care Financing Administration, "Medicare and Medicaid Data Book," 1982.

[4]Judith A. Sullivan, "IRS To Create Debtor File," *Government Computer News*, Nov. 8, 1985, pp. 1, 70.

[5]Louis Harris& Associates, Inc., *The Road After 1984: A Nationwide Survey of the Public and Its Leaders on the New Technology and Its Consequences for American Life*, (conducted for Southern New England Telephone for presentation at The Eighth International Smithsonian Symposium, December 1983), table 1-6.

eral Bureau of Investigation not keeping information about individuals confidential (82 percent viewed as serious invasion); and the Census Bureau not keeping information about individuals confidential (73 percent viewed as serious invasion). Yet, in a 1979 survey, 87 percent of respondents believed that government agencies were justified in using computers to check welfare rolls against employment records to identify people claiming benefits to which they are not entitled. However, they were less supportive (68 percent) of the IRS use of matching to check tax returns against credit card records.[6]

Public opinion polling results suggest that Americans recognize that a balance must be struck between individual rights and the protection of society. A majority of the public believes that there are some costs in terms of privacy that must be paid in order to have a more lawful society. In response to the statement: "In order to have effective law enforcement, everyone should be prepared to accept some intrusion into their personal lives, " 57 percent agreed and 36 percent disagreed.[7] Public opinion research also indicates that Americans have certain expectations about the scale of government monitoring activities. Americans assume that government investigations are predicated on evidence of individual wrongdoing and that procedural standards and safeguards exist for investigative behavior. The public overwhelmingly believes the police should not be able to tap the telephones of members of suspicious organizations without obtaining a court order. A large majority of the public is concerned about protecting records from examination by public authorities without a court order. Over 80 percent of the public believes that the police should not be able to examine the bank records of suspicious individuals without a court order.[8]

Computer matches can also conflict with the expectation of being treated as an individual. Computer matches are inherently mass or class investigations, as they are conducted on a category of people rather than on specific individuals. In theory, no one is free from these computer searches; in practice, welfare recipients and Federal employees are most often the targets.

[6]Louis Harris & Associates, Inc., and Alan F. Westin, *The Dimensions of Privacy: A National Opinion Research Survey of Attitudes Toward Privacy* (conducted for Sentry Insurance, 1979), table 9.3.

[7]Ibid., table 2.2.

[8]Ibid., table 8.3.

# BACKGROUND

## Technology

In conducting a computer match, one computer file is compared with another using software that instructs the computer to search for certain patterns, e.g., duplicate social security numbers, same names, identical addresses. Before a match is conducted, agency personnel need to determine whether the relevant data are formatted in a similar fashion on the two or more systems being matched. If not, then the data need to be reformatted or the software must be designed to take the differences into account.

Files can be compared either by using computer tapes of the record systems or by direct electronic linkages of computers. At the present time, the matching of tapes is the procedure commonly used. However, as systems become more compatible and costs drop, direct electronic linkages between/among systems are likely to increase.

During the match, computer files are compared on the basis of a specified data element as an identifier, generally the social security number. Experience from early computer matches suggested that social security numbers were often inaccurate. In order to ensure the effectiveness of a computer match, a search for erroneous social security numbers can be conducted before the match. Additionally, the

identifier used for the match can be the social security number plus another data element, such as the first few letters of a last name.

The social security number is not essential to computer matches as databases can also be searched for combinations of selected factors; however, a unique identifier makes matching far easier. In 1981, congressional legislation required that every member of a household receiving food stamps must have a social security number. Such a requirement makes matching more efficient because it is easier to identify duplicate or fraudulent recipients.

The resulting match produces information on individuals who are common to the two files; for example, an individual who has not repaid a Federal student loan may also be a Federal employee, or a physician may have billed Medicaid twice for the same service. Once the match has identified the files having duplicate or similar information, these files are considered "hits." The hits must then be verified manually to determine whether the same individual is really involved and whether there is cause to believe that the individual has committed fraud.

## Policy History

In the early 1970s, a few States began to use computer matching to check AFDC recipients against wage information from the State Employment Security agencies. The first major computer match at the Federal level was Project Match, announced in November 1977 by Joseph Califano, Secretary of the Department of Health, Education, and Welfare (HEW). Project Match compared computer tapes of welfare rolls and Federal payroll files in 18 States, New York City, the District of Columbia, and parts of Virginia. The goal was to detect government employees who were fraudulently receiving AFDC benefits. Privacy advocates in Congress, members of the Privacy Protection Study Commission, the American Civil Liberties Union, and others criticized the proposed match as a "fishing expedition."

There were disputes within the general counsel's office at HEW regarding the legal implications of conducting these matches, especially in light of the Privacy Act "routine use' provisions.' There were also disputes between HEW and the Civil Service Commission (CSC) and the Department of Defense (DOD), neither of which wanted to release its tapes because of the routine use provision.[10] The general counsel at CSC raised two concerns regarding the compatibility of the proposed match with the routine use provision of the Privacy Act: first, "it is evident that this information on employees was not collected with a view toward detecting welfare abuses," and second, "that disclosure of information about a particular individual at this preliminary stage is (not) justified by any degree of probability that a violation or potential violation of law has occurred."[11] CSC and DOD eventually released their tapes to HEW—CSC justifying the transfer on the argument that HEW could get the information under the Freedom of Information Act if it so chose, and DOD justifying the transfer as a new 'routine use' under the Privacy Act. HEW lawyers, themselves, were additionally concerned that the results of the match would need to be transferred to the employing departments for verification, which would also raise Privacy Act issues. As table 6 indicates, it was possible to justify under existing law all record transfers required by Project Match.

While Project Match was under way, an interagency advisory group of Federal personnel officials questioned whether Federal employees should be notified under the Privacy

[9]See Jake Kirchner, "Privacy-A History of Computer Matching in the Federal Government," *Computerworld*, Dec. 14, 1981, pp. 1-16. Section 3b of the Privacy Act establishes the conditions under which an agency can disclose personal information to another party without the prior consent of the individual. One of these conditions of disclosure is "for a routine use," defined as "the use of such record for a purpose which is compatible with the purpose for which it was collected" [3(a)(7)]. All routine uses are to be published in the *Federal Register*, including "the categories of users and the purpose of such use" [3(e)(4)(D)].

[10]For correspondence, see Kirchner, Op. cit., and pp. 122-125 of U.S. Congress, Senate, Hearings Before the Senate Subcommittee on Oversight of Government Management, Committee on Governmental Affairs, *Oversight of Computer Matching To Detect Fraud and Mismanagement in Government Programs* (Washington DC: U.S. Government Printing Office, Dec. 15-16, 1982) [hereafter referred to as the Cohen hearings].

[11]See Cohen hearings, op. cit., p. 123.

**Table 6.—Project Match Information Disclosures**

| Disclosure | Justification |
|---|---|
| **Health, Education, and Welfare Department disclosure of social security number and birth dates to other agencies** | **Exception in Privacy Act** |
| **Office of Personnel Management disclosure to Health,** Education, and Welfare Department | **Public interest outweighs personal privacy outlined** in the Privacy Act and information could be obtained under the Freedom of Information Act |
| Defense Department disclosure of military personnel on active duty to Health, Education and Welfare Department | Exception under "routine use" of the Privacy Act |
| State government disclosure of State Aid to Families With Dependent Children (AFDC) rolls to Health, Education, and Welfare Department | Privacy Act does not apply to States; no Federal law barring such disclosure |
| State government disclosure of State AFDC rolls to Federal employer agencies | New "routine use" published in the Federal Register based on original routine uses |
| Agencies disclosure of annotated work sheets to the Health, Education, and Welfare Department | HEW Inspector General Statute requiring agencies to respond to information requests by Inspector General |
| Agencies disclosure of civil or criminal proceedings to Health, Education, and Welfare Department | Exception in Privacy Act |
| Health, Education, and Welfare Department disclosure to State or local agencies | Exception in "routine use" of Privacy Act to assist States and localities enforce violated statutes |
| Agencies refer information and case to Department of Justice when lawbreaking is suspected | Exception under "routine use" or law enforcement exception of the Privacy Act |
| Agencies referral of cases to other agencies when lawbreaking is suspected or for investigation of government employees | For administrative action authorized by the "routine uses" of Privacy Act |

SOURCE Kenneth James Langan, "Computer Matching Programs A Threat to Privacy" *Columbia Journal of Law and Social Problems, VOl. 15, No 2, 1979,* pp. 149-150

Act of the record transfers. The Department of Justice argued against notification, saying, "We view Project Match as a law enforcement program, designed to detect suspected violations of various criminal statutes in (government) operations." 12 Opponents of the match pointed out that such a view was hardly consistent with the "routine use" concept.[13] By March 1978, Project Match had identified 7,100 employees who were possibly ineligible for welfare. But, it had also generated so much information that agency officials could not follow up adequately to determine the validity of that information.[14]

After Project Match was completed, Secretary Califano advocated more Federal use of matching and tried to access private sector company files. This increased public pressure for justification of matching under the Privacy Act, and OMB and the Carter White House began to take a more active role in the process. In late 1977, OMB sent a letter to Representative Richardson Preyer to explain the Administration's justifications for Project Match, concluding that 'the requirement of compatible purpose in the routine use is difficult and is ultimately largely a matter of judgment."5

While Project Match was being run, the White House was concurrently conducting its Privacy Initiative, following the 1977 report of the Privacy Protection Study Commission. The conflict between the goals of the Privacy Initiative and Project Match was not ignored within the White House, but remained unresolved. In response to concerns about Project Match's privacy implications, OMB took on the task of writing guidelines for computer matching, with input from the President Office of Telecommunications Policy and the White House Privacy Initiative.

In 1979, Congress required States to conduct wage matching for AFDC recipients. Because

"Kirchner, op. cit., p. 7.

[13]See testimony of John Shattuck of the American Civil Liberties Union, Cohen hearings, op. cit., p. 80.

[14]Laura B. Weiss, "Government Steps Up Use of Computer Matching To Find Fraud in Programs, " Congressional *Quarterly Weekly Report,* Feb. 26, 1983, p. 432.

"Kirchner, op. cit., p. 10.

computer matching was perceived as an efficient tool for managing benefit programs, States increasingly began to use it for a number of programs and with a number of sources, including private institutions such as employers and banks. One of the largest and best publicized of the State efforts occurred in Massachusetts in 1982 when welfare recipients were matched against bank records, identifying about 600 people who had bank accounts larger than regulations allowed. About 160 of those persons identified received termination notices. But for more than 110 of these 160 persons, the identification based on the computer match was later determined to be based on erroneous information, e.g., inaccurate social security number or bank account for burial expenses held in trust. '6

Since 1979, concern about the size and efficiency of the Federal Government and the increase in the Federal deficit has made management a policy priority for both Congress and the executive branch. One effect has been to encourage the use of computer matching, especially as a technique to detect fraud, waste, and abuse. In 1981, President Reagan established the President's Council on Integrity and Efficiency (PCIE), chaired by the Deputy Director of OMB, to enhance interagency efforts to reduce fraud and waste, and to give the inspectors general a direct link to the President. PCIE projects include: 1) a long-term computer matching project; 2) Project Clean Data (i.e., standardization of data elements); and 3) an inventory of State computer matching software packages. President Reagan has also formed the President's Council on Management Improvement, composed of the senior management official from each major department and agency (including central management agencies—OMB, the General Services Administration, and the Office of Personnel Management), the Assistant to the President for Policy Development, and the Assistant to the President for Presidential Personnel. Its purpose is to advise the President and to oversee agency implementation of management reforms.

In 1982, President Reagan established the President Private Sector Survey on Cost Control, popularly known as the Grace Commission, to study management problems in government. Its major finding was "that the Federal Government has significant deficiencies from managerial and operating perspectives, resulting in hundreds of billions of dollars of needless expenditures . . . "7 There have been criticisms of the Grace Commission's cost figures and its methodology .18 In 1982, the Reagan Administration also announced Reform '88, a program to increase efforts to reduce waste, fraud, and abuse, and to restructure the management and administrative systems of the Federal Government.

[16]Ross Gelbspan, "Computer Matching Stirs Up Criticism," *Boston Globe*, June 9,1985, p. A 1, cont. A4.

[17]Ellen Law, "Grace Reports To the President," *Government Computer News*, March 1984, p. 4.

"Steven Kelman, "The Grace Commission: How Much Waste in Government?" *The Public Interest, No.* 78, winter 1985, pp. 62-82.

# FINDINGS

## Finding 1

Although Congress has legislated general and specific restrictions on agency disclosure of personal information, it has also endorsed computer matching and other record linkages in various programmatic areas specified in several public laws. Thus, congressional actions appear to be contradictory.

As discussed in chapter 2, Congress has passed a number of laws that give an individual certain rights with respect to controlling the use of personal information, and that place restrictions on the ways in which agencies may legitimately use such information. These laws speak both to general agency practices (e.g., the Privacy Act of 1974) and to the practices of specific agencies, (e.g., Section 6103 of the Tax Reform Act of 1976).

Congress has also legislated a number of exchanges of information among agencies. Congressional concern with detecting fraud, waste, and abuse has resulted in several major legislative endeavors that have been viewed as authorizing computer matching. First is the establishment of inspectors general offices in a number of Federal agencies to identify and reduce fraud, waste, and abuse, and to identify and prosecute perpetrators (Public Law 94-452, Public Law 94-505, Public Law 97-252). The Departments of Health and Human Services, Energy, Defense, and 15 other Federal agencies have inspectors general. The inspectors general are potentially very powerful officers who:

> . . . have complicated reporting relationships involving department and agency heads, and Congress and its many committees. IGs can bypass department/agency general counsels and take matters directly to the Criminal Division of the Justice Department. They can initiate audits and investigations at any time, which can cover fraud, abuse, and any and all management deficiencies. [19]

Inspectors general employ a variety of techniques, including: 1) vulnerability assessments to assess the risk of loss in programs, 2) management control guides, 3) fraud bulletins and memos, 4) fraud control training, 5) hotlines for reports of wrongdoing, and 6) audit follow-up procedures. Matching, profiling, and front-end verification are used by inspectors general.

A second legislative endeavor that is perceived as encouraging data-sharing among agencies is the Paperwork Reduction Act of 1980 (Public Law 96-51 1), which gives OMB Federal information oversight authority and the responsibility y to promote the effective use of information technology. It establishes an Office of Information and Regulatory Affairs within OMB to carry out the purposes of the act, oversee agency compliance, and set up a Federal Information Locator System to register all information collection requests. OMB Circular A-130 was issued in December 1985 as an integrative policy statement on information resource management policies, including privacy and matching.[20]

A statute that may encourage the sharing of information within an agency is the Federal Managers Financial Integrity Act of 1982 (Public Law 97-255), which requires periodic evaluations of and reports on agency systems of internal control and action to reduce fraud, waste, abuse, and error. OMB Circular A-123 (October 28, 1981) complements the act by mandating an improvement in internal control systems, including a requirement that agency heads issue specific internal control directives and review plans for all components of their agencies. Inspectors general have the responsibility to review directives. OMB Assistant Director Wright and Comptroller General Bowsher have pledged that:

> OMB and GAO plan to work together very closely in implementing the Act and in assuring that the momentum already built up within the agencies for improved internal control is sustained.[21]

A fourth statute that encourages exchanges of personal information is the Debt Collection Act of 1982 (Public Law 97-365), which establishes a system of data-sharing between Federal agencies and private credit reporting agencies in order to increase the collection of delinquent nontax debts. The act permits agencies to:

1. refer delinquent nontax debts to credit bureaus to affect credit ratings;
2. contract with private firms for collection services;
3. require applicants for Federal loans to supply their taxpayer identification numbers (social security numbers);
4. offset the salaries of Federal employees to satisfy debts owed the government;
5. screen credit applicants against IRS files to check for tax delinquency;

[19] John D. Young, "Reflections On the Root Causes of Fraud, Abuse and Waste in Federal Social Programs," *Public Administration Review, 1983*, p. 366.

[20] Office of Management and Budget, "Management of Federal Information Resources," Circular No. A-130, Dec. 12, 1985.

[21] Office of Management and Budget, "Agencies to Tighten Internal Control Systems," OMB 82-26 (President Task Force on Management Reform), Oct. 8, 1982.

6. turn over to private contractors the mailing addresses of delinquent debtors obtained from IRS;
7. extend from 6 to 10 years the statute of limitations for collection of delinquent debts by administrative offset; and
8. charge interest, penalties, and administrative processing fees on delinquent nontax debts.

The law requires agencies to provide due process to individuals before using any of the newly authorized methods of collection. The law provides safeguards to preserve the confidentiality of taxpayer information, and civil and criminal penalties are included when taxpayer addresses are improperly disclosed. OMB estimates that the improved procedures and newly available tools will result in an additional $500 million in annual collections.[24] OMB has decided that:

> Rather than creating a new bureaucracy to implement the credit reporting provisions of the Debt Collection Act, the existing nationwide network of commercial and consumer credit bureaus will be under contract to provide this service for all departments and agencies.[25]

The statute requiring the most far-reaching data-sharing is the Deficit Reduction Act of 1984 (DEFRA) (Public Law 98-369), which requires the establishment of new State information systems for verification purposes and the use of verification in a number of federally funded State-administered programs. This 1,2 10-page law provides tax reforms and spending reforms, primarily by amending the Social Security Act and Internal Revenue Code. Provisions that are relevant to management and efficiency are in Subtitle C—' Implementation of Grace Commission Recommendations, " Section 2651.

The major changes in the Social Security Act mandated by DEFRA include requiring States or State agencies to: 1) have an income and eligibility system, 2) obligate recipients to supply their social security numbers and require States to use those numbers in the administration of programs, 3) compel employers to keep quarterly wage information, 4) exchange relevant information with other State agencies and with the Department of Health and Human Services, and 5) notify recipients and applicants that information available through the system will be requested and utilized. The programs that must participate in the income verification program are: AFDC; Medicaid; unemployment compensation; food stamps; and any State program under a plan approved under Titles I, X, XIV, or XVI of the Social Security Act. Under DEFRA, no Federal, State, or local agency may terminate, deny, suspend, or reduce any benefits of an individual until such agency has taken appropriate steps to independently verify information.

DEFRA provides certain procedural rights for the individual, including that the agency shall inform the individual of the findings made on the basis of verified information, and give the individual an opportunity to contest such findings. DEFRA makes a number of changes in the Internal Revenue Code, including that the Commissioner of Social Security shall, on request, disclose information on earnings from self-employment, wages, and payments on retirement income to any Federal, State, or local agency administering one of the following programs: AFDC; medical assistance; supplemental security income; unemployment compensation; food stamps; State-administered supplementary payments; and any benefit provided under a State plan approved under Titles I, X, XIV, or XVI of the Social Security Act. Information with respect to unearned income may also be disclosed from the IRS files to the above agencies.

In addition to these broad endorsements of and requirements for computer matches, there are a number of statutes that authorize specific computer matches (see table 7).

Congressional restrictions on agency disclosures of personal information and congres-

[24] Office of Management and Budget, "OMB Announces Progress in Administration's Debt Collection Effort ," OMB 82-32 (Reform '88 Communications), Dec. 15, 1982.

[25] Office of Management and Budget, "Government to Use Credit Bureaus to Cut Delinquent Debts; Delinquency Growth Halted, (OMB 83-29 ( Public Affairs Management), Sept. 23, 1983.

**Table 7.—Statutes Authorizing Specific Computer Matches**

Tax Reform *Act of* 1976, Public Law 94-455, permitted the Department of Health, Education, and Welfare to search the databanks of other Federal agencies to locate parents who fail to pay child support.

*Social Security Amendments of 1977,* Public Law 95-216, required States to use wage data in determining eligibility for Aid to Families With Dependent Children (AFDC) Program benefits by providing them access to earnings information held by the Social Security Administration (SSA) and State employment security agencies.

*Food Stamp Act Amendments of 1977,* Public Law 96-58, granted access to employer-reported wage information for recipients of supplementary security income (SSI) benefits.

*Food Stamp Act Amendments of 1980,* Public Law 96-249, amended the Internal Revenue Code and the Social Security Act to allow State food stamp agencies to obtain and use wage, benefit, and other information in SSA files and those of State unemployment compensation agencies.

*Food Stamp and Commodity Distribution Amendments of 1981,* Public Law 97-98, required States to obtain and use earnings information obtained from employers.

*Department of Defense Authorization Act of 1983,* Public Law 97-252, required the Secretary of Education to prescribe methods for verifying that individuals receiving any grant, loan, or work assistance under Title IV of the Higher Education Act of 1965 had complied with registration as necessary under the Military Selective Service Act.

*Deficit Reduction Act of 1984,* Public Law 98-369, required the Internal Revenue Service (IRS) to disclose information about an individual's unearned income to State welfare agencies and the SSA to verify the income of an applicant or beneficiary of the AFDC, SSI, and food stamp programs. (Presently, IRS is required to disclose only information on earned income.) The Deficit Reduction Act also requires States to maintain a system of quarterly wage reporting as part of its income verification system.

SOURCE Off Ice of Technology Assessment

sional authorizations of computer matching place agencies in a position where the legitimacy of either a disclosure or refusal to disclose can be challenged. A prime example is *Tierney v. Schweiker, 718* F.2d 449 (1983), which involved the Social Security Administration's (SSA) use of confidential tax return information maintained by IRS for purposes of verifying the income and assets of supplemental security income recipients. SSA was acting on its congressional mandate that SSA'S determinations of eligibility be based on "relevant information [that is] verified from independent or collateral sources and additional information [that is] obtained as necessary. "[24]

Two GAO reports[25] recommended that SSA use IRS tax information to verify eligibility. In deciding the case, Judge Abner Mikva recognized that:

> Much of the confusion , . . arises from conflicting signals given by the Congress. In 1972, when enacting the Social Security Amendments that instituted the Benefits program, Congress was concerned with ensuring that financially ineligible individuals not abuse the system. To this end, Congress directed the SSA to obtain as much information as possible to discover such ineligibility. In 1976, when expanding the confidentiality provisions as part of the Tax Reform Act of 1976, Congress made clear that tax information was to be absolutely confidential, subject to certain explicit exceptions. Although Congress created numerous exceptions, none was applicable to the information which SSA now seeks. When Congress speaks with two separate minds, the conflicting goals can present difficult dilemmas.[26]

In response to the OTA survey, 43 percent of agency components that reported participation in computer matching activities (16 out of 37) said that the matches were required or authorized by legislation. However, approximately one-third of the respondents cited general statutes such as an Inspector General Act, the Debt Collection Act, or an Omnibus Reconciliation Act. Another one-third cited explicit requirements for matching, such as the Uniform Code of Child Support or Title 7, U. S. C., chapter 51, "Food Stamp Program. " Another onethird cited more general authorization, e.g., Public Law 96-473, which requires the suspension of benefits for inmates of penal institutions and is given as the basis for matches between inmate records and social security files.

## Finding 2

It is difficult to determine how much computer matching is being done by Federal agencies, for what purposes, and with what results. However, OTA estimates that, in the 5 years from 1980 to 1984, the number of computer matches nearly tripled.

[24] 42 U.S.C. sec. 1383(3)(1)(B) as quoted in *Tierne-y v. Schweiker* 718 F.2d 449, 451 (1983).

[25] U.S. General Accounting Office, HRD 81-4, Feb. 4, 1981 and HRD 82-9, Jan. 12, 1982.

[26] *Tierney v. Schweiker 718* F.2d 449, 454 (1983).

There has been no accurate accounting of the number of matches that have been done at the Federal level. In part, this is a definitional problem. One distinction that affects reports of the amount of computer matching being done is that of "matching programs" versus "matches." The OMB guidelines define a "matching program" as:

> . . . a procedure in which a computer is used to compare two or more automated systems of records or a system of records with a set of non-Federal records to find individuals who are common to more than one system or set. The procedure includes all of the steps associated with the match, including obtaining the records to be matched, actual use of the computer, administrative and investigative action *on* the hits, and disposition of the personal records maintained in connection with the match. It should be noted that a single matching program may involve several matches among a number of participants.[27]

Based on this definition, there will be many more matches than there are matching programs, as one matching program may include a number of record sets (e.g., Office of Personnel Management (O PM) records with SSA records and OPM records with Farmers' Home Administration loans), and/or a matching program may involve a number of matches at certain intervals, e.g., yearly or monthly. However, this distinction between matching programs and matches has not always been recognized in accounts of numbers of computer matches.

A second important distinction in understanding reports on the scale of computer matching by Federal agencies is one made by OMB. Some compilations of computer matching at the Federal level include only those matches that fall under the OMB guidelines, others include both, and still others do not differentiate. OMB'S guidelines state that the following are not matching programs:

1. Matches that do not compare a substantial number of records, e.g., comparison of the Department of Education's Defaulted Student Loan database with the OPM'S Federal Employee database, would be covered; comparison of six individual student loan defaulters with the OPM file would not.
2. Checks on specific individuals to verify data in an application for benefits, done soon after the application is received.
3. Checks on specific individuals based on information that raises questions about an individual's eligibility for benefits or payments, done reasonably soon after the information is received.
4. Matches done to produce aggregate statistical data without any personal identifiers.
5. Matches done to support any research or statistical project where the specific data are not to be used to make decisions about the rights, benefits, or privileges of specific individuals.
6. Matches done by an agency using its own records .28

For the purposes of this report, the first three applications are considered front-end verification and are discussed in chapter 4. The fourth and fifth applications are not relevant to this inquiry. The sixth application does include a significant number of matching programs and matches that are relevant to this discussion, e.g., SSA and another component of the Department of Health and Human Services.

In addition to definitional problems, the rules for reporting matches may not require that all matches be reported. Notices of computer matching programs that meet the criteria in the OMB guidelines may appear in the Federal Register as a new routine use. However, if the agency providing the data believes that the system of records already contains such a use, then no additional notice in the F'ed-end Register is required. No notice is required for records that are matched within an agency.

There have been a number of attempts at determining the scale of computer matching. Figures range from 200 programs on upwards.

[27]Office of Management and Budget, "Privacy Act of 1974; Revised Supplemental Guidance for Conducting Matching Programs," *Federal Register*, vol. 47, No. 97, May 19, 1982, p. 21657.

[28]Ibid., p. 21757.

For example, in 1982 hearings on computer matching, Senator William Cohen estimated that:

> As of January 1982, Federal agencies had completed more than 85 matching programs and State government agencies are now performing approximately 170 matches involving public assistance records, unemployment compensation records, government employee files, and in some cases, the files of private companies. These projects involve the records of hundreds of thousands of citizens.[28]

At the same hearings, Thomas McBride, former Inspector General of the Department of Labor, testified:

> So my guess is we are talking about a population of roughly 500, more or less, routine recurring matches going on, some of them subject to Federal legislative action, some of them not.[30]

The Long Term Computer Matching Project of the President's Council on Integrity and Efficiency has issued three compilations of Federal computer applications to prevent/detect fraud, waste, and abuse. These compilations do not provide complete listings of computer matching programs.[31] They include those computer matches that agencies chose to report; some agencies submitted partial reports, others appear not to have responded at all, or to only one or two of the PCIE'S requests. Some of the reported matches are one time only, others are recurring. The first compilation was distributed in 1982[32] and reported 77 matches; the second was distributed in July 1984 as an expansion and update, and reported 162 matches; and the third was distributed in January 1986 as an update, and reported 108 matches.[33] (See table 8 for breakdown by agency.)

A 1985 GAO study, *Eligibil.z"ty Verification and Privacy in Federal Benefit* Programs: A *Delicate Balance,* reported that:

> Before 1976, only two benefit program-related Federal computer matching projects were conducted. However, recent inventories of Federal and State agencies' computer matching programs show that Federal agencies had initiated 126 benefit-related matches, 38 of which were recurring as of May 1984. State agencies, as of October 1982, had initiated more than 1,200 matching projects, most of them recurring.

[28]Cohen hearings, op. cit., p. 2.

[30]1 bid., p. 20.

[31]It does not appear that the PCIE inventory used the OMB guidelines' definition of computer matching programs. Some agencies reported matches within their agency, e.g., Department of Health and Human Services Black Lung and SSA Title II. Some agencies reported particular matches within a matching program.

[32]None of the compilations is dated. The phrase 'distributed in 1982" is used by PCIE in its second compilation to describe the first compilation.

[33]The low figures in the 1986 compilation can be attributed to two factors. The first is that some large agencies that previously had reported a number of matches did not respond, e.g., Departments of Labor, Defense, and Justice. The second factor is that many agencies have increased their use of computer screens and profiles rather than their use of computer matches. This latter factor will be discussed in ch. 4.

**Table 8.—Computer Matches Reported to the PCIE Long-Term Computer Matching Project**

| | 1982 | 1984 | 1986 |
|---|---|---|---|
| Department of Agriculture | 11 | 10 | 23 |
| Department of Commerce | 0 | 1 | 1 |
| Department of Defense | 0 | 30 | 0 |
| Department of Education | 1 | 1 | 0 |
| General Services Administration | 1 | 1 | 18 |
| Department of Health and Human Services | 29 | 58 | 55 |
| Department of Housing and Urban Development | 0 | 4 | 3 |
| Department of the Interior | 0 | 1 | 0 |
| Department of Justice | 8 | 5 | 0 |
| Department of Labor | 12 | 12 | 0 |
| National Science Foundation | 0 | 2 | 0 |
| Nuclear Regulatory Commission | 0 | 1 | 0 |
| Peace Corps | 0 | 1 | 0 |
| Pension Benefit Guaranty Corp. | 0 | 1 | 0 |
| Office of Personnel Management | 3 | 5 | 0 |
| Railroad Retirement Board | 0 | 8 | 1 |
| Small Business Administration | 1 | 1 | 0 |
| Department of State | 2 | 2 | 0 |
| Tennessee Valley Authority | 0 | 4 | 5 |
| Department of the Treasury | 0 | 3 | 0 |
| Veterans Administration | 9 | 11 | 2 |

SOURCE President's Commission on Integrity and Efficiency,

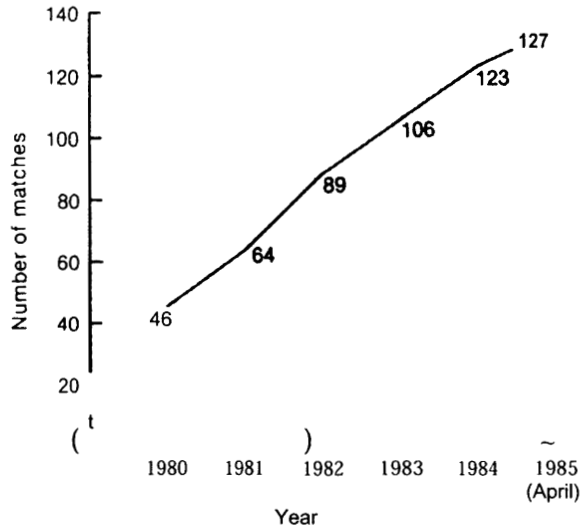
In response to the OTA survey of Federal agencies, 11 cabinet-level departments and 4 independent agencies reported conducting 110 matching programs[34] with a total of approximately 700 matches from 1980 to April 1985. The Departments of Energy and State were the only two cabinet-level departments that reported no matching programs. Of the 20 independent agencies surveyed, only three (NASA, Selective Service System, and Veterans Administration) reported any matching programs (see table 9 for a breakdown of matching programs by agency).

While the data from the responses to OTA and to PCIE are not directly comparable, the trend toward increased use of computer matches is clear (see figure 4). In the 5 years from 1980 to 1984, the number of computer matches nearly tripled.

From 1979 to 1984, OMB received only 56 reports on matching programs from Federal agencies. According to OMB records, there were 11 matches reported in 1979; 2 in 1980; 11 in 1981; 13 in 1982; 6 in 1983; and 13 in 1984. The OMB figures are obviously lower than the matching figures reported elsewhere because: 1) only those matching programs that fit the OMB definition are included; and 2) some agencies do not submit match notices under the routine use and systems of records, but instead fit matching programs into existing routine use and existing systems of records.

[34]Some of these matching programs are conducted within an agency and therefore do not fall within the OMB definition.

**Figure 4.—Computer Matches Conducted From 1980 to April 1985**




SOURCE: Off Ice of Technology Assessment

**Table 9.—Computer Matching Programs[a] Reported to OTA**

| | |
|---|---|
| Department of Agriculture | 33 |
| Department of Commerce | 1 |
| Department of Defense | 15 |
| Department of Education | 3 |
| Department of Health and Human Services | 1 |
| Department of Housing and Urban Development | 3 |
| Department of the Interior | 3 |
| Department of Justice | 6 |
| Department of Labor | 21 |
| Department of Transportation | 1 |
| Department of the Treasury | 14 |
| National Aeronautics and Space Administration | 1 |
| Selective Service | 1 |
| Veterans Administration | 7 |

[a]Some of these matching programs are conducted within an agency and therefore do not fall within the OMB definition.

SOURCE OTA Federal Agency Data Request

In determining the scale of computer matching activities at the Federal level, it is also important to consider the number of records that have been matched. In response to the OTA data request, information on number of records matched, number of hits, and percent of hits verified was provided for 20 percent of the matches reported. Despite this low response, the number of separate records used in the reported matching programs totaled over z billion; the total number of records matched was reported to be over 7 billion due to multiple matches of the same records.

## Finding 3

As yet, no firm evidence is available to determine the costs and benefits of computer matching and to document claims made by OMB, the inspectors general, and others that computer matching is cost-effective.

Before discussing the attempts to date at estimating costs and benefits, it is important to place computer matching within a context. Computer matching is a technique that has been used primarily to detect client fraud, which is only one component of fraud, waste, and abuse. In order to accurately determine the cost-effectiveness of computer matching, the extent of client fraud must first be documented. If client fraud accounts for only a small percentage of total fraud, waste, and abuse, then other techniques to detect other types of fraud, waste, and abuse maybe more cost-effective overall. In this respect, one author cited the 1978 Annual Report of the HEW Inspector General, which estimated that the Department lost between $5.5 and $6.5 billion through management inefficiencies, program misuse, and fraud. In this instance, management inefficiencies and program misuse accounted for 97 percent of the inspector general's estimate of losses, while client fraud accounted for only 3 percent.[36]

In response to the OTA survey, only 8 percent of the agencies that reported participation in computer matching activities (3 out of 37 agencies) said that they did cost-benefit analyses prior to computer matching. Eleven percent (4 of 37) reported doing cost-benefit analyses after matching.

Various individuals and organizations have asserted that computer matching is cost-effective, but have provided little or no specific information on actual costs and benefits. For example, Joseph Wright, OMB'S Deputy Director, reported in an OMB circular that:

> The IG's are wisely using this spectacularly effective technique to reap for the American public the savings that private industry has for many years been obtaining. Use of this technique will help assure that individuals who are not entitled to receive payments don't, making more money available for those who are deserving.[36b]

Likewise, the Grace Commission concluded that:

> Computer matching is an effective management tool for identifying fraud, waste, and abuse of government benefits, entitlements and loan programs. Computer matching is useful in other ways too, such as validating billings of large government contractors. . . Recommendations in the task force reports to correct information problems related to this issue provide opportunities for cost savings and revenue of $15.9 billion over 3 years ($11.3 billion when information gaps cited in other issues in the Report are netted out).[37]

In the 1982 Cohen hearings on computer matching, former Inspector General McBride of the Department of Labor testified that:

> The hits, the overpayments, for the big benefit programs run somewhere between 1.8 up to maybe 4 percent, depending on what program you are talking about. For AFDC, the hits are probably somewhere at the lower end, because they do a little better job of verification. Food stamps is a little higher. Unemployment insurance may be even higher, in some States particularly .38

In a 1983 article, Richard Kusserow, Inspector General of the Department of Health and Human Services, reported:

> Our own Project Spectre which matches Social Security beneficiary payments with Medicare death files has led to about $7.5 million in recoveries to date. Recoveries, in this case, covers all monies collected by our investigators, including checks not cashed but debited to the treasury. We project total savings over time to reach $25.2 million.[39]

In *Computer Matching in State Admim"stered Benefi"t Programs: A Manager's Guide*

---

[36] Young, op. cit., p. 362.

[36] OMB 83-14.

[36b] "President's Private Sector Survey on Cost Control, *A Report to the President* (1984), Part II: Issue and Recommendation Summaries, p. 82; see pp. 84-86 for examples.

[38] Cohen hearings, op. cit., p. 19.

[39] Richard P. Kusserow, "Fighting Fraud, Waste and Abuse," *The Bureaucrat*, fall 1983, p. 23.

*to Decision Making,'"* the quantitative benefits of computer matching include estimated savings and measures of grant reductions, collections, and corrections. The list of qualitative benefits is longer, including: increased deterrence, improved eligibility determinations, enhanced public credibility for benefit programs, more effective referral services, and improved databases.

The costs of computer matching vary according to the size of the record set, as well as the complexity, quality, and compatibility of the records. In *Computer Matching in State Administered Benefit Programs,* the quantitative costs include: hardware/software; computer processing time; space; supplies; personnel managers, data-processing staff, eligibility assistance workers, clerical workers, hearings officers, fraud investigators, collections staff, attorneys, and training staff; other public agency resources; and private institution resources. The qualitative costs include: reduced staff morale, heightened public concerns about "big brother, " increased political conflict, gamesmanship with numbers, operational inefficiencies, and diversion of resources. Definitions for these qualitative costs are not offered.

All agree that verification costs are the highest and the most difficult to compute. In *Computer Matching in State Administered Benefit Programs,* it is pointed out that:

> Follow-up is the most costly, labor-intensive part of the computer matching process. Most notably, it involves what can be a very tedious and time-consuming job of verifying hits. But it also involves other components such as making any necessary change in a recipient case status, calculating and pursuing overpayments, hearing appeals, making referrals to fraud units, and actually conducting criminal investigations and pursuing convictions.[41]

There is some disagreement as to how much verification, both in terms of number of hits verified and in terms of records and sources checked, is necessary. For example, the Department of Health and Human Services' Inspector General Kusserow has suggested that:

> For large matches, officials would have to analyze only a sample of the hits to verify the matching process. After doing this, officials should take corrective measures, proceeding cautiously against any individual where doubt exists.[42]

The PCIE Long Term Computer Matching Committee has developed some information on the costs of selected matches. For many of the matches, the information presented is very sketchy. The matches for which the PCIE offered the most complete information are listed in table 10.

David H. Greenberg and Douglas A. Wolf have recently completed a study[43] in which they constructed a cost-benefit framework (see table 11) and used it to evaluate the performance of computer wage-matching systems of welfare agencies in four areas: Camden County, New Jersey; Mercer County, New Jersey; San Joaquin County, California; and the State of New Hampshire. In each of their study sites, they reported that they obtained reliable and complete information on the costs of matching, but were unable to measure benefits as precisely. Additionally, there were some benefits, e.g., deterrent effects and positive effects on attitudes of affected parties, that they could not measure at all. Thus, they regard their test of the cost-effectiveness of wage matching to be a conservative one.

Greenberg and Wolf concluded from their four case studies that the benefits from computer matching outweighed the costs by "substantial amounts' '[44] (see table 12). If computer matching were as effective nationally, they suggested that "cost savings in the food stamp and AFDC programs would be approximately

[40]U.S. D-epartment of Health and Human Services, Office of Inspector General, *Computer Matching in State Administered Benefit Programs,* June 1984, p. 25.

[41] Ibid.

[42]Richard P. Kusserow, "The Government Needs Computer Matching To Root Out Waste and Fraud, " *Communications of the ACM,* vol. 27, No. 6, June 1984, p. 544.

[43]David H. Greenberg and Douglas A. Wolf, "Is Wage Matching Worth All the Trouble?' *Public Welfare,* winter 1985, pp. 13-20.

[44]I bid., p. 18.

**Table 10.—Examples of Cost/Benefit Analyses**

| Costs/benefits | Selected matches | | | | | |
|---|---|---|---|---|---|---|
| | DO L/TVA | IRS/DOL | OPM/SSA | OPM/OPM | RRB/HCFA | USAFIVA |
| Equipment costs | 1,500 | 125,000 | 10,950 | 2,291 | 6,124 | 1,000 |
| ADP staff costs | 1,200 | 25,000 | 3,213 | 2,142 | 1,831 | 1,150 |
| Staff verification costs | 4,500 | 1,000,000 | 94,163 | 12,968 | 15,763 | 96 |
| Travel and other costs | 10,000 | — | 39,416 | — | 10,028 | 100 |
| Cases found | 21 | 219 | 770 | 170 | 405 | 340 |
| Overpayments identified | 35,000 | 103,000 | 9,100,000 | 640,800 | 2,263,927 | 71,000 |
| Cases with recoveries made | 2 | 219 | — | — | 364 | — |
| Overpayments recovered | 2,500 | 139,000 | — | — | 993,118 | — |
| Overpayments prevented | — | — | 770 | 170 | — | 1,300 |
| Amount prevented | — | 50,000 | 4,089,600 | 46,300 | — | 274,000 |
| Questioned costs | — | — | — | — | — | — |
| Disallowed costs | — | — | — | — | — | — |

KEY: DOL = Department of Labor, TVA = Tennessee Valley Authority, IRS = Internal Revenue Service; OPM = Office of Personnel Management; SSA = Social Security Administration; RRB = Railroad Retirement Board; HCFA = Health Care Financing Administration, USAF = U S Air Force, VA = Veterans Administration.

SOURCE President's Council on Integrity and Efficiency Long Term Matchlng Committee, '(Draft/Summary of Federal Computer Applications for Prevention of Fraud and Abuse "

**Table 11 .—Costs and Benefits of Wage Matching**

*Benefits:*
Restitution of previous overpayments
Savings from food stamp disqualifications
Savings from benefit reductions and discontinuances:
- prevention of future overpayments
- administrative savings

Changes in behavior and attitudes:
- deterrent effects
- improved client attitudes
- improved staff morale
- improved relations with the public

*costs:*
Personnel costs (salaries and fringe benefits):
- income maintenance staff
  fraud investigative staff
- district attorney staff
- other

Materials and facilities costs:
- computers
- word processors
- forms
- general overhead such as office space, telephone, supplies

SOURCE: David H. Greenberg and Douglas A Wolf, "IS Wage Matching Worth All the Trouble?"Public *Welfare*, winter 1985, p 16

**Table 12.—Estimated Costs and Benefits of Computer Matching in Four Sites**

| | costs | Benefits | Ratio |
|---|---|---|---|
| Mercer County | $786,821 | $ 932,958 | 1.19 |
| Camden County | 753,662 | 1,452,367 | 1.93 |
| San Joaquin County | 308,128 | 762,355 | 2.47 |
| New Hampshire (DES Wage Crosshatch Project) | 264,856 | 707,316 | 2.67 |

NOTE" All figures are in annual terms pertaining mainly to 1982

SOURCE David H. Greenberg and Douglas A. Wolf, "IS Wage Matching Worth All the Trouble?" *PublK Welfare*, winter 1985, p t8

1 or 2 percent. "[45] However, they caution that this may not be the case because they chose wage-matching programs that were functioning well:

> For example, the employer-reported data used by these systems clearly were adequate in terms of coverage, content, and timeliness. Equally important: follow-up procedures were well-structured, adequate resources were available for follow-up, and supervisors were genuinely committed to the program. Without such conditions, it certainly is possible that wage matching could prove ineffective.[46]

## Finding 4

The effectiveness of computer matches that are used to detect fraud, waste, and abuse can be compromised by inaccurate data.

The Massachusetts case discussed earlier, in which 110 of the 160 termination notices that were sent following a computer match were based on erroneous information, is the best known example of use of inaccurate data. However, many matches experience some problems with inaccurate data, and, in part, computer matching can be effective in detecting errors in data.

[45] Ibid.
[46] Ibid.

One indicator, although not complete, of the quality of data used in computer matching is the percentage of hits verified as accurate. In response to the OTA survey, this percentage ranged from 0.1 to 100 percent. For example:

- The Department of Housing and Urban Development conducted computer matches to identify tenants in five different cities who had not reported all income when applying for federally assisted housing. The hit rates varied from about 6 to 54 percent, and the hit verification rates varied from 13 to 55 percent. The actual number of matches that resulted in valid hits ranged from 0.8 to 29 percent.
- The Department of Commerce Inspector General's office conducted a match to identify departmental employees who were collecting unemployment benefits. A total of 22,000 records were matched resulting in 98 hits, of which about 10 percent were verified.
- The Department of Education conducted a match to identify current and former Federal employees who were delinquent on student loans. About 10 million records were matched resulting in 46,860 hits, of which 100 percent were verified, according to Department officials.
- The Veterans Administration conducted a match to identify Federal employees and annuitants who were erroneously receiving VA compensation. About 15 million records were matched resulting in 5,166 hits, of which about 23 percent were verified.

For the majority of matches reported to OTA, information on hits verified was either unknown or unavailable.

Proponents of matching programs are taking measures to improve the quality of data used in matches. SSA has developed a computer software program to screen social security numbers and pull out inaccurate or incongruous numbers. Other agencies engaging in matching programs are likewise concerned. In response to the OTA survey, 68 percent (25 of 37) of the agencies indicating that they participated in matching programs said that procedures were used to ensure that the subject record files contain accurate information.

## Finding 5

There are numerous procedural guidelines for computer matching, but little or no oversight, follow-up, or explicit consideration of privacy implications.

Program personnel appear to have substantial discretion in deciding whether or not to use computer matching as an audit technique or means to detect fraud, waste, and abuse. There are few internal agency checks. The Inspector General's Office may be involved in planning a computer match; and the General Counsel's Office and the Privacy Act officer may be involved. But it appears that there are no agency or general policy guidelines regarding what types of information should be matched, against which records of what other agencies, and for what purposes. These substantive issues are rarely addressed.

For those matching programs that meet the OMB definition, agencies providing information "are responsible for determining whether or not to disclose personal records from their systems and for making sure they meet the necessary Privacy Act disclosure when they do. " In making this determination, agencies are instructed to consider the following:

- legal authority for the match;
- purpose and description of the match;
- description of the records to be matched;
- whether the record subjects have consented to the match; whether disclosure of records for the match would be compatible with the purpose for which the records where originally collected, i.e., whether disclosure under a' 'routine use' would be appropriate; whether the soliciting agency is seeking the records for a legitimate law enforcement activity; or any other provision of the Privacy Act under which disclosure may be made;
- description of additional information that may be subsequently disclosed in relation to "hits";

- subsequent actions expected of the agency providing information (e.g., verification of the identity of the "hits" or follow-up with individuals who are "hits"); and
- safeguards to be afforded the records involved, including disposition.

However, neither the source agency, the matching agency, nor OMB is accountable for the decision whether or not to disclose records for a matching program. For matching programs that do not fall under the OMB guidelines, there are no formal procedures or guidelines-one program manager may ask another for access to records for matching purposes, and no one else need know.

OMB has developed a number of procedural guidelines. The initial guidelines, *OMB Guidance to Agencies on Conducting Automated Matching Programs,* became effective on March 30, 1979. The purpose of the guidelines was "to aid agencies in balancing the government need to maintain the integrity of Federal programs with the individual's right to personal privacy. " Under the guidelines, a match was to be performed "only if a demonstrable financial benefit can be realized that significantly outweighs the costs of the match and any potential harm to individuals that could be caused by the matching program. " To this end, the guidelines required documentation of benefits, costs, potential harm, and alternatives considered to detect or curtail fraud and abuse or to collect debts owed to the Federal Government (see 5a of guidelines for listing). A report describing the match (see 9b.1 and 2 of guidelines for details) was to be submitted, 60 days before the match was initiated, to the Director of OMB, the Speaker of the House, and the President of the Senate. Necessary notices of system of records, new or altered systems, or routine use were to republished in the *Federal Register,* allowing 30 days for public comment. Any disclosures of personal information during the match were to be made in accordance with the "routine use" limitations noted in the *Federal Register.* Unless it was a continuing matching program, the guidelines stipulated that personal records should be destroyed or returned to the source agency within 6 months. The guidelines also suggested that matching should be done in-house by agency personnel, not by contractors.

The application of these guidelines was not very satisfactory for any party concerned. Agencies did not conduct cost-benefit analyses in a systematic fashion; instead, they were quickly estimated when asked for by OMB in order to comply with the letter of the guidelines. There was almost no public comment in response to matches proposed in the *Federal Register.* There was little congressional reaction to matching programs. There was minimal to no oversight by OMB; it processed the necessary paperwork, but never 'disapproved' a match. In part, OMB'S behavior can be attributed to the lack of clarity in the guidelines concerning its role. For example, it was not clear from the guidelines whether OMB had the authority to disapprove a match.

Based on the unsatisfactory experience under the 1979 guidelines, the PCIE'S Long Term Computer Matching Project decided that one of its first projects would be to revise the OMB guidelines. In conjunction with advice from PCIE, OMB'S *Revised Supplementary Guidance for Conducting Matching Programs* became effective May 1,1982. The 1982 guidelines simplified the administrative reporting requirements of the 1979 guidelines by eliminating the cost-benefit analysis, reducing the notice and reporting requirements, and exempting intra-agency matching programs. Publication of "routine uses" in the *Federal Register* was still required, but the 30-day public comment period for matching reports and advance notice to Congress and OMB were eliminated.

OMB and PCIE also developed a Model *Control System for Conducting Computer Matching Projects Involving Individual Privacy Data (1983).* The Model Control System is designed to provide procedural guidance to agencies conducting computer matching projects to help them comply with the Privacy Act and the OMB guidelines. The model includes 10 steps that agencies should follow:

1. define the match program,
2. determine the feasibility of the match,
3. establish matching and follow-up procedures,
4. confer with the agencies providing information,
5. publish routine use notice,
6. make a matching report,
7. obtain the agency data file,
8. conduct computer matching,
9. analyze and refine the raw hits, and
10. perform follow-up procedures.

Agencies are not required to follow the Model Control System, or to report to OMB on which procedures were followed.

In late 1983, OMB developed a *Computer Match Checklist* that must be on file for review by OMB, GAO, or other Federal entities. The checklist must be completed by both the agency providing information and the agency conducting the match immediately following *Federal Register* publication of an intent to match. Items on the checklist include: compliance with notification requirements, number of individuals whose records are to be matched, contractor involvement, and the date on which a cost/benefit analysis on the match will be available. Estimates of cost/benefit analyses are to be attached to the checklist.

In December 1985, OMB issued Circular A-130, *Management of Federal Information Resources,* which directs agencies to review annually every matching program in which they have participated, either as a matching or source agency, to ensure that the requirements of the Privacy Act, the OMB Matching Guidelines, and the OMB Model Control System and Checklist have been met. Additionally, agencies are to include in the Privacy Act Annual Report the number and description of matching programs participated in as a source or matching agency.

## Finding 6

As presently conducted, computer matching programs may raise several constitutional questions, e.g., whether they violate protection against unreasonable search and seizure, due process, and equal protection of the laws. But, as presently interpreted by the courts, the constitutional provisions provide few, if any, protections for individuals who are the subjects of matching programs.

The fourth amendment provides individuals the right "to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures. " The fourth amendment presumption, reinforced by case law and by the presumption of innocence additionally reflected in the fifth and sixth amendments, is that searches are not warranted unless there is indication of a crime. If there is probable cause of a crime and the individual's involvement, then a court may issue a search warrant. Fourth amendment case law has resulted in the concept of "expectation of privacy. "

The question of whether or not computer matches raise fourth amendment issues turns, in large part, on the 'expectation of privacy" that individuals have in records about them maintained by a third party, in this case primarily a government agency. Based on the Supreme Court ruling in United *States v. Miller, 425* U.S. 435 (1976), records that are held by a third party, and used by that party for administrative purposes, are considered the property of the third party. Under such circumstances, the individual does not have an assertible fourth amendment privacy interest in those records. Although *Miller* applied to records held by a bank, the logic of the holding may apply similarly to records held by the government.

In *Jaffess v. Secretary HE W, 393* F. Supp. 626 (S.D. N.Y. 1975), a district court allowed a computer match of recipients of veterans' disability benefits with those receiving social security benefits. The court held that the disclosure under the matching program was 'for the purpose of proper administration. Jaffess had not reported his social security income, and after the match his {eterans' benefits were reduced. He claimed that a constitutional right of privacy protected his records. The court rejected this claim:

> ... the present thrust of decisional law does not include within its compass the right of an individual to prevent disclosure by one governmental agency to another of matters obtained in the course of transmitting agency's regular functions.[47]

But, the legal question of what kind of fourth amendment "expectation of privacy" an individual has when he or she fills out a form and swears that the information provided is true and correct has not been specifically decided. Nor has the question of the privacy rights of Federal workers in information provided and maintained for employment purposes. In both instances, statutes, especially the Privacy Act, may give more precise legal guidance than the U.S. Constitution. However, the constitutional question could still be subject to further litigation.

A second fourth amendment issue that is raised by computer matches is the scope of the search. Computer matches are general electronic searches of, frequently, millions of records. Under the fourth amendment, searches are not to be overly inclusive—no 'fishing expeditions" or "dragnet investigations. " Yet, in matches, many people who have not engaged in fraud are subject to the computer search. If matches were to be considered a fourth amendment search, then some limitations on the breadth of the match and/or justifications for a match may be necessary. For example, the agency may need to show that a less intrusive means to carry out the search was not available, and that procedural safeguards limiting the dangers of abuse and agency discretion were applied. These may also be required under due process protections as discussed below.

A final fourth amendment issue that may be raised by computer matches is that of suspicion that criminal activity is occurring. If the purpose of a match is to produce evidence that someone has defrauded the government, then a computer match could be regarded as a search under the fourth amendment. Such a match may also conflict with the presumption of innocence, as reflected in the fourth and fifth amendments, if the individual is required to prove that he or she has not engaged in wrongdoing. If the purpose of a match is to detect and correct errors, and not to detect wrongdoing, then a match would probably not be regarded as a search under the fourth amendment.

The *due process* clause of the fifth[48] (Federal Government) and 14th (State governments) amendments ensures procedural protections before the government takes action against an individual. Generally, this clause has been held to require that individuals be given notice of their situation, the opportunity to be heard, and the opportunity to present evidence on their own behalves. In agency proceedings, this constitutional principle is given specific meaning in the Administrative Procedures Act (1946). Additional elements of due process that apply specifically to eligibility for benefit programs include: the right to a pre-termination hearing, placing the burden of proof *on* the gOVernment to prove ineligibility if the individual swears to eligibility, and entitlement to benefits pending resolution. These procedural due process protections were extended to welfare recipients in *Goldberg v. Kelly, 397* U.S. 254 (1970).

Under the 1979 OMB guidelines, notice of a proposed match is to be published in the Fed*eral Register 30* days before to allow time for comments. Many have questioned the adequacy of this, as the vast majority of individuals do not read the *Federal Register.* Additionally, there is evidence that agencies have not complied with the 30-day time period and that some agencies have provided notice *after* the match was well under way.[49] This requirement was eliminated in the 1982 OMB guidelines. DEFRA now requires more specific no-

[47] Kenneth James Langan, "Computer Matching Programs: A Threat to Privacy?" *Columbia Journal of Law and Social Problems,* vol. 15, No. 2, 1979, pp. 158-159.

[48] It does not specifically provide for equal protection, but the Court ruled in *Bolling* v. *Sharpe* (347 U.S. 497, 19854) that "the concepts of equal protection and due process, both stemming from our American ideal of fairness, are not mutually exclusive" and that the fifth amendment also provided equal protection.

[49] See Cohen hearings, op. cit.

tice prior to some matches. It is important to recognize that notice can take place at various points in the matching process, i.e., before the match occurs, once an individual appears as a "hit," and prior to any outside verification. Notice can also be provided rather passively, e.g., a statement on a form, or requiring the active acknowledgment of the individual. Based on results of the OTA survey, 8 percent (3 out of 37 agency components) of the agencies reporting that they participated in computer matching said that individual subjects of the match had provided written consent prior to a match.

Once a match has taken place, the resulting "hits" are further investigated in order to verify their status. At this time, these individuals may not be given notice of their situation, or the opportunity to be heard and present evidence on their own behalves. They may not be notified until and unless the agency decides to take some action against them. Based on the Court's ruling in *Goldberg,* due process would require a hearing for an individual whose benefits are to be terminated or lowered based on information from computer matching. Such hearings may be quasi-judicial in nature, but the individual would not have the right to a lawyer or jury, the burden of proof would be on the individual, and the individual may incriminate himself or herself in these hearings. If such hearings are the starting point for an investigation leading to criminal charges, then it maybe necessary to conduct them in a more formal judicial setting.

The *equal protection* clause of the 14th and, by implication, the fifth amendments prohibits the States and Federal Government from creating legal categories and taking actions that discriminate against members of that category (e.g., race, national origin, and gender). Economic status has never been regarded as a *suspect* classification,'" and therefore the government interest in subjecting welfare recipients to computer matching would only need to be rationally related to a legitimate purpose of the government. In this case, the purpose, i.e., detecting fraud, waste, and abuse, would probably be regarded as legitimate, and the means chosen, i.e., computer matching, rationally related.

Despite this development of constitutional decisions, matching may conflict with the equal protection clause in that categories of people, not individual suspects, are subject to these electronic searches. In the computer matching that has been done to date, two groups of people—welfare recipients and Federal employees—have been used frequently. This is true despite arguments by supporters of matching that computer matches are effective tools in a number of situations. Although the Grace Commission and others have recognized the usefulness of matching in detecting fraud, waste, and abuse in government contracting, it has not been used to any significant extent for this purpose. DEFRA, in its section incorporating the Grace Commission recommendations, did not require or endorse the use of matching in government contracting.

## Finding 7

The Privacy Act as presently interpreted by the courts and OMB guidelines offers little protection to individuals who are the subjects of computer matching.

The Privacy Act gives individuals certain rights of notice, access, and correction in order that they may control information about themselves. It also places certain requirements on agencies to make certain that the information they maintain is relevant, timely, and complete.

Under the Privacy Act, the individual has the right to prevent information being used without his or her consent for a purpose other than that for which it was collected. An exception to this rule is if information falls within a "routine use" of the particular record system. Under the OMB Matching Guidelines, matching can be considered such a routine use; therefore, individual consent is not required. Many argue that matching of information is

"*see* **Dandridge** *v.* **Williams,** *397* U.S. 471 (1970) and San *Antonio Independent* **School** *District v. Rodriguez, 411* U.S. 1 (1973).

not consistent with the legislative intent that information should be used only for the purpose collected. As table 6 indicated, it is quite easy to find justification in the Privacy Act for disclosures of information for matching purposes.

Additionally, the Privacy Act requires agencies to 'collect information to the greatest extent practicable directly from the subject individual when the information may result in adverse determinations about an individual's rights, benefits, and privileges under Federal programs" [see.e(2)]. In computer matching, information that will be used to determine whether benefits should be eliminated, decreased, or increased is collected from third parties-not from the individual.

Although not specifically prohibited in the Privacy Act, the legislative history reflects censure of a national data center. The linking of systems in computer matching can be regarded as moving towards a de facto national data center or national recipient system. Additionally, new computerized databases are being created solely for the purpose of providing information for computer matches and other record searches. The Federal Government, under the auspices of the inspectors general, is developing a national computerized file of deceased individuals (who have no rights under the Privacy Act) for screening beneficiary records and preventing payments to deceased persons. Two other examples mentioned previously are the Medicaid Management Information System and the proposed IRS Debtor Master file. The State wage reporting systems, required under the proposed DEFRA regulations, could also be regarded as the first stage of a national data system.

The OMB guidelines require that the files used for matching be returned to the custodian agency or destroyed. However, since there is no oversight of this, records could be used for additional purposes.

## Finding 8

The courts have been used infrequently as a forum for resolving individual grievances over computer matching, although some organizations have brought lawsuits.

It does not appear likely that the courts will protect individual privacy in computer matching programs .51 There are at least four reasons. The first is that the courts have not extended constitutional protections for computerized records, and the fourth amendment "search and seizure" doctrine has not been applied. The second reason is that courts only require rationality in such programs, i.e., that the means used be reasonably related to a legitimate government purpose. The purpose of achieving efficiency and detecting fraud, waste, and abuse is a legitimate one. With respect to the choice of means, courts have traditionally given deference to administrative discretion. The third reason is that when courts balance individual privacy against the public interest, the weight generally favors the public interest—all else being equal. The fourth reason is that the damage requirements of the Privacy Act are so difficult to prove that they act as a deterrent to its use.

Additionally, with large-scale computer matching, no one individual is sufficiently harmed to litigate a claim and most individuals are not even aware of the match. The cases that have gone to court have generally been brought by welfare rights organizations. These cases include:[52]

> *15, 844 Welfare Recipients v. King, 474* F. Supp. 1374 (D. Mass., 1979)–State welfare agency was required to restore benefits to recipients whose aid had been terminated either by fraud investigators improperly acting as caseworkers, or by caseworkers improperly acting as fraud investigators.
>
> *Tierney v. Schweiker, 718* F. 2d *449* (D.C. Cir., 1983)–Coerced signatures to notice-and-consent forms, extracted from SS1 recipients in preparation for an IRS matching, were invalidated because the agency action violated IRS confidentiality rules.

---

[51] Langan, op. cit., p. 175.

[52] *See: Henry Korman, "Creating the Suspicious Class—Surveillance of the Poor by Computer Matching," unpublished paper, August 1985, esp. pp. 52-53.

*Greater Cleveland Welfare Rights Organization v. Bauer, 462* F. Supp. 1313 (N. D. Ohio, 1978)–An Ohio wage match was invalidated insofar as subject AFDC recipients were not informed of use of their social security numbers as identifiers in the match.

*Lessard v. Atkins,* CA 82-3389-MA (D, Mass., Apr. 23, 1985)–Defendants in a bank match case agreed to both the use of secondary identifiers and enhanced follow-up investigations that plaintiffs argued were required by Federal law.

## Finding 9

Computer matches are conducted in most States that have the computer capability. At least four-fifths of the States are known to conduct computer matches, most in response to Federal directives.

In many respects, the personal information gathered by State agencies is more sensitive and more extensive than that gathered by Federal agencies. [51] Many Federal agencies fund programs that are administered through the States (or local educational agencies). The Federal agencies do not store individually identifiable information on all of the beneficiaries of these programs, but the States do. Federal auditors regularly have access to individually identifiable information to monitor program effectiveness, but the personal data on all participants is not stored in Federal agencies themselves.

At the State level, the following information is typically stored: income or business taxpayer records in the revenue department; driving records in the Department of Motor Vehicles; public assistance in the welfare agency; drug and alcohol treatment records in the appropriate agencies; communicable diseases and abortions in the Department of Health; treatment at State institutions in the Departments of Health, Mental Health, or Public Health; current earnings in the quarterly reports submitted by employers (a few States require reporting less often) to the unemployment security office; criminal records and criminal intelligence in the State police or Department of Public Safety; educational, financial aid, and vocational training information in the Department of Education; occupational information in the various State licensing boards (attorneys, beauticians, auctioneers, boxers, vendors, physicians, etc.); patient information and physicians earnings records in the State agency administering Medicaid; suspicions of child abuse in the appropriate State agency; and birth records of adoptees in the adoption agency.

Most matching occurs in programs that are federally funded or controlled by Federal law. For example, States conduct matches in unemployment insurance programs to detect fraudulent and duplicative payments, and to monitor employers' contributions. Forty-one States reported conducting such matches, and 23 States reported matching unemployment insurance records with other jurisdictions. [54] Less than 20 States report matching for workers' compensation programs. [55] In public assistance programs, States generally match recipient files against quarterly wage reports submitted by employers to detect recipients who are receiving wages over an allowable limit. An OTA survey of eight States revealed that six (California, Colorado, Georgia, Illinois, Indiana, and Michigan) conducted such matches, while two States (Florida and Minnesota) did not. DEFRA now requires that this be done by all States.

Other examples of State matching activities include:

- Thirty-seven States submit social security numbers of welfare recipients to SSA for computerized verification that the numbers are accurate.
- At least two States, Massachusetts and Maryland, have authorizations in their laws for the public assistance program to conduct computer matches against the accounts of all bank customers in the State.

[51] Information for this section is derived from Robert Ellis Smith, *Report on Data Protection and Privacy in Seven* Selected States, OTA contractor report, February 1985.

[54] See U.S. Department of Labor Inspector General, *Inventory of Computer Matching Activities in State Labor and Related Agencies, 1982.*

[55] Ibid.

- The Iremigration and Naturalization Service is encouraging States to match motor vehicle, welfare, and unemployment files with its databank of current registered aliens. Colorado, Illinois, and California have agreed. California must approve new regulations before this can be done, and the regulations have not yet been published.
- California, Minnesota, and several other States conduct Project Intercept. Lists of persons owing money to the State—either in delinquent taxes, welfare overpayments or frauds, faulty unemployment compensation, etc.—or those reported delinquent in child support payments are submitted to the public assistance agency (or any other agency making periodic payments) so that the amount owed is offset against the State payments. This is also done with tax refund checks (not only in the States, but by the IRS as well).
- Many States compare their lists of recipients, whether public assistance, unemployment compensation, or other payment programs, against comparable lists of recipients in neighboring jurisdictions, to determine who is "double-dipping." Examples are Virginia's unemployment compensation records matched with those of Maryland and the District of Columbia; or Indiana's records matched with those of Kentucky.

There are other generic exchanges of personal data by most States that are significant, although they may not be classified strictly as "matches." Many of them predate the current Federal initiative on matching, which began in 1978. They include:

- Motor vehicle departments in 49 States provide lists of young, male drivers to the Selective Service System for matching against lists of men who have registered for a military draft. Objections, based on invasion of privacy, were expressed in many States. Some laws or regulations governing DMVS seem to prohibit such disclosures. But in the end, the Selective Service System had nearly 100 percent participation.
- More than 80 percent of the motor vehicle departments disclose driving records and accident reports to Dataflo Systems, a division of Equifax, Inc., so that Dataflo can computerize the data and market it to insurance companies. The abstract includes social security number, driver's license number, birth date, physical description, restrictions on the permit, and a chronological list of violations. An insurance company can then query one of five regional computers operated by Dataflo.
- Motor vehicle departments also disclose suspended or revoked licenses to the National Driver Register operated by the U.S. Department of Transportation in Washington and, in turn, query the system when persons apply for drivers' licenses.

Just about all motor vehicle departments rent mailing lists of licensees and of automobile owners to mailing list firms and other marketers. A report by the Secretary of State of Illinois in 1983 stated that 44 States answered in the affirmative when surveyed on whether they rent mailing lists. The other six States did not respond. Many States, however, have regulations or laws limiting, if not fully prohibiting, such disclosures.

- Every State with a State income tax has an agreement with the IRS to exchange computerized data on its taxpayers with IRS and to receive comparable information from IRS.

An analysis of State matching activities in light of State Privacy Acts or Fair Information Practices Acts indicates that the presence of such laws does not deter computer matching. However, it often assures that there is a review of a State agency's decision to match, that there are specific procedures to follow, and that information is checked for accuracy. The critical factor in determining the extent of matching at the State level appears to be the size of the population. States with larger populations engage in more computer matching than States with smaller populations.

## Finding 10

All Western European countries and Canada are using computer matching or record linkages, to an increasing degree, as a technique for detecting fraud, waste, and abuse.

In general, the specific uses of matching in Western Europe and Canada are similar to those in the United States—primarily in social welfare programs.[56] In Western European countries, computer matching and other record linkage issues are handled within the context of data protection laws and oversight. In general, European data protection laws require the advice or consent of the data protection agency before any records can be Linked. A brief review of matching activities in different countries follows.

### Canada

The Canadian Privacy Act of 1982 does not address computer matching specifically, but does contain the principle that information should be used only for the purpose for which it was collected. The Canadian Privacy Commissioner, John W. Grace, has spoken out strongly on the privacy implications of matching. As he sees it:

> That computer-matching is carried on in the name of efficiency, good government and law enforcement makes it potentially a more, not less, dangerous instrument in the State's hands."

Specific instances of matching include: opening Federal databanks to obtain information for collecting alimony and child support payments from recalcitrant fathers, Revenue Canada's matching of a provincial voters' list with tax records to identify individuals who had not filed tax returns, and matches by the Canadian Employment and Immigration Commission to detect overpayment of unemployment insurance benefits.

### Sweden

Under Section 2 of the Data Act, specific permission is required from the Data Inspection Board (DIB) for the linkage of files that contain "personal data procured from any other personal file, unless the data are recorded or disseminated by virtue of a statute, a decision of the Data Inspection Board, or by permission of the person registered. " DIB evaluates all proposals for record linkages and has approved an estimated 80 to 90 percent of the proposed record linkages. In reviewing proposals, DIB looks especially at the purpose of the match and the quality, e.g., timeliness, accuracy, and completeness, of the data to be used. In general, DIB is opposed to linkages of very sensitive personal information, e.g., alcoholism and drug addiction records, and linkages where the users do not know why personal information was originally collected.

DIB has not always been successful at preventing record linkages. For example, when the tax authorities sought information on income from interest and dividends from the banks, DIB said that the banks were not licensed to divulge such information to the tax authorities. Regardless, the banks gave the information to the tax authorities. DIB sought to prosecute the banks under the Data Act and the case is still under appeal.

### France

The National Commission on Informatics and Freedoms (CNIL) has to authorize record linkages. In general, CNIL is opposed to linkages because of the principle that data should be used only for the purposes for which they were collected. In contrast to other countries, there are few plans for record linkages.

### Federal Republic of Germany

The Republic's Federal Data Protection Act contains a general prohibition against the dissemination of personal data from one public body to another, unless the release of the information "is necessary for the legitimate accomplishment of the tasks for which the dissemination unit or the recipient is competent. "

---

[56] Information for this section is derived from David H. Flaherty, "Data Protection and Privacy: Comparative Policies, " OTA contractor report, January 1985.

[57] Privacy Commissioner, *Annual Report, 1983-84*, p. 3.

Computer linkages among social services occur frequently and do not have to be reported to the Data Protection Commissioners. Most linkages of social service data outside the social service administrations are prohibited by the Social Code unless the information is necessary to prevent premeditated crimes, to protect public health under certain circumstances, to implement specific stages of the taxation process, and to assist the registered alien authorities.

## Finding 11

Computer matching raises a number of policy questions that warrant congressional attention, including availability of records for matching, approval before matches, notice for individuals, requirement of cost-benefit analysis, and verification of hits.

In designing policy for computer matching, consideration of the following factors is important:

*Records to be made available for computer matches and for what purposes.* —Currently, there are few restrictions on the systems of records that can be used. If a "routine use" can be crafted to justify the match, then almost any Federal system can be made available. The primary exception to this is IRS information, but this restriction can be circumvented somewhat by matching with a system of records that has already been matched against IRS information. Another long-standing exception has been private sector information; however, a number of new Federal and State laws now allow for such access.

In determining what records should be available, several possibilities exist. One is to make all records available for all matches. Another is to prohibit the use of some systems of records, e.g., health information, bank records, or IRS records. A third is to make the availability of records dependent on the purpose of the match. The difficulty with this alternative, which may be otherwise attractive because it allows flexibility, is that it could easily evolve into a system similar to what currently exists where routine use exceptions are not carefully scrutinized. If the use of records is to depend on the purpose of the match, then the purposes that would legitimate the use of particular systems of records need to be specifically established in advance of proposals to match.

Another issue in determining what records are to be available is the quality of records used in computer matching. Inaccurate records detract from the effectiveness of computer matching and increase the problems individuals experience as a result of a match. Record systems could be required to meet specific data quality standards prior to being used in a computer match.

*Approval required before a match takes place.* —Both a process for approving matches and a substantive review of the purpose of the match must be considered. In terms of process, one task is to check on and oversee program managers' decisions to match. This check could be carried out within an agency, as often appears to be the case at present, by a formal executive branch review process, or by review by a legislative body. In addition to the process, criteria need to be developed to determine the appropriateness of matching under the circumstances. Such criteria could be based on both the privacy interests involved and the management interests.

*Notice to individuals.* –This depends in part on the purposes of notification. Originally, notice as part of due process was viewed as a means of empowering the individual. If an individual knew what was to take place, he or she could take measures to try to stop the action. This original goal seems to have been replaced with a more passive view of notice. In part this may be attributed to the lack of options available to an individual who is dependent on government benefits or employment. If this is indeed the case, i.e., that individuals could be told of an action with no recourse, its implications need to be acknowledged.

There are limitations to the present system of placing notices in the *Federal Register.* Other alternatives include placing a notice on the original application form, having an indi-

vidual sign a consent form at the time of application, writing all individuals prior to the match, and writing to obtain signed consent prior to the match.

An additional question is when to notify individuals–before they become part of the program, before the match, after matching has produced a hit, or after the hit has been verified?

*Requiring cost-benefit analysis.* –Originally, cost-benefit analyses were required prior to a match. Currently, cost-benefit analyses are to be filed with OMB following a match. Agencies have not welcomed the requirement of doing cost-benefit analyses. In part, this is because there are many qualitative costs that are difficult to measure. In part, it is because many of the quantitative costs are difficult to separate from other administrative costs. In determining what kind of a cost-benefit analysis to require, questions of time of submission, review, and components to be addressed need to be answered.

*Verification of* hits. -Other than for matches conducted under DEFRA, there are no requirements on verifying hits. Again, this involves two issues—the process of verification and the substance of what is to be verified. Specific questions include: do all hits have to be verified or only some predetermined percentage; what sources are to be used in verifying hits; if there is a discrepancy in information received, how is it resolved; and what is the role of the individual in the verification process?

*Appropriate action to be taken against an individual who has submitted false in formation.*—Presently, the individual is given an administrative hearing and can then be subject to criminal charges. If the purpose of the hearing is indeed to refine evidence for criminal proceedings, then it may be more appropriate to conduct the hearing in a formal judicial setting. Alternatively, the use of evidence from a computer match could be prohibited from criminal proceedings, allowing its use only in civil proceedings.

Chapter 4

# Computer-Assisted Front-End Verification

# Contents


| | Page |
|---|---|
| summary | 67 |
| Introduction and Background | 67 |
| Findings | 68 |
| Finding1 | 68 |
| Finding2 | 74 |
| Finding3 | 78 |
| Finding4 | 80 |
| Finding5 | 81 |


Tables


| *Table No.* | Page |
|---|---|
| 13. Computerized Databases Used for Front-End Verification | 73 |
| 14. Examples of State Front-End Verification Programs | 75 |


Figures


| FigureNo. | *Page* |
|---|---|
| 5.Current Database Linkages | 69 |
| 6. Composite of Data Linkages Through Computer Matches by AFDC Programs in Various States | 70 |
| 7. A Representative Income and Eligibility Verification System (IEVS) for a State Food Stamp Agency as Required by the Deficit Reduction Act of 1984 | 76 |

Chapter 4

# Computer-Assisted Front-End Verification

## SUMMARY

Whereas computer matching involves comparing records after an individual is already receiving government benefits or services, front-end verification is used to certify the accuracy and completeness of personal information at the time an individual applies for government benefits, employment, or services. Like computer matching, any large-scale application of front-end verification is dependent on computers and telecommunication systems.

OTA found that:

- The use of front-end verification is creating a de *facto* national database covering nearly all Americans. The technological requisites for front-end verification lead to the establishment of individual databases for verification purposes and to the connection of these databases through on-line telecommunication linkages.
- There is no comprehensive information on the use of front-end verification by Federal agencies. Front-end verification is used by many States, mostly in federally funded programs, and is initiated or required by the Federal Government. Legislation, either recently enacted and/or proposed, will expand the use of front-end verification at the Federal as well as the State level.
- Front-end verification raises due process and privacy issues that have not been systematically studied.
- There has been no comprehensive study of how to conduct front-end verification in the most cost-effective manner and with the highest possible data quality.
- There are no general Federal regulations, either statutory or administrative, guiding the use of front-end verification. In designing guidelines, a number of factors warrant consideration, including:
  - —the responsibility for determining access to and record quality of the databases used for verification purposes;
  - —the frequency of front-end verification, i.e., routine or selective;
  - –the rights of individuals;
  - –the types of information used; and
  - —the possible requirement of a cost-benefit analysis.

## INTRODUCTION AND BACKGROUND

Computer-assisted front-end verification is used to certify the accuracy and completeness of personal information by checking it against similar information held in a computerized database, generally of a third party. It may involve certifying information that the individual has supplied, checking a database to determine if there is additional relevant information, or both. Front-end verification is used when an individual initially applies for government benefits, employment, credit, contracts, or some other government program or service. In the past, such verification was done manually on a random basis or when the accuracy of information provided was suspect. Today, the number of applications and details to be verified makes manual verification prohibitive in terms of cost and time; however, computerized databases and on-line networking make it possible to carry out such verification routinely.

Front-end verification is similar to computer matching in that it involves an electronic search for the purpose of ensuring the accuracy and completeness of information to maintain

the integrity of government programs. However, front-end verification differs from computer matching in four ways: 1) information is verified on an individual basis, rather than for a category or class of people; 2) information is verified before an individual receives any government benefits or employment; 3) its purpose is to prevent and deter, rather than to detect and punish; and 4) it is done most effectively at the time of the initial transaction, and thus accelerates the trend to on-line data linkages. For these reasons, some of the policy issues (e.g., data quality, cost-effectiveness, and administrative discretion) are essentially the same for both front-end verification and computer matching. However, other issues, such as due process and privacy concerns, are different for front-end verification than for matching.

Computer-assisted front-end verification can be done in two ways–by batch processing or by a direct on-line inquiry. If batch processing is used, the agency compiles (usually on magnetic tape) all information needing a specific type of verification, either at the end of the day or week, and sends it to the relevant source for verification. A tape-to-tape match reveals inconsistencies in the data. The second method is a direct on-line inquiry from an agency terminal to the computerized source database as each individual case is considered. An immediate on-line response reveals inconsistencies in the data. Because of its speed and efficiency, the trend is toward more direct on-line verification. For example, the Department of Health and Human Services found that 73 percent of front-end verification in the Aid to Families With Dependent Children (AFDC), food stamp, and Medicaid programs at the State level was conducted on-line.[1]

[1]U.S. Department of Health and Human Services, Office of inspector General, Catalog *of Automated Front-End Eligibility Verification Techniques: A Project of the President Council on Integrity and Efficiency, OAI-85-H-51,* September 1985, p. 13.

# FINDINGS

## Finding 1

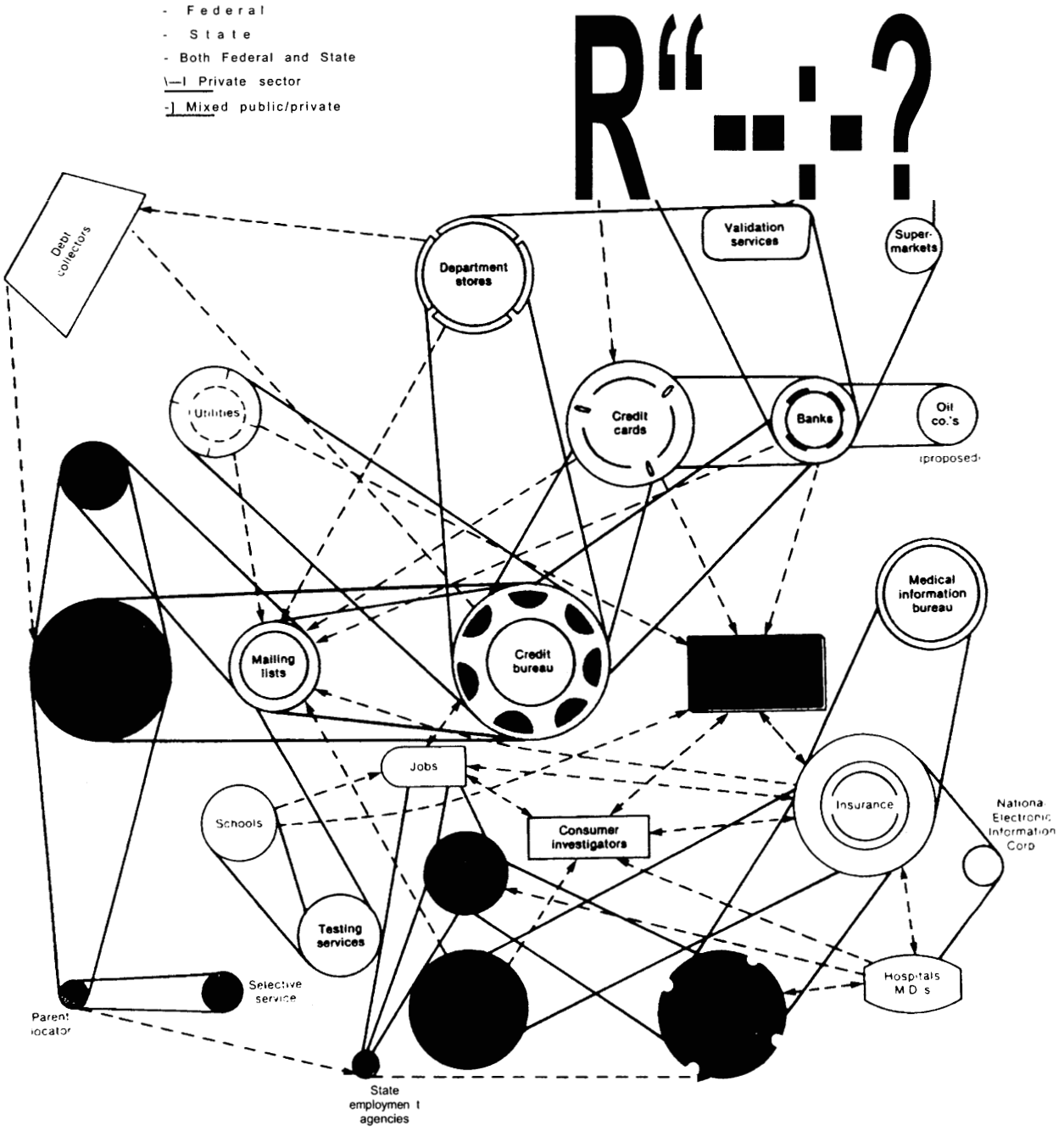
The use of front-end verification is creating a de ***facto*** national database covering nearly all Americans. The technological requisites for front-end verification lead to the establishment of individual databases for verification purposes and to the connection of these databases through on-line telecommunication linkages.

This de *facto* national database is not a centralized database in the sense that all information is contained in one mainframe computer housed in one building. Instead, the present dominant approach is to create a "virtual" central databank by electronically (via direct on-line linkages[2] or exchange of computer tapes) combining and comparing information from several separate, usually remote, record systems. If enough separate record systems are queried, the result can be the creation of a ***de*** *facto* electronic dossier on specific individuals. See figures 5 and 6 for attempts to portray the current state of computerized linkages among separate databases.

Part of the explanation for this decentralized approach to databanks and dossiers, rather than a centralized approach, is that advances in computer and data communication technology have reduced the technical and cost barriers to such interconnections. However, part of the explanation is also political in nature. The decentralized approach reflects the fragmented and complex structure of the executive branch of the Federal Government. Although Federal agencies may collect and use similar information on individuals, they also collect information that is specific to their missions and would prefer to maintain their own

[2]On-line telecommunication linkages involve data communications, the contents of which are not protected by existing statutory (e.g., Title III of the Omnibus Crime Control and Safe Streets Act) and constitutional prohibitions on the interception of phone calls. See U.S. Congress, Office of Technology Assessment, *Federal Government Information Technology: Electronic Surveillance and Civil Liberties, OTA-CIT-293* (Washington, DC: U.S. Government Printing Office, October 1985).

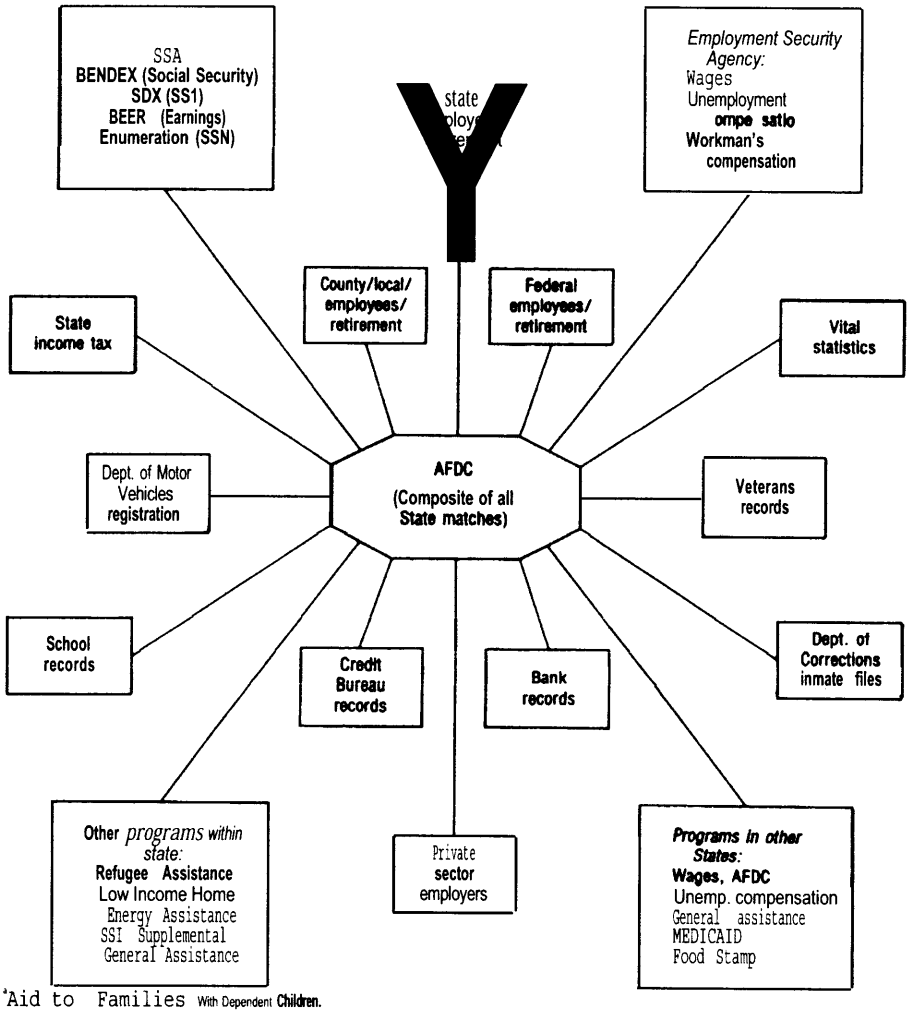
**Figure 5.—Current Database Linkages**




**NOTES** Solid lines=automated exchanges, dotted lines=manual exchanges

**SOURCE** *The Privacy Journal,* April 1984, p 5

**Figure 6.–Composite of Data Linkages Through Computer Matches by AFDC*Programs in Various-States**




*Aid to Families With Dependent Children.

**NOTE" No single State has all of these links, but each link** occurs in at least one State. With a few exceptions, however, these types of sources could be available in every State

SOURCE U S Department of Health and Human Services, Office of Inspector General, Inventory of State Computer Matching Technology, and GAO observation.

databases for their clients or employees. Additionally, the decentralized approach reflects incremental responses to policy problems. Databases usually are created to deal with a specific problem as seen at a particular time. Rarely is the opportunity taken to review related problems and look for a common solution.

The decentralized approach also reflects political concerns frequently expressed about centralized databanks and dossiexs. Indeed, when proposals for various national databanks were first made 15 to 20 years ago, the reaction was quite negative. Concern was expressed that, even if central databanks were technically fea-

sible, they might be more open to abuse, and might consolidate power and control in the Federal Government.[3] Since that time, few proposals for national databanks of personal information have been made or seriously considered. In cases where there has been a serious debate, the common result has been a decentralized approach. Two cases in point are the Interstate Identification Index (known as Triple I), run by the Federal Bureau of Investigation (FBI), and the National Drivers Register (NDR) run by the Department of Transportation's National Highway Traffic Safety Administration (NHTSA).

In both of these situations, proposals to maintain central databanks (on criminal history records and motor vehicle operator records, respectively) run by the Federal Government were strongly opposed by various States and civil liberty groups and ultimately defeated, even after partial implementation. In both cases, a decentralized index approach was adopted (with support from the States and civil liberty groups) as an alternative to the central databank approach. In the index approach, the Federal Government (in these examples, the FBI and NHTSA) maintains, in effect, an index to records in State record systems. Only names and identifiers are contained in the index–it does not include information about specific offenses, charges, and dispositions (for criminal history records indexed by the Triple I) or about specific driver violations and license suspensions (for vehicle operator records indexed by NDR).

The NDR contains 10 million records with information on drivers' licenses that have been revoked or suspended in various States. NDR is a voluntary Federal/State cooperative program to aid States in exchanging information about the driving records of certain individuals. Currently all States participate in reporting license withdrawals, submitting names to be checked against the NDR file, or both. NDR has been in operation since 1961 under the authority of Public Law 86-660, which directed the Secretary of Commerce to establish a register of all names of individuals reported by the States for revocation of a driver's license because of driving while intoxicated or violation of a highway safety code involving loss of life. Until 1982, reports on license withdrawals and denials contained descriptive information about the individual and details of the adverse action taken. The National Drivers Register Act of 1982 (Public Law 97-364) requires that the content of the Federal NDR file be limited to minimal, personal, identifying information with case-specific information being maintained only by the State instituting the adverse action. The 1982 law also converted NDR to a fully automated system.

The FBI's Triple I, which became operational on February 7, 1983, contained 9,268,332 records as of May 1, 1985.[4] Triple I is essentially an index of persons with criminal history records on file at the FBI and/or in State criminal history record repositories. For each person listed, Triple I includes only information on personal descriptors, identifying numbers, and the location(s) of the criminal history record(s). At present, use of Triple I is limited to criminal justice and criminal justice employment purposes, although the question of noncriminal justice use (primarily for employment and licensing checks) has not been resolved (see app. A at the end of this report for further dis-

[3] See U.S. Congress, House Committee on Government Operations, Special Subcommittee on Invasion of Privacy, *The Computer and Invasion of Privacy*, hearings, 89th Cong., 2d sess., July 25, 27, 28, 1966 (Washington, DC: U.S. Government Printing Office, 1966); and U.S. Congress, Senate Committee on the Judiciary, Subcommittee on Administrative practice and Procedure, *Invasion of Privacy*, hearings, 89th Cong., February 1965 to June 1966 (Washington, DC: U. S. Government Printing Office, 1965-67).

[4] FBI response to OTA Federal Agency Data Request. Also see U.S. Department of Justice, Federal Bureau of Investigation, Technical Services Division, *Statement of Work for NCIC 2000 (2K) Project—PHASE I: A Comprehensive Study To Define: System Requirements, Functioned Design and System Specs (Consistent With a Rigorous Environmental Analysis Evaluation)*, January 1985, p. A9; and David F. Nemecek, "The Interstate Identification Index (III)," *Interface*, SEARCH Group, Inc., 101.9, No. 1, summer 1984, pp. 10-11.

cussion). If authorized criminal justice agencies obtain a "hit" or match on Triple I, the agencies obtain the actual criminal history record information from the FBI (for Federal offenders and offenders from States not yet participating in Triple I) or from State criminal record repositories (for Triple I participants). Triple I inquiries are made electronically via the National Crime Information Center's (NCIC) communication lines and, if a hit occurs, are referred or switched automatically to the appropriate holder of the original criminal history record. Records are provided by one or a combination of the following: on-line via NCIC, electronically from a State via the National Law Enforcement Telecommunications System, or by mail from the FBI or State repository.

Triple I represents an alternative to the now-defunct Computerized Criminal History (CCH) file previously maintained in NCIC. By including index entries for computerized criminal history records maintained by the FBI's Identification Division, as well as records from participating States, Triple I has been able to facilitate access to and exchange of over 9 million criminal history records, compared to the roughly 2 million records contained in the old NCIC/CCH file. However, there still are several unresolved issues concerning Triple I—noncriminal justice use, record quality, and policy oversight. These are discussed in further detail in appendix A to this report.

The decentralized approach in these instances is generally perceived as minimizing adverse impacts on Federal-State relations, since the States retain primary control over the source records. Also, the risk of abuse or misuse by the Federal Government is thought to be lessened, since there is no central file. However, authorized Federal, State, and local agencies can determine, via the index, the location of records of interest and request such records directly from the State record repositories. Thus, a dossier on any given individual can be compiled by consolidating various records from separate State agencies. It is also possible for Federal agencies to run a longer list of persons against the index to see if there are any matches, or "hits," and then follow up to obtain more detailed information.

Agencies may also maintain a centralized index of individuals whose records are maintained in their computerized databases. For example, the OTA survey revealed that the Immigration and Naturalization Service (INS) has a Central Index System (CIS) of 152 million records that contains file location, immigration status, and biographical data on individuals of interest to INS. On-line access to CIS is provided at ports of entry, file control offices, border patrol headquarters, and other agencies involved in intelligence or law enforcement. On an average, 600 users generate 100,000 file accesses per day.

Although electronically linked, on-line databases are distributed in a physical sense, they constitute a centralized database in a practical sense. As more and more systems automate and have on-line communication capability, this virtual database will grow. There are a number of computerized databases that are accessible by selected government agencies for computer-assisted verifications—for example, the computer files of the FBI's NCIC and those of the Bureau of the Customs' Treasury Enforcement Communication System. INS maintains a number of computerized record systems —the Anti-Smuggling Information System, the Central Index System, the Non-Immigrant Information System, the Student School System, and the National Automated Immigration Lookout System. The Social Security Administration (SSA) also maintains a number of databases for verification purposes—the State Data Exchange, the Beneficiary and Earnings Data Exchange, the Third Party Query, and the Enumeration Search and Verification Response System. Additionally, private sector firms, such as credit bureaus and medical insurers, maintain a number of centralized databases that are accessible by government agencies. See table 13 for a description of these databases.

Centralized databases are also created from existing decentralized databases. One example is the IRS's Debtor Master File, which was

**Table 13.—Computerized Databases Used for Front-End Verification**

*National Crime Information* Center (A/C/C) .-There are 12 files containing a total of 16,395,662 files (as of 5/1/85) that can be accessed through the NCIC system.[a] The 12 files include: the Interstate Identification Index (III) File, the Stolen Securities File, the Stolen Guns File, the Stolen Articles File, the Stolen Vehicles File, the Stolen License Plates File, the Wanted Persons File, the Missing Persons File, the Stolen Boats File, the Canadian Warrant File, the U.S. Secret Service Protective File, and the Unidentified Persons File. NCIC functions as a nationwide computerized Information service for Federal, State, and local criminal justice agencies.

*Treasury Enforcement Communication System (TECS).*–Includes a range of information on persons suspected of, or wanted for, violations of U.S. Customs or related laws —e. g., persons suspected of or wanted for thefts from international commerce, and persons with outstanding Federal or State warrants, The Border Enforcement System is the major component and is used to: assist Customs and the Immigration and Naturalization Service (INS) personnel screen persons and property entering and exiting the United States; provide investigative data to Customs or other agency law enforcement or intelligence officers; and aid i n the exchange of data with other Federal, State, or local law enforcement agencies. As of May 1, 1985, the Border Enforcement System included computerized records on over 2 million persons.

*Nonimmigrarft Iformation System (NI/S),* —Contains over 32 million records on foreign visitors, diplomats, and students for purposes of tracking their movements, The system has been operational since January 1983, The student/schools subsystem became operational in August 1984 and tracks 500,000 students at 15,000 schools,

*Anti-Smuggling Information System (AS/S).* —Incorporates 750,000 records containing information relating to alien smugglers, including names (and aliases), addresses, phone numbers, and license plates.

*National Automated Immigration Lookout System (NAILS).*—Provides on-line information for the detection of inadmissible persons and others of particular interest to INS and other law enforcement agencies. Presently contains 40,000 records.

*State Data Exchange (SDX—Social Security Administration* [SSA]).–Contains 7.5 million records with title XVI information extracted from the supplemental security record, as well as Medicaid eligibility data for specified States. SDX has been in operation since December 1973 and is accessible by State Welfare/Human Resources Departments for use in adminitration of income maintenance and Medicaid programs.

*Beneficiary and Earnings Data Exchange (BENDEX–SSA),* –Contains *64* million records with information on title II eligibility, Medicare entitlement, wage data, and eligibility entitlement to other SSA-administered programs. BENDEX has been in operation since 1968 and is accessible by State Welfare/Human Resources Departments for use In administration of income maintenance programs.

*Third Party Query (TPQY—SSA).* —Contains the 7.5 million SDX records and the 64 million BENDEX records. TPQY has been in operation since November 1984 and is accessible for purposes of speeding up the SSA-administered benefit verification process by all State, local, and Federal agencies that administer a health and/or income maintenance program (including commercial vendors).

*Enumeration Search Verification and Response System (ESVARS—SSA).* –Contains identification data for every social security number that has been issued. There are 280 million base records, which are expanded to 420 million iterations because of name changes, duplicate cards, and such, ESVARS has been in operation since Apr. 1, 1985 and is accessible by all SSA employees who need to verify social security numbers and Federal, State, local, and private agencies that justify their need to verify social security numbers.

[a]For furt her di sc uSSi on see app A at the end of this report Also see US Con gress Office o' Technology Assessment, *An Assess men t o' A lternati ves for a National Computerized Criminal History* System OTA CIT-161 (Springfield VA National Technical Information Service October 1982)

SOURCE Off ice of Technology Assessment

created in 1986 using information from the databases of a number of agencies. The Debtor Master File was authorized in the Deficit Reduction Act. The purpose of the Debtor Master File is to aid in administering the offset of tax refunds to collect on delinquent Federal debts, such as student loans.[5] The 1986 Debtor Master File contains the names of 750,000 individuals who are indebted to at least one of the following agencies: the Departments of Education, Housing and Urban Development, or Agriculture; the Veterans Administration; and the Small Business Administration. Preoffset notices were sent to these individuals and resulted in payments from 41,000 persons totaling $14 million.[6]

As the exchange of information becomes faster and easier, there will be pressure to increase computer connections and on-line processing. The Deficit Reduction Act and the establishment of Income Eligibility Verification Systems (IEVS) is a good example (see app. E of this report). Under the rules issued by the De-

[5]U.S. Department of the Treasury, Internal Revenue Service, "Privacy Act of 1974: System of Records, " *Federal Register*, vol. 50, No. 195, Oct. 8, 1985, p. 41085.

[6]See David Bumham, " I.R.S. To Withhold Tax Refunds Owed Loan Defaulters, " *New York Times*, Jan. 10, 1986, pp. Al, Al 1; Keith B. Richburg, "Agencies Give Defaulters' Names to IRS, " *Washington Post*, Jan. 10, 1986, p. A21; and Judith A. Sullivan, "IRS To Collect Agencies' Debts, " *Government Computer News*, Sept. 13, 1985, pp. 1, 16.

partments of Labor, Agriculture, and Health and Human Services,[7] IEVS would contain wage and benefit data from State Wage Information Collection Agencies; wage, benefit, and other income data from SSA; and unearned income data from the Internal Revenue Service (IRS). The Deficit Reduction Act requires each State to establish an Income Eligibility Verification System. The rules do not interpret this as mandating a physical system, but a logical process that would assure timely and efficient exchange of data. Compatibility to allow exchanges of data among various IEVS is a possibility. The Deficit Reduction Act also requires each State to collect quarterly wage reports from all employers and to establish a State Wage Information Collection Agency that will maintain records of social security numbers; full name; quarterly wages; and employer's name, address, and identifier. As of 1982, 12 States did not collect wage information on a quarterly basis.[8]

The result of IEVS will be uniformity among State systems. The Department of Agriculture has agreed that State Wage Information Collection Agencies should collect the following information: social security number; full name; quarterly wages; and employer's name, address, and identifier. Additionally, the need to follow specific guidelines in accessing IRS-and SSA information will also create more uniform systems throughout the States, and is tantamount to the establishment of a de *facto* wage and eligibility recipient system. In the congressional debates on the Deficit Reduction Act there was no explicit discussion of such a system.

[7]Departments of Labor, Agriculture, and Health and Human Services, "Income & Eligibility Verification Procedures for Food Stamps, Aid to Families With Dependent Children, State Administered Adult Assistance, Medicaid and Unemployment Compensation Programs: Final Rule, " *Federal Register,* vol. 51, No. 40, Feb. 28, 1986, pp. 7178-7217.

[8]U.S. Congress, Hearings Before the Senate Committee on Governmental Affairs, Subcommittee on Oversight of Government Management, *Oversight of Computer Matching To Detect Fraud and Mismanagement in Government Programs,* Dec. 15-16, 1982 (Washington, DC: U.S. Government Printing Office, 1982), p. 14.

## Finding 2

There is no comprehensive information on the use of front-end verification by Federal agencies, although the Federal Government is increasingly requiring front-end verification in many federally funded programs administered by the States. Recently enacted legislation will expand the use of front-end verification at the Federal as well as the State level.

Because the personal information provided by applicants for government programs is often inaccurate or incomplete, front-end verification is useful for checking eligibility for Federal benefit programs, checking on current debts and earnings for loan applicants, and checking financial and criminal histories for employment applicants.

The existence of the numerous computerized databases discussed above would seem to indicate that many agencies use front-end verification. However, only two agencies–the Bureau of Indian Affairs in the Department of the Interior and the Veterans Administration—responded affirmatively to the OTA survey's question on front-end verification. In part, the small number of affirmative responses to the question may be attributed to a lack of understanding of what would be termed "front-end verification. "

Until recently, there was almost no information on State use of front-end verification. However, the Department of Health and Human Services has recently completed a survey of automated front-end eligibility y verification applications currently used or being developed at the State level for use in AFDC, food stamp, Medicaid, and unemployment insurance programs. With a 92 percent response rate from the States, the survey found 75 front-end verification applications being used in AFDC, food stamp, and Medicaid programs in 36 States, and 53 front-end verification applications being used in unemployment insurance programs

in 36 States.[9] The primary data checked in these front-end verifications include duplicate benefits, earned income, and work history. Examples of some front-end verification programs appear in table 14.

There has been a marked increase in State use of front-end verification in Federal welfare programs. Federal statutes, most notably the Deficit Reduction Act, now require front-end verification in certain programs. The Deficit Reduction Act requires States to use front-end verification in administering the food stamp, AFDC, unemployment compensation, Medicaid, and SSA'S adult assistance programs (titles I, X, XIV, XVI). The sources that will be used most frequently for verifying information are: the agency's own data sources, as a check on duplicate benefits; SSA'S State Data Exchange System (SDX), which contains a listing of all supplemental security income recipients in the State; the SSA'S Beneficiary and Earnings Data Exchange (BENDEX), which contains wage data and eligibility entitlements to SSA programs; SSA'S Enumeration Verification System (EVS), which contains information on social security numbers; IRS files for earned and unearned income; INS files for immigration status; and State wage data systems (see fig. 7).

Under the rules developed by the Departments of Labor, Agriculture, and Health and Human Services, States are required to develop a statewide IEVS, and to use SSA and IRS systems for verifying additional information. Examples of front-end verification required under the Deficit Reduction Act include verification of: social security numbers through BENDEX, SDX, or EVS; unearned income through IRS with subsequent verification from the individual or source of unearned income; and income/wages through IEVS.[10]

[9] U.S. Department of Health and Human Services. *Catalog of Automated Front-End Eligibility Verification Techniques*, op. cit.

[10] See app. E of this report.

**Table 14.— Examples of State Front-End Verification Programs**

*Nevada.* —The Welfare Referral System under development will provide the caseworker with information about the applicant's receipt of income assistance benefits, wages, and unemployment compensation benefits (UC B). When an applicant comes into the local office, the worker will enter the applicant's name, social security number, and other data into the "key file. " This information will be matched on-line against welfare and wage and UCB data (welfare refers to Aid to Families With Dependent Children (A FDC), food stamps, Medicaid, child support, and social services). A hardcopy of the match WIII be generated and transmitted to the worker.

Georgia.—At the time of application, the eligibility worker does an on-line check of the current recipient database to detect any duplicate benefits. In addition, this match is also run during the batch processing of the application that occurs immediately prior to payment. Results are received prior to eligibility certification. This batch match also accesses statewide records of closed benefit cases. The duplicate benefit check is part of Georgia's larger Public Assistance Reporting System (PARIS) designed to cotlect, store, and generate information utilized by the AFDC. food stamp, and Medicaid programs.

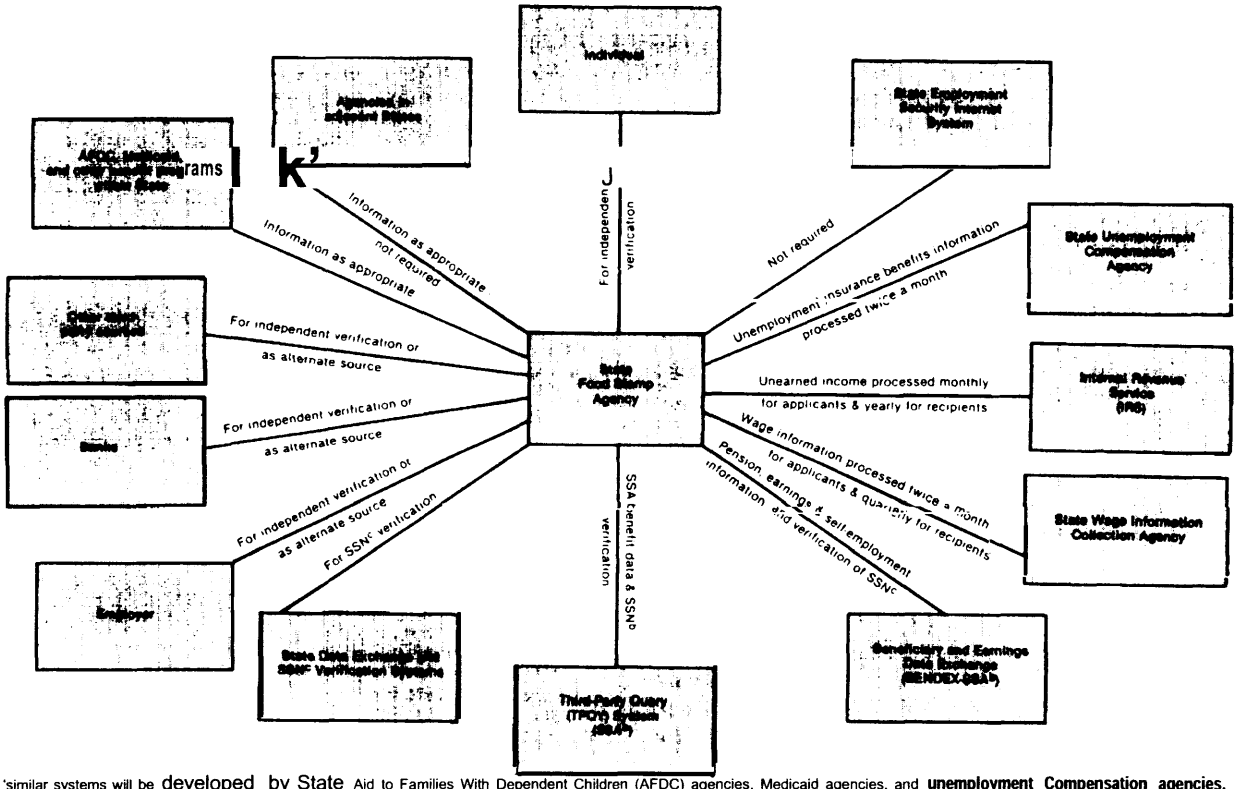
New York.—As a new subsystem of the Welfare Management System, the Resource File Integration automatically provides front-end matching of all applicants for public assistance against the State wage file. The wage data is available on-line to eligibility workers. To assure that local workers take action on the information, a resolution code indicating the action is required before any further processing can take place. Future plans call for adding State UCB data to the resource file. This system is used statewide except in New York City, which has a slightly different system providing the same information by overnight batch processing.

*Florida.* —Information on individuals who are known to have been involved in labor disputes and who have committed benefit fraud is stored in the claim history file When an individual applies for unemployment compensation benefits, employees automatically perform an on-line match between this data and applicant data when they enter data from a new application. Positive hits generate flags that prevent any payments from being made until the issue is resolved.

SOURCE U S Department of Health a-rid Human Services, Office of Inspector General, Catalog of *Automated Front-End Eligibility* Verification Techniques, OAI-85-H-51 September 1985

The Debt Collection Act requires applicants for Federal loans to supply their taxpayer identification number (for individuals, their social security numbers), and requires agencies to screen credit applicants against IRS files to check for tax delinquency. Circular A-70 of the Office of Management and Budget (OMB)

**Figure 7.—A Representative Income and Eligibility Verification System (IEVS) for a State Food Stamp Agency as Required by the Deficit Reduction Act of 1984[a]**




[a]similar systems will be developed by State Aid to Families With Dependent Children (AFDC) agencies, Medicaid agencies, and unemployment Compensation agencies. as well as for the Adult Assistance Program in the Territories,
[b]Social Security Administration.
[c]Social security number.

SOURCE: Office of Technology Assessment.

mandates that Federal agencies must conduct a credit screen on a potential candidate before issuing a contract, grant, loan, or loan guarantee.

With debt collection and with credit screening, the Federal Government is relying on private sector databases for verifying the information. As presently planned, five companies, including TRW Information Services, will develop databanks on individuals' credit and debt information from private and governmental sources, and two companies, TRW and Dun & Bradstreet, will do likewise for commercial firms.11 Dun & Bradstreet's Director of Corporate Government Services was quoted as saying:

> Private lenders, banks, etc., who are Dun & Bradstreet subscribers can get this data, too. So, if you don't pay the Feds, from now on it'll affect your commercial credit rating, too.[12]

There has also been an increased effort to require criminal history record checks for job applicants in sensitive categories, e.g., day-care providers for children. Congress included a pro vision in the Continuing Appropriation Act of 1985 (Public Law 98-473) requiring that States establish procedures to provide for nationwide criminal history checks for all operators and

[11]"Front-End Credit Screening: How an Ounce of Prevention Could Avoid Billions in Cure," *Government Executive*, January 1985, pp. 34-35.

[12]Ibid, p. 35.

employees of child-care facilities.[13] States were to have such procedures in place by September 30, 1985. '4 According to the Office of the Inspector General, U.S. Department of Health and Human Services, as of November 1984, 3 States (California Georgia Minnesota) had statutes requiring FBI criminal record checks on day-care providers, 24 States conducted statewide criminal record checks on day-care providers, and 20 States were anticipating new legislation authorizing such criminal record checks.[15] There has also been growing interest in implementing criminal record checks for teachers, youth group leaders, and elder-care providers. 'G

IRS files are also considered to be valuable sources of information for many record linkages because of the variety of information on file (e.g., address, earned income, unearned income, social security number, and number of dependents) and because the information is relatively up to date. As a general rule, returns and return information are to remain confidential, as provided for in Section 6103 of the Tax Reform Act of 1976. Under this section, information may be disclosed for tax and audit purposes and proceedings, and for use in criminal investigations if certain procedural safeguards are met.

Additionally, Section 6103(1) allows for the disclosure of return information for purposes other than tax administration. The list has grown considerably since 1976, and includes disclosures to: SSA and the Railroad Retirement Board (Public Law 94-455, 1976); Federal loan agencies regarding tax delinquent accounts (Public Law 97-365, 1982); the Department of Treasury for use in personnel or claimant representative matters (Public Law 98-369, 1984); Federal, State, and local child support enforcement agencies (Public Law 94-455, 1976); and Federal, State, and local agencies ad-ministering certain programs under the Social Security Act or Food Stamp Act of 1977 (Public Law 98-369, 1984). Section 2651 of the Deficit Reduction Act also amends Section 6103(1) of the Tax Reform Act and allows return information from W-2S and unearned income reported on 1099s to be divulged to any Federal, State, or local agency administering one of the following programs: AFDC; medical assistance; supplemental security income; unemployment compensation; food stamps; State-administered supplementary payments; and any benefit provided under a State plan approved under Titles I, X, XIV, or XVI of the Social Security Act. Section 6103(m) of the Tax Reform Act also provides for disclosure of taxpayer identity information to a number of agencies, including the National Institute for Occupational Safety and Health and the Secretary of Education.

Pressure to extend the list of agencies that can access IRS information has intensified with interest in record linkages to detect fraud, waste, and abuse; to register men for the Selective Service; and for any program that needs a current address for an individual. The IRS's position is that its goal is to maintain a voluntary tax system and that public perception that tax information be confidential is important to maintaining a voluntary system. Thus, the IRS is, in principle, opposed to disclosing tax information.

The potential for expanding the use of front-end verification for government programs, loans, and employment is enormous, as evidenced by the Reagan Administration's proposed Payment Integrity Act that would require front-end verification in 12 new programs, including Pen Grants, guaranteed student loans, school lunches, health education loans, veterans' programs, Department of Housing and Urban Development housing programs, and railroad retirement. Additionally, the Administration would expand the types of data

[13] U.S. Department of Health and Human Services, *Model Child Care Standards Act- Guidance to States To Prevent Child Abuse in Day Care Facilities*, Washington, DC, January 1985, p. 2.

[14] 1 bid., p. 3.

[15] 1 bid., p. 27.

[16] See, for example, Adrian Higgins, "Day Care Worker Checks Getting Mixed Reviews, *Arlington Journal*, Sept. 6, 1985, p. A7; Linda Lantor, "Fairfax Schools To Tighten Employee Screening, " *Arlington Journal*, Sept. 10, 1985; p. A4; and Andee Hochman, "Youth Workers Face Additional Screening; Change Follows Spate of Sex Abuse Cases, *The Washington Post*, Sept. 23, 1985, pp. D1-D2.

available for verification beyond those specified in the Deficit Reduction Act to include alien status, government wages and pensions, veterans' benefits, and railroad retirement.

Another section of the proposed Payment Integrity Act would set up a Health Insurance Verification System that would enable federally funded health care programs to access third-party insurance files to verify information supplied by the person applying for insurance payments. The Federal programs include Medicaid, Medicare, Veterans, Indian Health, Black Lung, and Maternal and Child Health. The third-party insurance files to be accessed include private insurance companies, health maintenance organizations, self-insured employer-based plans, State and local employee health plans, Federal health insurance programs, and Federal and State workers' compensation.

There are presently a number of front-end verification pilot projects being conducted at the Federal level or at the State level with Federal funds. One is the Systematic Alien Verification for Entitlements (SAVE) system operated by the Immigration and Naturalization Service. State welfare agencies can access SAVE to determine if an applicant is a legal alien. Such information was previously verified by sending individual forms to INS. SAVE in this way saves time for the applicant, although State laws generally require welfare agencies to act on an application within 10 days. However, INS also regards it as a "policing tool, as indicated by this statement in an INS memo about SAVE:

> Success will be measured by the number of criminal prosecutions resulting from these efforts; the dollars of cost avoidance; and the number of unentitled aliens identified and removed or barred from benefit rolls .17

Another pilot project is Project Checkmate in the District of Columbia. In this project, AFDC applicants are screened against credit bureau records providing information on income, resources, bank accounts, credit balances, and employment.[18]

## Finding 3

Front-end verification raises due process and privacy issues that have not been systematically studied.

Under traditional due process principles, it is arguable that individuals should be notified that information they provide will be verified by third-party sources.[19] In many of the front-end verification programs currently being used, individuals are not informed or are only informed indirectly, i.e., they are told that information may be verified, but not when or how. They are often left with the impression that they will be responsible for bringing proof to verify information, not that the agency will verify information from other sources (see box B).

The Deficit Reduction Act and the Debt Collection Act include requirements that agencies give some notice to individuals. The Deficit Reduction Act requires agencies to notify applicants at the time of application and periodically thereafter that information about them will be exchanged and used to verify income and eligibility. Under the proposed rules, it is not clear how this will be done ("in writing at application, but not necessarily on the application form' or how specific will be the information that is provided to the individual.

"As quoted in American Civil Liberties Union, "computer Matching-Focus Paper, " September 1985, p. 5.

[18]U.S. Department of Health and Human Services, *Catalog of Automated Front-End Eligibility Verification Techniques*, op. cit., p. 44 .

[19]Procedural due process traditionally means that an official government action must meet certain standards of fairness to an individual. This generally includes the rights of adequate notice and of a meaningful opportunity to be heard prior to a decision. In deterrnining the level of procedural due process that is appropriate, three issues are considered: 1 ) is there a threat to life, liberty, or property interests; 2) what are the interests of the government and of the individual; and 3) what procedures are cost-justified. See Kenneth C. Davis, *Administrative* Law *Treatise*, 2d ed. (San Diego, CA: K.C. Davis Publishing, 1979); Kenneth C. Davis, *Discretionary Justice: A Preliminary Inquiry* (Urbana: University of Illinois Press, 1969); and Ernest Gellhorn and Barry B. Boyer, *Administrative Law and Process* (St. Paul, MN: West Publishing, 1981).

Box B.—Example of Front-End Verification Notice

## Penalty Warning

THE INFORMATION PROVIDED ON THIS FORM WILL BE SUBJECT TO VERIFICATION BY FEDERAL, STATE AND LOCAL OFFICIALS. IF ANY IS FOUND INACCURATE, YOU MAY BE DENIED FOOD STAMPS AND/OR BE SUBJECT TO CRIMINAL PROSECUTION FOR KNOWINGLY PROVIDING FALSE INFORMATION.

ANY MEMBER OF YOUR HOUSEHOLD WHO INTENTIONALLY BREAKS ANY OF THE FOLLOWING RULES CAN BE BARRED FROM THE FOOD STAMP PROGRAM FOR 6 MONTHS AFTER THE FIRST VIOLATION, 12 MONTHS AFTER THE SECOND VIOLATION, AND PERMANENTLY FOR THE THIRD VIOLATION. THE INDIVIDUAL CAN ALSO BE FINED UP TO $10,000, IMPRISONED UP TO 5 YEARS, OR BOTH. A COURT CAN ALSO BAR AN INDIVIDUAL FOR AN ADDITIONAL 18 MONTHS FROM THE FOOD STAMP PROGRAM. THE INDIVIDUAL MAY ALSO BE SUBJECT TO FURTHER PROSECUTION UNDER OTHER APPLICABLE FEDERAL LAWS.

DO NOT give false information, or hide information, to get or continue to get food stamps.

DO NOT trade or sell food stamps or authorization cards.

DO NOT alter authorization cards to get food stamps you're not entitled to receive.

DO NOT use food stamps to buy ineligible items, such as alcoholic drinks and tobacco.

DO NOT use someone else's food stamps or authorization cards for your household.

## Your Signature

I understand the questions on this application and the penalty for hiding or giving false information or breaking any of the rules listed in the Penalty Warning. My answers are correct and complete to the best of my knowledge.

I understand that I my have to provide documents to prove what I've said. I agree to do this. If documents are not available, I agree to give the Food Stamp office the name of a person or organization they may contact to obtain the necessary proof.

Your signature

Today's date

Witness if you signed with an X

You or your representative may request a fair hearing either orally or in writing if you disagree with any action taken on your case. Your case may be presented at the hearing by any person you choose.

We will consider this application without regard to race, color, sex, age, handicap, religion, national origin or political belief.

FORM FNS-385 (7-83) *Previous Editions Obsolete*

Page 5

From prototype of food stamp application approved by the Office of Management and Budget. Actual forms vary by State.

In 1983, OMB issued its Guidelines *on the Relationship of the Debt Collection Act of 1982 to the Privacy Act of 1974.*[20] The guidelines specify that before an agency discloses information to a consumer reporting agency, the agency head or designee must review and validate the disclosure, must have given notice to the debtor of the overdue debt and its intention to disclose, must have given the individual time to file for review, and must have published

[20] Apr. 11, 1983 (effective Mar. 30, 1982) (43 FR 15556).

a notice in the *Federal Register* identifying those systems of records from which they intend to disclose. Disclosure should be limited to that information directly related to the identity of the debtor and the history of the claim. Although under the act the consumer reporting agencies receiving records are exempt from criminal liability for misuse of information, the guidelines indicate that it would be appropriate to incorporate assurances to this effect in service contracts between Federal and consumer reporting agencies. The guidelines also clarify that nothing in the wording of the Debt Collection Act authorizes agencies to share information among themselves or to use information obtained under this act for any other purpose.

In general, it can be a simple process to notify applicants that information they provide will be verified before benefits are granted and which databases will be searched for verification of which data elements. Some even envision verification being completed while the individual waits. However, there is some question whether notice is useful for the individual under these circumstances. The purpose of notice is to give the individual information so he or she can act.[21] In the case of front-end verification, notice generally leaves the individual only one recourse if he or she does not want the information verified, and that is to withdraw the application.

The exchanges of personal information necessitated by front-end verification may conflict with the Privacy Act principles that information should be collected directly from the individual and that information collected for one purpose should not be used for another purpose without the consent of the individual. Although in front-end verification information may originally be collected directly from the individual, additional information is provided from outside sources. Moreover, the information being used to verify information provided by the individual is being used for a purpose other than that for which it was originally collected.

[21]Davis, op. cit., 1979.

With respect to access to IRS information, Sections 6103(1) and (m) of the IRS code specify procedures that parties are to follow. Moreover, Federal, State, and local employees outside of IRS who handle IRS information are subject to the same criminal liabilities as IRS employees for misuse or disclosure of the information. The IRS also puts out a publication, *Tax Iformatjon Security Guidelines for Federal, State, and Local Agencies* (Publication 1075; Rev. 7-83), that describes the procedures agencies must follow to ensure adequate protection against unauthorized disclosure.

An additional due process question that is raised by verifying information from governmental or private sector (e.g., TRW or Dun & Bradstreet) databanks is: what recourse does the individual have if the information is false? Specifically, can the individual sue the databank owner or operator? The Privacy Act provides means by which individuals can take action against a Federal agency. The Fair Credit Reporting Act may provide a vehicle by which an individual could take action against a credit reporting agency. However, in other circumstances, statutes may not provide a legal means by which individuals can challenge false information and individuals would need to rely on common law defamation suits.

## Finding 4

There has been no comprehensive study of how to conduct front-end verification in the most cost-effective manner and with the highest possible data quality.

The high costs of computer matching (e.g., verifying large numbers of hits, holding hearings, and prosecuting wrongdoers) are not incurred in front-end verification. However, front-end verification has its own costs. It may add to the caseworker's time in processing an application, although it may save somewhat in subsequent administrative time. Front-end verification will increase budgets devoted to automated data processing and telecommunications. There are also some high initial overhead costs in terms of developing the databases used for verification (e.g., State Income

Verification Eligibility Systems) and getting them on-line, and ongoing costs of keeping them up to date.

The Department of Health and Human Services' survey of front-end eligibility verification techniques at the State level asked respondents about both developmental and operating costs. Most States were not able to provide the information as they were not keeping track of the administrative time devoted to verification. [22]

The major savings associated with front-end verification result from the avoidance of payments. The General Accounting Office reported that a New York State program that matched welfare applications with tax records to verify income avoided paying over $27.5 million, and that front-end verification in AFDC and food stamp programs in Arkansas saved $5 to $8 million.[23] In neither case was a detailed cost-benefit analysis available.

Another projected saving is a reduction in efforts to detect fraud, waste, and abuse for those already enrolled in government programs, as these individuals would have been initially screened by front-end verification. However, front-end verification would not eliminate the need to use other techniques (e.g., computer matching) because even when information is verified initially, frequent status changes (e.g., address and income) may necessitate later verification.

The President's Council on Integrity and Efficiency has projected that the eligibility verification required by the Deficit Reduction Act will save $1 billion over 5 years. The Congressional Budget Office did a gross estimate that confirmed this figure, but did not specify categories or figures for costs and savings.[24]

[22] Interview with Liz Handley, Project Director, Department of Health and Human Services Front-End Eligibility Project, Apr. 9, 1985.

[23] U.S. General Accounting Office, *Eligibility Verification and Privacy in Federal Benefits Programs: A Delicate Balance*, HRD-85-22, Mar. 1, 1985.

[24] U.S. Department of Health and Human Services, Office of the Inspector General, Semiannual *Report to the Congress*, Apr. 1, 1985 -Sept. 30, 1985.

The costs of front-end verification are directly tied to data quality. The timeliness of data used is an especially critical issue; for example, wage data are often between 3 and 6 months out of date by the time they are available from State wage reporting agencies. Unearned income from the IRS is not reported until a month after the end of the tax year and would not be processed and available for verification purposes until many months later. Other income data can likewise be stale. Some front-end verification systems, such as those required in the Deficit Reduction Act, require workers to manually check information that appears false. However, the costs associated with front-end verification will increase with each subsequent verification.

## Finding 5

At the present time, there are no policy guidelines for use of computer-assisted front-end verification.

There are no general Federal guidelines, statutory or administrative, guiding the use of front-end verification. The OMB computer matching guidelines specifically exclude from their purview record searches that are conducted at the application stage. The Deficit Reduction Act due process requirements for notice, verification, and hearings may provide a model for more general guidelines. In designing policy guidelines, the following factors warrant consideration:

1. *The responsibility for determining access to and record quality of the databases used for verification purposes.*

It is noteworthy that the FBI has taken the position that it has a responsibility only for the quality of the Triple I index entries, and not for the State criminal history records on which the index entries are largely based. Likewise, NHTSA officials have stated that the quality of driver's license records maintained by the States (and indexed in the NDR) is not the responsibility of NHTSA.

When records are maintained in a central Federal records repository, access and dissem-

ination generally follow applicable Federal laws and regulations. However, under a decentralized index approach, record access and dissemination are much more complicated. There are wide differences in State laws and regulations on record access and dissemination, ranging all the way from so-called "open record" States such as Florida, where many personal records maintained in State files are open to public access at a modest fee, to very restrictive States like Massachusetts, where access and dissemination are tightly controlled.

This wide disparity in approach is especially true with respect to criminal history records, but also affects many other kinds of personal records maintained in State repositories. This contributes to inconsistent and incomplete exchange of record information. In some of the Federal social service and welfare programs, Congress has addressed this problem by requiring States to collect and exchange information as a condition of Federal funding, as discussed earlier. But in other areas such as criminal history records, while Congress previously has taken action to encourage enactment of State laws, there are wide differences among the many State laws that have been enacted.

2. *The frequency of use of front-end verification, i.e., routine or selective.*

If it is conducted routinely (e.g., for all benefit programs and Federal employment, loans, and contracts), the societal implications of subjecting to scrutiny all information submitted to the government by individuals would need to be considered. Any possible long-term societal effects, such as increased distrust between citizens and government, loss of individual responsibility, and a sophisticated governmental information infrastructure would need to be weighed against the significant budgetary savings that may be achieved by routine verification.

If front-end verification is used selectively (e.g., by law, OMB regulations, or court decisions) rather than routinely, then consideration must be given to the criteria for selecting Federal programs that may use it, the approval process for each use, and the societal groups that will be most affected. Another alternative for doing selective verification would be to select particular individuals rather than particular programs. The individuals selected for front-end verification could be chosen by a computer profile. However, profiling raises additional policy issues, as will be discussed in chapter 5.

3. *The rights of individuals.*

Based on due process principles, as well as traditional information privacy principles, individuals should be given some notice of verification and some means to challenge information if discrepancies should appear as a result of verification. There are a number of ways in which compliance with these principles could be achieved. Individuals could be informed in writing or verbally at the time they submit an application that the information supplied will be verified. Additionally, they could be given a range of details concerning the sources to be accessed in the process. Individuals wanting more details on the process or wishing to contest verification could be advised by the caseworker whom they should consult within the agency and when.

If front-end verification reveals problems with the information provided by the individual, then a process of further checking the validity of information and informing the individual of the problems could be started. The degree of individual involvement and the depth of validation may vary based on agency directives or the goodwill of caseworkers, and therefore may need to be specified in the regulations.

Once these principles are recognized in procedural protections, there may also be a need to ensure that agencies are providing the requisite notices and hearings. Some method of enforcement or automatic accounting could also be specified in the regulations. Such oversight could be conducted within the agency or by some outside body.

With respect to involving the individual in the verification of information, the Department of Education is conducting an experimental

program, the Pen Grant Electronic Pilot.[25] Under this project, Pen Grant applicants can correct or verify information on their Student Aid Reports through computer facilities at institutions or financial aid services that participate in the project. Applicants can now make corrections on their Student Aid Reports and mail them back to the Department of Education.

**4.** *The types of information used.*

This question involves whether the use of some types of information (e.g., medical history or criminal history) should be prohibited because of their sensitivity. The use of such information could be prohibited, or its use could be restricted to particular verifications, for example, use of criminal history information in screening day-care workers.

Additionally, front-end verification raises a separate and potentially more serious issue because the information is being used to make an immediate, or near immediate, decision. In order for front-end verification to be most effective, information should be up to date, accurate, and complete. However, the information in some categories, for example, unearned income and checking accounts, may change so often that the data contained in computerized databanks will rarely be up to date. Additionally, the record quality of many existing databanks that could be used in front-end verification (e.g., computerized criminal history records) is questionable.

5. *The possible requirement of a cost-benefit analysis.*

Because a major purpose of front-end verification is to cut programmatic costs, documentation of how front-end verification will achieve this may be necessary. If a cost-benefit analysis were to be required, the categories of costs and benefits to be included could be specified in regulations. The detail to which costs and benefits should be analyzed could also be specified. The degree of detail may vary depending on the category; for example, administrative costs may be more difficult to compute than telecommunication costs.

Cost-benefit analyses could be used within an agency or program for internal improvements in ongoing front-end verifications. They could also be distributed among agencies or programs for development of new front-end verifications. Additionally, they could be used within an agency or by an outside body as part of a process of approval of new front-end verifications or review of ongoing ones.

[25] U.S. Department of Education, office of Postsecondary Education, "Invitation To Participate and Closing Date for Participation in Pen Grant Electronic Pilot, *Federal Register*, vol. 50, No. 141, Tues., July 23, 1985.

Chapter 5

# Computer Profiling

# Contents


| | Page |
|---|---|
| summary | 87 |
| Background | 87 |
| Findings | 89 |
| Finding 1 | 89 |
| Finding 2 | 91 |
| Finding 3 | 93 |
| Finding 4 | 94 |

Chapter 5

# Computer Profiling

## SUMMARY

While computer profiling is not currently a subject of major policy debate, the potential policy issues raised by the future growth of computer profiling are important. In computer profiling, a record system (or record systems) is searched for a specified combination of data elements, i.e., the profile. Profiling involves the use of inductive logic to determine indicators of characteristics and or behavior patterns that are related to the occurrence of certain behavior.

A profile is developed by a government agency to select characteristics of types of individuals, and to determine the probabilities of such individuals engaging in activities or behavior of interest to that agency. For example, the Drug Enforcement Agency (DEA) has developed profiles of the types of persons more likely to be engaging in illegal drug activity; the Internal Revenue Service (IRS) has developed profiles of categories of taxpayers more likely to be under-reporting taxable income; and the Federal Bureau of Investigation (FBI) has developed profiles of violent offenders. Profiles can be valuable tools for investigative, administrative, and intelligence purposes because they reduce the population that is of interest to an agency, and thus may increase the agency's efficiency and effectiveness.

OTA found that:

- Federal agencies are currently using computer profiling and it is likely that its use will expand in the near future.
- Important privacy and constitutional implications are raised by computer profiling because prople may be treated differently before they have done anything to warrant such treatment.
- The validity of computer profiles in accurately selecting the desired subset of individuals is subject to debate, and thus also raises questions about the relevancy of data used and the appropriateness of using computer profiles for certain decisions.
- At the present time, there are no policy guidelines for agency use of computer profiling.

## BACKGROUND

Before computers were used to process and store information, systematic data on large numbers of individuals were not retained (or if retained were not readily accessible). Moreover, there was no easy means to analyze the data that did exist in order to construct profiles. Information technology in general-and computers in particular-have removed these constraints. Detailed, historical information on individuals can be compiled from various computerized databases. Computers can be used to analyze complex and disparate information and, based on that analysis, to design a profile. Additionally, computers can be used to search a record system on the basis of a profile. These technological changes make profiles both more powerful and more available. Most importantly, technology is now making possible many new profiling applications for which judgments of social acceptability have yet to be made.

Profiling involves the use of inductive logic to determine indicators of characteristics and/or behavior patterns that are related to the occurrence of certain behavior. A judgment is

made about a particular individual based on the past behavior of other individuals who appear statistically similar, that is, who have similar demographic, socioeconomic, physical, or other characteristics. Generally, in the Federal Government, the behavior of interest is actual or potential violation of a law or administrative regulation.

In the past, and as is often still the case, people who appeared suspicious or acted strangely were often watched more carefully and their stories were verified from outside sources. Searches through Federal record systems were often conducted on the basis of a list of characteristics that experience had shown were problematic. Such profiles were often crude and could easily lead to the stereotyping of individuals. Today, profiling is much more sophisticated as a result of advances in behavioral psychology and statistics. As most behavior is complex, sophisticated modeling may be done to determine the interrelations among certain indicators. There are two general models of profiling. One is singular profiling, which models distinct characteristics or activities, e.g., sex, age, income, or number of dependents. When these characteristics appear together or in a certain pattern, that individual is flagged by the profile. The second model of profiling is aggregative profiling, which is based on the frequency with which selected factors appear across cases. This model is designed to find systematic and repetitive violators. '

Profiles have been used for decisionmaking in a variety of areas, ranging from insurance and advertising to motor vehicle or real estate licensing to entrance to the medical and legal professions. Profiles used range from those that are benign and socially acceptable (e.g., granting driver's licenses to 16 year olds, who inmost States are judged to be physically and mentally mature enough to drive a car) to those that are discriminatory and socially unacceptable (e.g., denying rental housing to minorities or students or denying professional employment opportunities to women).

'Gary T. Marx and Nancy **Reichman,** "**Routinizing** the Discovery of Secrets, *American Behavioral Scientist,* vol. 27, No. 4, March/April 1984, pp. 429-431.

Profiles have been used by the government to help agencies uncover possible misrepresentation of eligibility to receive Federal funds or benefits, possible noncompliance with or violation of agency regulations, and possible violation of civil or criminal statutes. In the government, profiles can be created, to some extent, for the convenience of implementing public policies, as they replace subjective judgments with objective decisionmaking criteria. Profiles can be useful during any stage of an agency's interaction with individuals. For example, in eligibility benefit programs, profiling may be used at the application stage to determine if an applicant is likely to misrepresent his or her income, or at the redetermination stage to ascertain if it is likely that an individual's status has changed. In law enforce ment, profiling may be used in discovering likely suspects (e.g., airplane hijackers) or in determining an appropriate sentence for some one convicted of a crime. Profiles can be valuable tools for investigative, administrative, and intelligence purposes because they reduce the population that is of interest to an agency, and thus may increase the agency's efficiency and effectiveness.

Because computer profiling may result in selected individuals being treated differently from those not selected, it has raised a number of policy questions involving civil, constitutional, and equal rights considerations. The primary conflict is between the rights of the individuals selected (e.g., equal protection and due process) and the purpose of the government in using computer profiles and their effectiveness in achieving that purpose. No matter how sophisticated the profile, the question of treating people differently before they have acted remains.

Computerized profiling also introduces some very important new policy issues. If the use of computer profiling in the Federal Government were to be expanded, the long-term societal effects on behavior patterns, and the possible effects on individuality and creativity, would warrant attention. Additionally, the va-

lidity of computer profiles in accurately selecting the desired subset of individuals is subject to debate, and thus also raises questions about the relevancy of data used and the appropriateness of using computer profiles for certain decisions.

# FINDINGS

## Finding 1

Federal agencies are currently using computer profiling and it is likely that its use will expand in the near future.

Federal agencies have developed profiles for a number of purposes, mainly for identifying individuals most likely to be involved in an illegal activity or most likely to misrepresent their financial or personal situation in applying for a Federal benefit. The OTA survey revealed that 16 Federal agencies presently use computer profiling. For example, the IRS uses computer-generated generic profiles to identify potential compliance deficiencies; the Department of Education uses profiles, based on criteria including taxes paid, marital status, and size of household, to select Pen Grant applicants for validation; the Bureau of Indian Affairs profiles the public social service support and facilities usages and needs of individual corporate groups of Indians for budgetary planning and allocation of resources; and the Federal Reserve Board uses surveys of retailers and consumers to obtain statistical data concerning financial status and behavior of households and businesses, access to and use of consumer credit, asset holdings, financial practices, effect of charge card transactions, and the like.

According to the OTA survey, some agencies are planning to add this capability to existing systems. For example, the redesign of the Treasury Enforcement Communications System, known as TECS II, will incorporate profiling. The U.S. Army Criminal Investigation Command is considering developing a system of profiling potential victims and criminal offenders for use in the conduct of crime prevention surveys and in the development of investigative leads. Some agencies have conducted pilot programs of profiling that are no longer in use, for example, the Office of the Inspector General in the Department of Energy developed, with DOE Defense Programs, a pro file of the "Insider Criminal. "

The use of profiles for law enforcement purposes has been widely documented. Computers were not necessarily used in preparing these, but they are illustrative of the type of computer profiles already under development. The Drug Enforcement Agency (DEA) has developed a profile of airplane passengers likely to be smuggling drugs, and a profile to detect those transporting marijuana on trains.[3] The Coast Guard has a profile of vessels likely to be smuggling drugs into the country.[3] The Customs Bureau also has a "smuggler's profile."[4] The Federal Aviation Administration used a hijacker profile as part of its screening program at domestic airports until it began routine searches of all carry-on items and magnetometer screening of all passengers.[5]

The FBI has developed numerous profiles, including those of various violent criminals and serial murderers. This work is being expanded under the auspices of the FBI National Center for the Analysis of Violent Crimes. Also, based in large part on interviews with felons convicted of serial murders, the FBI has developed profiles of serial murderers, especially

[3]See, for example, *United States v. Johnston*, 497 F.2d 397 (9th Cir. 1974) and *United States v. Chadwick*, 393 F. Supp. 763 (D. Mass. 1975).

[4]Note, "High On the Seas: Drug Smuggling, the Fourth Amendment, and Warrantless Searches at Sea, *Harvard Law Review*, vol. 93, 1980, p. 725.

[5]See, for example, *United States v. Klein*, 592 F.2d 909 (5th Cir. 1979), and *United States v. Asbury*, 586 F.2d 973 (2d Cir. 1973).

[6]Note, "The Airport Search and the Fourth Amendment: Reconciling the Theories and Practices, " *U. C. L. A.—Alaska Law Review*, vol. 7, 1978, p. 307.

serial sex murderers.' The FBI is currently developing software for preparing computerized profiles of violent offenders, based on the concept already implemented for arson offenders in the computer-assisted Arson Information Management System (AIMS).[7] In 1983, the Office of Juvenile Justice and Delinquency Prevention of the Department of Justice funded the University of Pennsylvania School of Nursing to identify the variables that fit profiles of rapists, child molesters, and sexually exploited children.'

In the 1970s, the Law Enforcement Assistance Administration funded "pre-delinquency' programs to create computer models to identify those young people who were likely to become delinquent. The computer models or profiles included factors that were common among known delinquent youths, such as area of residence, family situation, school performance, ethnic group, and medical history. Young people who most closely matched the profile were to be given special treatment. In 1983, the Office of Juvenile Justice and Delinquency Prevention funded the Rand Corp. to develop strategies based on the "pre-delinquency' presumption.

Computer profiles can also be used as a way of avoiding errors in Federal Government eligibility and benefit programs and as a way of allocating scarce investigative resources. Based on a computer profile, caseworkers can determine during the application process which applicants may need more careful checking. Characteristics often associated with errors could include basic factors such as age, race, or education level; some combination of factors; or more indirect factors, such as length of family separation, residency, or living with a specified relative. In 1979, the Supplemental Security Income's Office of Family Assistance reported that the following characteristics were used in error-prone profiles: earned income, home ownership, age 26 to 40, recent separation, bank account, and overdue redetermination of benefits.'

In eligibility benefit programs, computer profiles or screens can also be used to search databases of recipients prior to conducting a computer match. The records that were selected by the profile would be the only ones subject to computer matching. A smaller number of records would then be matched. If the computer profile was effective in selecting those records most likely to contain errors, then the percentage of verifiable hits would increase. In this way, computer profiles or screens may make computer matching more effective and efficient. Additionally, cuts in the Federal budget may increase the pressure to use computer profiling not only to detect and prevent fraud and errors, but also to allocate the time of caseworkers or investigators.

There has been no survey of the use of computer profiles in social service programs at the Federal level. The President's Council on Integrity and Efficiency (PCIE) has released three inventories of Federal computer applications to prevent/detect fraud, waste, and mismanagement. The applications include matches, profiles, edits, scans, screens, analyses, and extracts. If one adopts the PCIE categorization, there were no profiles used prior to 1982, 13 profiles used in the period 1982-83, and five profiles used in the period 1984-85. '0 However, agencies have sometimes placed computer applications that appear to be profiles in a different category, e.g., Project Sonoma— Welfare Fraud Profile is listed as a match. Some computer screens appear to be based on a computer profile (e.g., a Department of Education screen designed to identify, by selected criteria, guaranteed student loans

'Robert K. Ressler, Ann W. Burgess, Ralph B. D'Agonstino, and John E. Douglas, "Serial Murder: A New Phenomenon of Homicide, " September 1984.

'AIMS deals both with past activities, in developing profiles on arson incidents, and possible future activities, in profiling arson-prone properties and suspects. See U.S. Fire Administration, *Arson Information Management System: Users Manual and Documentation,* Apr. 2, 1984.

"'Pre-Delinquent Funding: Deja Vu, " Privacy *Journal,* April 1984, p. 3.

"'Use of Error Prone Profiles, " *Eligibility Simplification Project,* October 1980.

'0U.S. Department of Labor, Office of Inspector General, "Inventory of Federal Computer Applications To Prevent/Detect Fraud, Waste and Mismanagement. Original distributed July 1982; supplements distributed July 1984 and January 1986,

maintained by State Guaranty Agencies that are in excess of the regulatory maximum of 10 years), while others do not (e.g., prescription payments made by Blue Cross and Blue Shield, screened to ascertain whether that company was computing and claiming Medicaid prescription drugs in accordance with Federal procedures).

Information on State use of computer profiles is also sketchy. The Carter Administration's Eligibility Simplification Project reported on the use of error-prone profiles, primarily at the State level. According to its study, West Virginia had used computer profiling, or a selective case action system, for Aid to Families With Dependent Children (AFDC), food stamp, and Medicaid cases, based on a quality-control sample generated monthly by the computer. The profile was based on a statistical method of evaluating previous error situations and was modified periodically. Reportedly, from 1973 to 1976, the case error rate and payment error rate declined by 20 percent." The Eligibility Simplification Project found similar results with the use of error-prone profiling in South Carolina and New Hampshire. The Eligibility Simplification Project found that other States appeared to be experimenting with the use of such profiles in determining social service eligibility. A survey of seven States conducted for OTA in 1984 revealed that computer profiling was not used by those States."

## Finding 2

Important privacy and constitutional implications are raised by computer profiling because people may be treated differently before they have done anything to warrant such treatment.

Computer profiles involve categorizing people based on selected criteria, and then selecting a subset of these people for special treatment. The equal protection guarantees of the fifth and 14th amendments were designed to ensure that individuals were treated in a manner similar to other individuals, and that the government not treat individuals differently simply because they were members of a group. Although the government can classify people for special treatment, it cannot do so based on impermissible criteria (e.g., race, religion, or national origin), nor can it use a classification to arbitrarily burden a group of individuals. In computer profiling, the criteria used might be those that are already viewed as discriminatory under existing law—e.g., race, religion, national origin, and sex. For example, in DEA's drug courier profile, being Hispanic has appeared as one of the criteria. With sophisticated profiling, it may also be possible to use a number of related indicators rather than a category whose use would be illegal.

The equal protection clauses may also require that the criteria on which the profile is based be related to the behavior in question; otherwise, the selected group may be arbitrarily burdened. Additionally, the government program would need to be rationally related to achieving a legitimate purpose such as detecting fraud, waste, and abuse or apprehending drug smugglers.

The use of computer profiling may also conflict with the due process clauses of the fifth and 14th amendments that protect an individual against arbitrary treatment and provide an individual with certain procedural guarantees. Some argue that computer profiles eliminate the discretion and arbitrariness of investigative authorities, caseworkers, and parole officers. Others respond that profiles merely replace a crude form of profiling (hunches, for example) with a more sophisticated one. In either case, the due process clauses require rules and procedures to limit discretion and protect individuals from arbitrary treatment. In some instances, use of computer profiling may not provide for adequate rules and procedures.

With respect to the use of profiles in eligibility programs, Senator William Cohen reported that:

"Ibid.

"Robert Ellis Smith, "Report on Data Protection and Privacy in Seven Selected States, " OTA contractor report, February 1985. The seven States are California, Florida, Indiana, Minnesota, New York, Texas, and Virginia.

> We have profiles that have been developed by computer, and disability payments that have been discontinued with no human contact coming about until such time as those cases are appealed to an administrative law judge. Two-thirds of the cases appealed are being reversed.[13]

The extreme result of a computer profile would be that benefits are terminated, which would not occur without a hearing. The more common result would be that a selected individual is subject to a more thorough investigation than others because he or she fits a profile. To some extent, this individual is regarded with suspicion based on the profile. Individuals may not know that they are being treated differently, and even if they do, may not know why.

With respect to the use of computer profiles in law enforcement, the primary issue is whether fitting a profile constitutes probable cause or reasonable suspicion and is reason to search or detain an individual. In determining whether an investigative stop is lawful, the courts balance the need for the search against the intrusion to the person. To justify the intrusion, law enforcement agents must be able to identify specific and articulable f acts that show the intrusion is reasonably warranted."

There have been a number of court cases involving the use of the drug courier profile, and, hence, this will serve as an example of the legal issues that arise with use of profiles for law enforcement purposes. Although this profile is not currently generated by a computer nor are computers necessarily used to search relevant databases, the legal issues would be similar whether or not a computer was involved. Agents typically use the drug courier profile as a tool in conducting surveillance on a group of people, generally those boarding or departing a plane. If agents see a person whose behavior fits a number of criteria in the profile, then they follow the person. If agents believe it is justified, they stop the individual, identify themselves as law enforcement agents, and request to see identification. Based on the information revealed and the behavior of the person, the agents may then "request" that the suspect accompany them to an office in the airport. There the person is told that he or she is suspected of carrying drugs, advised of his or her rights, and asked for permission to search his or her luggage and person.[15]

In cases in which the sole or primary justification for an investigative stop has been the drug courier profile, the lower courts have not been consistent in their rulings. For example, in *United States v. McCaleb, 552* F.2d 717 (6th Cir. 1977) and *State v. Washington, 364 So.* 2d 958 (La. 1979), the courts reversed the appellants' convictions based on investigative stops triggered by meeting a drug courier profile because their activities were too consistent with innocent behavior. In *United States v. Vasquez, 612* F.2d 1338, an investigative stop based in part on a profile was judged valid.

In 1979, the Supreme Court ruled on two instances involving the use of the drug courier profile. In the first case, *United States v. Mendenhall, 446* U.S. 544, the Court ruled that the investigative stop of Mendenhall, which was based on her fitting characteristics of the drug courier profile, was constitutional. However, the majority did not agree on why it was constitutional, giving little guidance to the lower courts on the acceptability of the profile in establishing justification for an investigative stop. One month later, the Court handed down

---

"Senate Committee on Governmental Affairs, Subcommittee on Oversight of Government Management, *Oversight of Computer Matching To Detect Fraud and Mismanagement in Government Programs*, hearings, Dec. 15-16, 1982 (Washington, DC: U.S. Government Printing Office, 1982), p. 17.

"*Terry v. Ohio, 392* U.S. 1 (1968).

"For a description of the profile, its use, and court cases, see William V. Conley, "*Mendenhall* and *Reid:* The Drug Courier Profile and Investigative Stops, " *University of Pittsburgh Law Review*, vol. 42, summer 1981, pp. 835-867; Hon. Mark A. Costantino, Vito A. Cannavo, and Ann Goldstein, "Drug Courier Profiles and Airport Stops: Is the Sky the Limit?" *Western New England Law Review*, vol. 3, 1980, p. 175; Philip S. Greene and Brian W. Wice, "The D.E.A. Drug Courier Profile: History and Analysis," *South Texas Law Journal*, vol. 22, spring 1982, p. 261; Kathleen Mahoney, "Drug Trafficking at Airports—The Judicial Response, " *University of Miami Law Review*, vol. 36, 1981, p. 91; and Francis Karl Toto, "Drug Courier Profile Stops and the Fourth Amendment: Is the Supreme Court's Case of Confusion in Its Terminal Stage?" *Suffolk University Law Review*, vol. 25, 1981, p. 217.

a second decision dealing with the drug courier profile, Reid v. Georgia, 448 U.S. 438. In this case, the Court held that the investigative stop of Reid, based on his matching characteristics of the drug courier profile, was not constitutional. The Court described the drug courier profile as "a somewhat informal compilation of characteristics believed to be typical of persons unlawfully carrying narcotics.""

Based on these two cases, the legal status of the present drug courier profile is in question. Moreover, the Reid opinion may imply that the constitutionality of the profile could turn on its sophistication. If this is true, then the use of computer-generated profiles in law enforcement may be considered a more valid investigative tool than the more informal profiles.

Federal court decisions since Mendenhall and *Reid* have not clarified the status of the use of a drug courier profile in an investigative stop. " In 1981, in *United States v. Cortez,* 101 S. Ct. 690, the Supreme Court approved use of a profile by border patrol agents to detect the smuggling of illegal aliens from Mexico to the United States.

## Finding 3

The validity of computer profiles in accurately selecting the desired subset of individuals is subject to debate, and thus also raises questions about the relevancy of data used and the appropriateness of using computer profiles for certain decisions.

Profiles vary in their complexity and in the formality of statistical techniques on which they are based. Because computers are such powerful tools in analyzing and manipulating vast quantities of data, it is likely that profiles will become even more complex and formal. Regardless of their complexity and formality, profiles by definition are prone to some degree of error, as they are merely probability statements.

In formal profiles, when a general population is characterized and a profile developed, the profile is only a statistical average of that general population. The similarities among the population will be accentuated, while the differences will be ignored. If the profile was based on a sufficiently large population, it will have *some* value in selecting those of interest, but there will also be some margin of error in the profile. The types of errors will be false positives (identifying those who fit the profile, but do not fit the category sought) and false negatives (passing by those who do not fit the profile, but do fit the category sought). In developing the profile, the statistician will incorporate the degree of error that the user is willing to tolerate.

The more informal, crude profiles are greatly influenced by the experience and concerns of those who develop them. For example, in the case of the drug courier profile, the criteria that make up the profile have varied over time and with the city in which DEA agents are working. Some subset of the following are generally considered as the profile: the use of small bills for ticket purchase, travel to and from major drug import centers, travel for short periods of time, absence of luggage or empty luggage, travel under an alias, unusual itinerary, unusual nervousness, use of public transportation, making a phone call after deplaning, leaving a fictitious callback telephone number with the airline, attempting to conceal that someone is waiting for them or that they are traveling with someone, purchase of a one-way ticket, Hispanic origin, youth, luggage without identification tags, ticket purchased at the last minute or late arrival, and deplaning last. There is no record establishing how and why these characteristics have come to be included in the profile. There may also be some criteria that DEA keeps confidential.

The OTA survey asked agencies to provide both information on the development and testing of profile programs and any evaluation reports. Of the 16 agencies that reported profil-

[18] *Reid v. Georgia, 448* U.S. 438, 440.

[19] See: *United States v. Fry,* 622 F.2d 1218 (5th Cir. 1980), *United States v. Robinson, 625* F.2d 1211 (5th Cir. 1980), and *United States v. West,* 495 F. Supp. 871 (D. Mass. 1980).

ing activities, none had this information available. There are no known studies of the degree of error in profiles used in eligibility verification programs.

A principal policy issue involves determining the accuracy of a computer profile and its effectiveness in achieving the desired outcome. The cost-effectiveness of computer profiles has never been systematically studied. There are a number of costs that may need to be considered: 1) developmental costs, including research, testing, validation, and evaluation; 2) computer costs, including hardware and software; and 3) administrative costs, including follow-up on individuals who fit the profile. The costs to individuals who may needlessly be subject to investigation may also need to be considered. Additionally, as with computer matching, there may be hidden or secondary costs that need to be examined.

There are also a number of benefits that need to be considered, primarily increasing the effectiveness and efficiency of an investigation because the relevant population has been narrowed, and preventing and deterring illegal behavior.

Some information is available on the effectiveness of profiling for law enforcement purposes. None contains specific cost-benefit categories or figures. A 1981 FBI evaluation of psychological profiling found that, of 192 cases examined, in 77 percent the profile helped focus the investigation, in 20 percent it helped locate possible suspects, and in 17 percent the profile actually identified the suspect. (Totals exceed 100 percent since more than one type of assistance may apply to a single case.) The vast majority of cases were murder or rape investigations. '8

There are some sketchy statistics on the effectiveness of the drug courier profile in selecting persons carrying drugs. In *United States v.* Van Lewis, 409 F. Supp. 535 (E.D. Mich. 1976), testimony from DEA revealed that agents at the Detroit airport had searched 141 persons in 96 encounters, found narcotics in 77 of these encounters, and arrested 122 persons. Forty-three of the searches in which narcotics were found were nonconsensual. In 15 of the 25 consent searches, no illegal narcotics were found. 'g In testimony in *United States v. Price, 599* F.2d 494 (2d Cir. 1979), a DEA agent stated that about 60 percent of those he stopped, based on the drug courier profile, were carrying narcotics. However, it appears that no national statistics are available on the effectiveness of the drug courier profile.

## Finding 4

At the present time, there are no policy guidelines for agency use of computer profiling.

The use of computer profiling raises a number of important policy questions. In determining the appropriate use of computer profiling, a number of factors warrant consideration, including:

1. The *nature of the decision* for which the profile is used. In other words, under what circumstances is it appropriate to use computer profiling? In answering this question, two distinctions may prove helpful. The first is the government purpose in using profiling-e. g., detection of fraud, waste, and abuse; detection of violent criminals; and detection of discrimination. It may be appropriate to use computer profiling for all of these purposes and for any other purposes. Alternatively, the dangers of categorizing people and the speculative nature of profiles may outweigh their general use, but not their use for specific purposes.

   The second distinction is whether only one individual, or one group or class of individuals, is subject to the computer profile. A profile may provide the key by which a database of many individuals is searched. One individual may also be selectively compared to a profile. Because an individual may be affected differently

"Federal Bureau of Investigation, **"Evaluation of the Psychological Profiling Program,"** December 1981.

"Conley, *"Mendenhall* and Reid, " op. cit., p. 839.

under the two circumstances, different standards could be considered for its use.

2. The *nature and source of the data* used. To be consistent with equal protection law, one could argue that computer profiles should not include criteria traditionally considered discriminatory, e.g., race, religion, national origin, or sex. It may also be necessary to eliminate or restrict the use of attributes that may substitute for the overtly discriminatory criteria. Additionally, it may be necessary to restrict the use of results of sophisticated invasive or intrusive psychological or physiological tests, e.g., genetic testing, in profiles.

   In setting standards for the use of data, it may also be helpful to consider the source of the data in determining its relevance for a profile. For example, it may not be appropriate for IRS profiles to include information not provided by the taxpayer or not directly relevant to financial matters.

3. The *rights of individuals,* with respect to both decisions based on profiles and being the subject of profiling, regardless of use. Should individuals be informed that their records are being searched on the basis of a profile or that they are being compared to a profile? If they do not want to be subject to profiling, what are their remedies? If an individual is accorded different treatment because of the way he or she compares to a profile, what rights does he or she have and how can they be implemented?

4. The *accuracy* of the profile. Given that profiles themselves are prone to errors, some testing may be necessary prior to the use of a profile. Independent validation and testing of any software program used for profiling may be necessary to determine bias and accuracy. If profiles are to be used, guidelines may need to be developed for validation and testing. It may be necessary that this testing be done by a group (or groups) other than the one that developed the profile. Although it maybe difficult to get an exact accounting of costs and benefits, some outlining of the significant costs and benefits that are expected could also be done.

   With respect to the drug courier profile, William Conley has suggested that testing should be done in two steps. First, establishing the percentage of those previously arrested who displayed a particular characteristic. Second, determining what percentage of all airplane passengers exhibit the same characteristic.'"

"'Ibid., p. 863.

Chapter 6

# Policy Implications

# Contents


| | Page |
|---|---|
| summary | 99 |
| Introduction | 100 |
| Policy Problems | 104 |
| PolicyActions | 107 |
| Action 1: Maintaining the Status Quo | 107 |
| Action 2: Problem-Specific Actions | 108 |
| Action 3: Institutional Changes | 113 |
| Action 4: Consideration of a National Information Policy | 122 |


## Table


| *Table No.* | **Page** |
|---|---|
| **15. Selected** Institutional Changes for Information Policy I?toposed in the 99th Congress | 123 |

Chapter 6

# Policy Implications

## SUMMARY

All governments collect and use personal information in order to govern. Democratic governments moderate this need with the requirements to be open to the people and accountable to the legislature, as well as to protect the privacy of individuals. In the United States, these needs are recognized in the Constitution and various public laws.

In 1974, Congress passed the Privacy Act to address the tension between the individual's interest in privacy and the government need to know. Since the act was passed, there have been dramatic changes in the scale and scope of technological innovations applied to records and record systems, primarily as a means to detect fraud, waste, and abuse, and to aid in law enforcement investigations. New technological applications—most notably the widespread use of microcomputers, computerized record searches, and computer networking—have multiplied within Federal agencies, and have expanded the opportunities for inappropriate, unauthorized, or illegal access to and use of personal information. Individual rights and remedies, as well as administrative responsibilities, are not clear under current policies. At the same time, there is stronger public concern for privacy and more support for legislative protections than there was in the past.

OTA'S analysis of Federal use of electronic record systems revealed a number of common policy problems. First, new applications of personal information have undermined the goal of the Privacy Act that individuals be able to control information about themselves. Second, there is serious question as to the efficacy of the current institutional arrangements for oversight of Federal compliance with the Privacy Act and related Office of Management and Budget (OMB) guidelines. Third, neither Congress nor the executive branch is providing a forum in which the privacy, management efficiency, and law enforcement implications of Federal electronic record system applications can be fully debated and resolved. Fourth, within the Federal Government, the broader social, economic, and political context of information policy, which includes privacy-related issues, is not being considered.

Overall, OTA has concluded that Federal agency use of new electronic technologies in processing personal information has eroded the protections of the Privacy Act of 1974. Many applications of electronic records being used by Federal agencies, e.g., computer profiling and front-end verification, are not explicitly covered either by the actor subsequent OMB guidelines. Moreover, the use of computerized databases, electronic record searches and matches, and computer networking is leading rapidly to the creation of a de *facto* national database containing personal information on most Americans. And use of the social security number as a de *facto* electronic national identifier facilitates the development of this database. Absent a forum in which the conflicts generated by new applications of information technology can be debated and resolved, agencies have little incentive to consider privacy concerns when deciding to establish or expand the use of personal record systems.

Additionally, OTA'S analysis of electronic record systems and their effect on individual privacy has confirmed once again the complexity of Federal information policy. Its broad social, economic, and political implications need systematic policy study.

OTA identified a range of policy actions for congressional consideration:

1. Congress could do nothing at this time, monitor Federal use of information technology, and leave policymaking to case law and administrative discretion. This

would lead to continued uncertainty regarding individual rights and remedies, as well as agency responsibilities. Additionally, lack of congressional action will, in effect, represent an endorsement of the creation of a de *facto* national database and the use of the social security number as a de facto national identifier.

2. Congress could consider a number of problem--specific actions. For example:
   - establish control over Federal agency use of computer matching, front-end verification, and computer profiling, including agency decisions to use these applications, the process for use and verification of personal information, and the rights of individuals;
   - implement more controls and protections for sensitive categories of personal information, such as medical and insurance;
   - establish controls to protect the privacy, confidentiality, and security of personal information within the microcomputer environment of the Federal Government and provide for appropriate enforcement mechanisms;
   - review agency compliance with existing policy on the quality of data/records containing personal information, and, if necessary, legislate more specific guidelines and controls for accuracy and completeness;
   - review issues concerning use of the social security number as a *de facto* national identifier and, if necessary, restrict its use or legislate a new universal identification number; or
   - review policy with regard to access to the Internal Revenue Service's (IRS) information by Federal and State agencies, and policy with regard to the IRS's access to databases maintained by Federal and State agencies, as well as the private sector. If necessary, legislate a policy that more clearly delineates the circumstances under which such access is permitted.
3. Congress could initiate a number of institutional adjustments, e.g., strengthen the oversight role of OMB, increase the Privacy Act staff in agencies, or improve congressional organization and procedures for consideration of information privacy issues. These institutional adjustments could be made individually or in concert. Additionally or separately, Congress could initiate a major institutional change, such as establishing a Data Protection or Privacy Board or Commission.
4. Congress could provide for systematic study of the broader social, economic, and political context of information policy, of which information privacy is a part.

## INTRODUCTION

All governments collect and use personal information in order to govern. Democratic governments moderate this need with the requirements to be open to the people and accountable to the legislature, as well as to protect the privacy of individuals. Advances in information technology have greatly facilitated the collection and uses of personal information by the Federal Government, but also have made it more difficult to oversee agency information practices and to protect the rights of individuals.

In the 1960s, Congress and the executive branch began the first modern reexamination of the effects of government information collection on individual privacy and agency accountability. This occurred in response to two factors: first, the explosion in information activities necessitated by the Great Society programs; and second, the introduction in Federal agencies of large mainframe computers for information storage and retrieval. This reexamination went on for a number of years, and included, most prominently, the 1966 and 1967

hearings on the reposal to establish a National Data Center. the 1971 Senate Committee on the Judiciary hearings on Federal databanks,[2] the 1973 Department of Health, Education, and Welfare's Advisory Committee on Automated Personal Data Systems,[3] and the 1972 project on databanks sponsored by the Russell Sage Foundation and the National Academy of Sciences.[4]

The reexamination of government information collection, computers, and privacy culminated in the 1974 joint hearings of the Senate Committee on Government Operations, Ad Hoc Subcommittee on Privacy-and Information Systems and the Senate Committee on the Judiciary, Subcommittee on Constitutional Rights; [5]-and hearings of the House Committee on Government Operations.[6] These hearings coincided with Watergate and its revelation of how those in power could use and abuse personal information, especially that held by the IRS and the Federal Bureau of Investigation, for their own personal advantage. The result of these hearings was the enactment of the Privacy Act of 1974, which established rights and remedies for individuals who are the subjects of agency recordkeeping and specified requirements that Federal agencies were to meet in handling personal information. In addition, OMB was assigned responsibility for overseeing agency implementation of the act.

Technology.—At the time the Privacy Act was debated and enacted, there were technological limitations on how agencies could use individual records. The vast majority of Federal record systems were manual. Computers were used only to store and retrieve, not manipulate or exchange, information. It was theoretically possible to match personal information from different files, to manually verify information provided on government application forms, and to prepare a profile of a subset of individuals of interest to an agency. However, the number of records involved made such applications impractical.

In the 12 years since enactment of the Privacy Act, at least two generations of information technology have become available to Federal agencies. Advances in computer and data communication technology enable agencies to collect, use, store, exchange, and manipulate individual records, as well as entire record systems, in electronic form. Specifically:

- Microcomputers were not used at all by Federal agencies in the 1970s. Agencies responding to the OTA survey reported a few thousand microcomputers in 1980, with a dramatic increase to over 100,000 in 1985.
- Computer matching was not used by Federal agencies until 1976, and from 1980 to 1984 there was almost a threefold increase in the number of computer matches. Computer matching has become routine in a number of programs, especially eligibility benefit programs.
- Use of computer-assisted front-end verification, especially with on-line computer searches, has intensified in the 1980s, particularly following the requirements of the 1984 Deficit Reduction Act.
- The widespread use of computerized data-

[2]U.S. Congress, House Committee on Government Operations, Special Subcommittee on Invasion of Privacy, *The Computer and Invasion of Privacy,* hearings, 89th Cong., 2d sess,, July 26, 27, and 28, 1966 (Washington, DC: U.S. Government Printing Office, 1966); U.S. Congress, Senate Committee on the Judiciary, Subcommittee on Administrative Practice and Procedure, *Invasions of Privacy (Government Agencies),* hearings, 89th Cong., 2d sess., part 5, Mar. 23-30 and June 7-9, 14, and 16, 1966 (Washington, DC: U.S. Government Printing Office, 1967); and *Computer Privacy Hearings,* 90th Cong., 1st sess., Mar. 14-15, 1967 (Washington, DC: U.S. Government Printing Office, 1967).

[3]U.S. Congress, Senate Committee on the Judiciary, Subcommittee on Constitutional Rights, *Federal Data Banks, Computers and the Bill of Rights,* hearings, 92d Cong., 1st sess., Feb. 24-25 and Mar. 2,3,4,9, 10, 11, 15, and 17, 1971, part 1 (Washington, DC: U.S. Government Printing Office, 197 1).

[4]U. S. Department of Health, Education, and Welfare, Secretary's Advisory Committee on Automated Personal Data Systems, *Records, Computers and the Rights of Citizens* (Washington, DC: U.S. Government Printing Office, 1973).

[5]Alan F. Westin and Michael A. Baker, *Databanks in a Free Society (New* York: Quadrangle/New York Times Book Co., 19'72).

[6]U.S. Congress, Senate Committee on Government Operations, Ad Hoc Subcommittee on Privacy and Information Systems, and Committee on the <Judiciary, Subcommittee on Constitutional Rights, *Privac.y—The Collection, Use and Computerization of Personal Data,* joint hearings, 93d Cong., 2d sess., June 18-20, 1974 (Washington, DC: U.S. Government Printing Office, 1974).

[7]U.S. Congress, House Committee on Government Operations, *Privacy Act of 1974* (Report 93-1416), 93d Cong., 2d sess. (Washington, DC: U.S. Government Printing Office, 1974).

bases, electronic record searches and matches, and computer networking is leading rapidly to the creation of a de *facto* national database containing personal information on most Americans. And use of the social security number as a de *facto* electronic national identifier facilitates the development of this database.

- In the 1970s, manual profiling was used by a few agencies, especially for law enforcement purposes. In the 1980s, computers can be used to generate profiles, and software programs can search databases on the basis of these profiles. The use of computer profiling is expanding beyond law enforcement per se to include various management programs, such as those designed to detect fraud, waste, and abuse.

These technological advances have opened up many new possibilities for improving the efficiency of government recordkeeping; the detection and prevention of fraud, waste, and abuse; and law enforcement investigations. At the same time, the opportunities for inappropriate, unauthorized, or illegal access to and use of personal information have expanded. Because of this expanded access to and use of personal information in decisions about individuals, the completeness, accuracy, and relevance of information becomes even more important. Additionally, it is nearly impossible for individuals to learn about, let alone seek redress for, misuse of their records. Even within agencies, it is often not known what applications of personal information are being used. Nor do OMB or relevant congressional committees know whether personal information is being used in conformity with the Privacy Act.

Information Technology and Fair Information Principles.–The core of the Privacy Act of 1974 is the code of fair information principles. Twelve years later, it is important to review these principles in light of current information technology applications and administrative practices. Although there are a number of iterations of the code of fair information principles, the model for the Privacy Act was the one developed by the Department of Health, Education, and Welfare's Advisory Committee on Automated Personal Data Systems, and hence will serve as the basis for the analysis here.

The first principle is that there must be no personal data recordkeeping system whose very existence is secret. Ensuring that all record systems containing personally-identifiable information are cataloged for the public record depends on each agency carefully monitoring its record systems. In an age of electronic record systems, it is difficult for an agency to keep an accurate catalog of all record systems, both because of the number of systems and because of the continual electronic changes and manipulations. Additionally, the multiplication of personal data systems makes it difficult for an individual to be aware of all the systems whose existence is public.

There are two types of record systems whose status under the Privacy Act is unclear. The first is a personal information system maintained on a microcomputer. Privacy Act officers are unsure of their responsibilities in this area and are looking for either legislative or OMB clarification.[7] The question is whether records maintained on microcomputers are analogous to 'desk notes, which are not covered by the Privacy Act, or whether they are of a different character because they can be retrieved by others and easily disseminated.

The second type of record system whose status is unclear is one that is developed as a result of electronic record searches-primarily computer matches, computer profiles, or computer screens. All electronic record searches generate a new file of those who appear in both systems or who meet the criteria of a profile or screen. Agencies argue that the Privacy Act notice procedures would not apply to these because they are only temporary systems that are destroyed in the process of verification,

[7]Panel on "Privacy Problems Relating to Computer Security, Seventh Annual Symposium on the Freedom of Information and Privacy Acts, sponsored by the Office of Personnel Management Government Executive Institute, Washington, DC, August 1985.

and, therefore, are not record systems under the Privacy Act.

The second principle of fair information practice is that there must be a way for an individual to find out what information about him or her is in a record and how it is used. Technology makes the first requirement of this principle even more important for individuals because more information is being collected from third parties as a result of computerization and on-line searches. While technology could offer individuals more ways to learn what is in their records, OTA found that no agencies have yet offered individuals computer access to their personal information.

Technology has also affected the requirement that there must be a way for an individual to find out how personal information is used. With computerization, the matching of records, searching of files based on profiles, and verifying of information with numerous other record systems have become routine for many record systems. The fact that the uses of information in government databases are increasing does not necessarily mean that individuals will not find out about such uses; however, OTA'S research indicates that agencies have generally not informed individuals, at least not in a direct fashion.

The third principle, that there must be a way for an individual to prevent information about him or her that was obtained for one purpose from being used or made available for another purpose without his or her consent, is affected most dramatically by new applications of technology. The principle includes not just knowledge of the uses of information, but also a means to prevent uses. Given the scale of government recordkeeping and the number of administrative uses of information, it appears to be extremely difficult for an individual to take action.

In computer matching, front-end verification, and computer profiling, information that was collected for one purpose, such as personnel or tax, is being used for another purpose, e.g., detection of fraud, registration for selective service, or payment of child support. In some cases, this principle has been overriden by legislation that has authorized the exchanges. In these instances, the legislative history reveals little explicit consideration of the effect on the fair information principles of the Privacy Act. In the majority of cases, these new uses of information have not been authorized by legislation, but instead have been justified under the routine use exemption of the disclosure pr~visions in the Privacy Act. This exemption has been used for such a large number of information exchanges and for so many types that it now appears to mean that all uses of Federal records are permitted except those that are expressly prohibited.

The fourth principle of fair information practice is that there must be a way for an individual to correct or amend a record of identifiable information about him or her. This principle has become even more important in an age of electronic recordkeeping because more information is collected from parties other than the individual and because information is added to files at indeterminate periods. The increased exchanges and uses of information by Federal agencies make it more difficult to determine what information is maintained and how it is used; therefore it is harder for an individual to corrector amend records. On the other hand, in an age of electronic recordkeeping, it is possible that corrections to individual files could be negotiated via a home computer or agency computer, and agreed upon changes made directly into the system. Based on OTA'S research, it appears that no agency is using computers and telecommunications to provide new ways for an individual to amend records.

The fifth principle is that any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuse of the data. It is from this principle that the maxim that information must be accurate, timely, relevant, and complete has been taken [Public Law, 93-579, Sec. 3(e)(5)]. With electronic record systems, data are collected, manipulated, and exchanged much more quickly than in paper systems. The speed of exchanges

and large number of users make it more difficult to determine who is responsible for data reliability and use. Once again, the technology offers at least a partial solution in that audit trails can be built into systems. In addition, systems can be programmed to automatically purge records or separate data elements after a specified period of time. OTA found that agencies were not, on the whole, making use of the technology to ensure record quality, and were conducting few reviews of record quality.

Public Opinion.— In general, Americans do not believe that there are adequate safeguards for protecting the privacy of information about people.' The percentage of the public believing that personal information about them is being kept in files not known to them has increased from 44 percent in 1974 to 67 percent in 1983. Most Americans, from two-thirds to three-fourths, believe that agencies that release information they gather to other agencies or individuals are seriously invading personal privacy. Yet, a significant percentage of the public believes that public and private organizations do share personal information. Most Americans, 84 percent, believe that master files of personal information could be compiled "fairly easily, " and 78 percent would regard this as a violation of their privacy.

There is increasing public support for additional government action to protect privacy. In 1978, two-thirds of the public responded that laws could go a long way to help preserve privacy. Sixty-two percent thought it was very important that there bean independent agency to handle complaints about violations of personal privacy by organizations. In 1982, over 80 percent of the public supported the major principles of the code of fair information principles. In 1983, large majorities of the public supported the enactment of new Federal laws to deal with information abuse, including laws that would require that any information from a computer that might be damaging to people or organizations must be double-checked thoroughly before being used, and laws that would regulate what kind of information about an individual could be combined with other information about the same individual.

[1]For a more complete discussion of public opinion and privacy, see ch. 2.

## POLICY PROBLEMS

OTA's analysis of Federal agency use of electronic record systems, specifically for computer matching, computer-assisted front-end verification, and computer profiling, revealed a number of common policy problems.

First, new applications of personal information have undermined the goal of the Privacy Act that individuals be able to control information about themselves. As a general principle, the Privacy Act prohibits the use of information for a purpose other than that for which it was collected without the consent of the individual. New computer and telecommunication applications for processing personal information facilitate the use of information for secondary purposes, e.g., use of Federal employee personnel information for locating student loan defaulters, or use of Federal tax information for evaluation of a Medicaid claim.

The expanded use and exchange of personal information have also made it more difficult for individuals to access and amend information about themselves, as provided for in the Privacy Act. In effect, the Privacy Act gave the individual a great deal of responsibility for ensuring that personal information was not misused or incorrect. Technological advances have increased the disparity between this responsibility and the ability of the individual to monitor Federal agency practices. For example, individuals may not be aware that information about them is being used in a computer match or computer profile, unless they monitor the ***Federal*** Register for notices of such uses or unless questions about their personal information arise as a result of the application. In computer-assisted front-end verification, individuals may be notified on an application form that information they provide

will be verified from outside sources, but are unlikely to be told which sources will be contacted.

Additionally, new computer and telecommunication capabilities enable agencies to exchange and manipulate not only discrete records, but entire record systems. At the time the Privacy Act was debated, this capability did not exist. The individual rights and remedies of the act are based on the assumption that agencies were using discrete records. Exchanges and manipulations of entire record systems make it more difficult for an individual to be aware of uses of his or her record, as those uses are generally not of immediate interest to the individual.

Second, there is serious question as to the efficacy of the current institutional arrangements for oversight of Federal agency compliance with the Privacy Act and related OMB guidelines. Under the Privacy Act, Federal agencies are required to comply with certain standards and procedures in handling personal information—e.g., that the collection, maintenance, use, or dissemination of any record of identifiable personal information should be for a necessary and lawful purpose; that the information should be current, relevant, and accurate; and that adequate safeguards should be taken to prevent misuse of information.

OMB is assigned responsibility for oversight of agency implementation of the Privacy Act. Prior studies by the Privacy Protection Study Commission (1977), U.S. General Accounting Office (1978), and the House Committee on Government Operations (1975 and 1983) have all found significant deficiencies in OMB'S oversight of Privacy Act implementation. For example, under the Privacy Act, information collected for one purpose should not be used for another purpose without the permission of the individual; however, a major exemption to this requirement is if the information is for a "routine use'—one that is compatible with the purpose for which it was collected. Neither Congress nor OMB has offered guidance on what is an appropriate routine use; hence this has become a catch-all exemption permitting a variety of Federal agency information exchanges.

More specifically, OTA found that OMB is not effectively monitoring such basic areas as the quality of Privacy Act records; the protection of Privacy Act records in systems currently or potentially accessible by microcomputers; the cost-effectiveness of computer matching and other record applications; and the level of agency resources devoted to implementation of the Privacy Act. OTA also found that neither OMB nor any other agency or office in the Federal Government is, on a regular basis, collecting or maintaining information on Privacy Act implementation. Given the almost total lack of information on Federal agency personal information activities, OTA conducted its own one-time survey of major Federal agencies and found that:

- the quality (completeness and accuracy) of most Privacy Act record systems is unknown even to the agencies themselves, few (about 13 percent) of the record systems are audited for record quality, and the limited evidence available suggests that quality varies widely;
- even though the Federal inventory of microcomputers has increased from a few thousand in 1980 to over 100,000 in 1985, few agencies (about 8 percent) have revised privacy guidelines with respect to microcomputers;
- few agencies reported doing cost-benefit analyses either before (3 out of 37) or after (4 out of 37) computer matches; authoritative, credible evidence of the cost-effectiveness of computer matching is still lacking; and
- in most Federal agencies the number of staff assigned to Privacy Act implementation is limited; of 100 agency components responding to this question, 33 reported less than 1 person per agency assigned to privacy and 34 reported 1 person.

Additionally, OTA found that there is little or no government-wide information on or OMB oversight of: 1) the scope and magnitude of computer matching, computerized front-end verification, and computer profiling activities; z) the quality and appropriateness of the per-

sonal information that is being used in these applications; and 3) the results and cost-effectiveness of these applications.

Third, neither Congress nor the executive branch is providing a forum in which the privacy, management efficiency, and law enforcement implications of Federal electronic record system applications can be fully debated and resolved. The efficiency of government programs and investigations is improved by more complete and accurate information about individuals. The societal interest in protecting individual privacy is benefited by standards and protections for the use of personal information. Public policy needs to recognize and address the tension between these two interests.

Since 1974, the primary policy attention with respect to Federal agency administration has shifted away from privacy-related concerns. Interests in management, efficiency, and budget have dominated the executive and legislative agenda in the late 1970s and early 1980s. Congress has authorized information exchanges among agencies in a number of laws, e.g., the Debt Collection Act of 1982 and the Deficit Reduction Act of 1984. In these instances, congressional debates included only minimal consideration of the privacy implications of these exchanges.

A number of executive bodies have been established to make recommendations for improving the management of the Federal Government, e.g., the President's Council on Integrity and Efficiency, the President Council on Management Improvement, and the Grace Commission. All have endorsed the increased use of applications such as computer matching, front-end verification, and computer profiling in order to detect fraud, waste, and abuse in government programs. However, these bodies have given little explicit consideration to privacy interests. Some executive guidelines remind agencies to consider privacy interests in implementing new programs, but these are not followed up to ensure agency compliance.

In general, decisions to use applications such as computer matching, front-end verification, and computer profiling are being made by program officials as part of their effort to detect fraud, waste, and abuse. Given the emphasis being placed on Federal management and efficiency, agencies have little incentive to consider privacy concerns when deciding to establish or expand the use of personal record systems. As a result, ethical decisions about the appropriateness of using certain categories of personal information, such as financial, health, or lifestyle, are often made without the knowledge of or oversight by appropriate agency officials (e.g., Privacy Act officers or inspectors general), OMB, Congress, or the affected individuals.

Fourth, within the Federal Government, the broader social, economic, and political context of information policy, which includes privacy-related issues, is not being considered. The complexity of Federal Government relations—within executive agencies, between the executive and legislature, between the Federal Government and State governments, and between the Federal Government and the private sector—is mirrored in interconnecting webs of information exchanges. This complexity and interconnectedness is reflected in a myriad of laws and regulations, most of which have been enacted in a piecemeal fashion without consideration of other information policies.

Some of these policies may be perceived as being somewhat inconsistent with others, e.g., the privacy of personal information and public access to government information. Some laws and regulations may only partially address a problem, e.g., Federal privacy legislation does not include policy for the private sector or for the flow of information across national borders. In other instances, issues that are inherently related and interdependent, such as privacy and security, are debated and legislated in separate forums with only passing attention to their relationship.

Additionally, the Federal Government information systems, as well as its information policy, are dependent on technological and economic developments. Federal funding for research and development and Federal financial

and market regulations will have significant implications for these developments. Yet, uner the present policymaking system, there is no assurance that these implications will be considered. Likewise, the international information policy environment, as well as international technological and economic developments, affects domestic information policy; yet these factors are not systematically considered in the existing policy arenas.

## POLICY ACTIONS

Overall, OTA has concluded that Federal agency use of new information technologies in processing personal information has eroded the protections of the 1974 Privacy Act. Many of the electronic record applications being used by Federal agencies, e.g., computer profiling and front-end verification, are not explicitly covered by either the act or subsequent OMB guidelines. Even where applications are covered by statute or executive guidelines, there is little oversight to ensure agency compliance. More importantly, neither Congress nor the executive branch is providing a forum in which the conflicts-between privacy interests and competing interests, such as management efficiency and law enforcement—generated by new applications of information technology can be debated and resolved. Absent such a forum, agencies have little incentive to consider privacy concerns when deciding to establish or expand the use of personal record systems.

OTA has identified a range of policy actions for congressional consideration, including maintaining the status quo, problem-specific actions, institutional changes, and consideration of a national information policy. These policy actions are discussed below.

### Action 1: Maintaining the Status Quo

Congress could do nothing at this time, monitor Federal use of information technology, and leave policymaking to case law and administrative discretion.

The implication of maintaining the status quo is that the present policy problems and confusion will continue. It is likely that the policy emphasis on management efficiency; on detection and prevention of fraud, waste, and abuse; and on effective law enforcement will continue to take precedence over privacy-related concerns. This emphasis will most likely result in an increased use of current applications of information technology in Federal agencies for record searches such as computer matching, computer-assisted front-end verification, and computer profiling. In addition, it is likely that new applications will be developed.

Without congressional action, individuals will continue to be unaware of the majority of uses and disclosures of personal information by Federal agencies because there will be no notice other than that which appears in the *Federal Register*. If an individual has a question about agency practices and procedures, it is difficult for him or her to find the appropriate person to contact in a Federal agency. If an individual wishes to challenge an agency use of personal information, he or she will not have clearly defined or effective recourse because of the problems with the damage remedies of the Privacy Act.

Additionally, absent congressional action, there will be a lack of information available to Congress and the American people, as well as within agencies, concerning the scale and scope of technological applications applied to records and record systems in Federal agencies. This will make it even more difficult for Congress to be aware of current or proposed agency practices in order to exercise effective oversight. Moreover, the lack of information will aggravate the existing difficulties in monitoring the quality, e.g., accuracy and completeness, of personal information that is used and exchanged by Federal agencies.

If Congress does not address the problems resulting from Federal agency applications of

new information technology in processing personal information, then Federal agency staff will be left to interpret the meaning of the fair information principles in an electronic age. This would undermine a primary goal of the Privacy Act because it would increase the discretion of administrative agencies in handling personal information. Additionally, this would not meet the need expressed by some agency staff for more specific guidance from either OMB or Congress.

Most importantly, lack of congressional action will, in effect, represent an endorsement of the creation of a *de facto* national database containing personal information on most Americans, and an endorsement of the use of the social security number as a *de facto* national identifier. Current legislation, such as the Deficit Reduction Act of 1984, has accelerated what had been the gradual development of a national database because of the increased data searches and creation of computerized databases authorized by this legislation. Individual authorizations such as these have been largely unnoticed by the public. However, without consideration of the overall societal and political implications, these authorizations taken together could lead to personal information practices that most of the American public would find unacceptable.

### Action 2: Problem-Specific Actions

**Congress could also consider a number of** problem-specific actions, dealing with computerized record searches, specific categories of information (social security number, tax information, and medical or other sensitive information), microcomputers, and record/data quality.

There are a number of procedural and substantive changes that Congress could legislate. In fashioning such changes, it would be easiest for Congress to deal with specific problem areas. Each of these will be discussed below. These changes are not mutually exclusive. Indeed, to provide the most comprehensive protection for personal information, it maybe necessary to legislate in all of these areas.

*A. Establish control over Federal agency use of computer matching, front-end verification, and computer profiling, including agency decisions to use these applications, the process for use and verification of information, and the rights of individuals.*

In order to do this Congress could, in effect, require congressional approval for every record search involving personal information. This would entail amending the "routine use" provision of the Privacy Act to eliminate matching and other record searches from this exemption. As a result, agencies would need to obtain congressional authorization each time they wished to search records containing personal information. Although this approach would enable Congress to monitor record searches and to limit agency discretion in deciding to search records, it may involve a prohibitive time investment for Congress or be a de *facto* prohibition on such searches. Federal agencies likely would be opposed to such an approval process, as they might perceive it as unnecessary interference in internal agency affairs.

Alternatively, Congress could authorize general record searches, but establish explicit standards and procedures. This would require amending the Privacy Act in at least three possible ways:

1. Amend the "routine use" provision to allow record searches under specific circumstances and with specific types of records. In this way, Congress would establish the criteria under which matches and other searches could be done, and the types of records that could not be used in these searches (e.g., medical files or tax and security clearance records).
2. Specify the due process protections (e.g., notice, right to a hearing, right to confidentiality of results, or right to counsel) for persons whose records are to be searched, and the time when due process protections come into effect (e.g., before the match, after the match but before verification, and after verification).
3. Require a cost/benefit analysis before and after every match.

Although establishing standards and procedures may be more workable and realistic than requiring congressional approval for every record search, it does not provide any mechanism to ensure that agencies have complied with the general standards. Based on the experience of agency record searches to date, it appears that oversight and enforcement are essential.

In addition to any of the above amendments, or as an alternative, Congress could require agencies to adopt a 5-year plan for detecting fraud, waste, and abuse. In this way, agency proposals to search record systems would be placed within a context. Agencies would then need to justify record searches as a technique according to criteria such as purpose, cost, and alternatives considered. Such plans could be subject to congressional approval. Again, this would likely be ineffective without critical review, oversight, and enforcement.

Also, in addition to the above, Congress could amend the Privacy Act to require the social security number on all Federal, State, and local government forms. This might improve the accuracy of information used in matching, and might reduce the costs of verifying hits. However, it seems unwise to adopt this action without considering the problems with using the social security number as an authenticator and identifier, and the problem of endorsing a national identifier.

*B. Implement more controls and protections for sensitive categories of personal information, such as medical and insurance.*

Statutes provide specific protection in many areas where personal information is collected and used—e.g., banks, credit agencies, educational institutions, and criminal history repositories. Based on *United States v. Miller, 425* U.S. 435 (1976), if there is no specific statutory basis for an individual's right with respect to a particular type of personal information held by another party, the individual may not be able to assert a claim about how that information is used.

The Privacy Protection Study Commission (PPSC) analyzed the privacy implications of the recordkeeping practices in a number of areas, including insurance, employment, and medical care, and made recommendations for policy. Very few of these recommendations resulted in legislation, although some were embodied in voluntary codes by organizations such as insurance companies and employers.

Medical information is still an area in which an individual's interests are not protected by statute. In 1977, PPSC recommended that "now is the proper time to establish privacy protection safeguards for medical records."[g] The Commission was led to this conclusion by the changing conceptions of the medical record and increased automation. Although many bills to protect medical information have been introduced, none has yet passed. The Federal Government collects, maintains, and discloses a great deal of sensitive medical information. Agencies involved include, for example, the Department of Health and Human Services (HHS), the Occupational Safety and Health Administration, the Environmental Protection Agency, and the Veterans Administration. Agencies collect medical information for purposes such as delivering services, providing cost reimbursements, and conducting research. Legislation could address these and other needs.

Legislating for a specific type of information or specific organizational entity on a piecemeal basis is not without its problems. OTA'S research indicates that it is difficult to isolate collection of personal information in this way. Instead, the information infrastructure is complex and constantly overlapping. Needs, interests, and programs converge at many points.

*C. Establish controls to protect the privacy, confidentiality, and security of personal information within the microcomputer environment of the Federal Government and provide for appropriate enforcement mechanisms.*

[g]Privacy Protection Study Commission, *Personal Privacy in an Information Society* (Washington, DC: U.S. Government Printing Office, 1977), p. 290.

Agencies appear to be dealing with microcomputer policy on an ad hoc basis. This approach results in variation in the protection afforded personal information by Federal agencies. In establishing policy for the use of microcomputers withing Federal agencies, it is necessary to address the management, data integrity, security, confidentiality, and privacy aspects.

OTA'S companion report, Management, Security, and Congressional Oversight,[10] analyzes in detail the management, data integrity, and security aspects of information systems policy, including for microcomputers. Briefly, there are four general kinds of measures to protect information systems. First are administrative security measures, such as requiring that employees change passwords every few months; removing the passwords of terminated employees quickly; providing security training programs; storing copies of critical data off-site; developing criteria for sensitivity of data; and providing visible upper management support for security. Second are physical security measures, such as locking up diskettes and/or the room in which microcomputers are located, and key locks for microcomputers, especially those with hard disk drives.

There are also numerous technical measures to assure security, including audit programs that log activity on computer systems; security control systems that allow different layers of access for different sensitivities of data; encrypting data when they are stored or transmitted, or using an encryption code to authenticate electronic transactions; techniques for user identification; and shielding that prevents eavesdroppers from picking up and deciphering the signals given off by electronic equipment.

Lastly, there are legal remedies to discourage information system abuse, generally known as computer crime, and to prosecute perpetrators. Because computerized information is intangible, its abuses do not fit neatly into existing legal categories, such as fraud, theft, embezzlement, and trespass. This makes computer crime a different kind of criminal act needing special legislative attention. Concern with protecting the privacy of personal information is related to computer crime in that such crimes may involve unauthorized access to personal information.[11]

However, there are important aspects of privacy protection that are not addressed by the security measures discussed above. The Privacy Act establishes individual rights of knowledge, access, and correction, and places requirements on agencies to maintain records in a certain fashion, and to use and disclose records for certain purposes. These procedural and substantive protections are limited to records containing personal information that are "contained in a system of records. " A system of records is defined as "a group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual" [See.3(a)(5)]. It is unclear which records maintained on microcomputers come under this definition. Once this has been determined, it will be necessary to provide a means of monitoring these records to ensure that the individual rights of knowledge, access, and correction are provided.

*D. Review agency compliance with existing policy on the quality of data/records containing personal information, and, if necessary, legislate more specific guidelines and controls for accuracy and completeness.*

A central aspect of Federal records policy, as embodied in the Privacy Act and Paperwork Reduction Act, is that records should be complete and accurate. Through the provisions in these acts, Congress has recognized the importance of record quality both to management efficiency and to the protection of individual

[10] U.S. Congress, Office of Technology Assessment, *Federal Government Information Technology: Management, Security, and Congressional Oversight*, OTA-CIT-297 (Washington, DC: U.S. Government Printing Office, February 1986).

[11] For further discussion of computer crime issues and policy options, see ibid., especially ch. 5. Also see U.S. Congress, Office of Technology Assessment, *Federal Government Information Technology: Electronic Surveillance and Civil Liberties*, OTA-CIT-293 (Washington, DC: U.S. Government Printing Office, October 1985).

rights. Agency decisions based on inaccurate or incomplete information can lead to wasteful or even harmful results. Many Federal record systems are now computerized. While computerized systems offer the potential to improve record quality, undetected or uncorrected errors can be disseminated more quickly and widely—with potentially serious consequences.

Based on available evidence, including the results of the OTA survey, OTA has concluded that most Federal agencies do not maintain statistics on record quality or conduct audits of record quality. While many agencies have policies and procedures intended to ensure record quality, they do not measure actual quality levels (by comparing record contents with primary information sources), and thus do not have a complete basis for knowing whether or not problems exist.

OTA asked Federal agencies (major components of all 13 cabinet departments plus 20 independent agencies) for the results of any record quality audits conducted on Privacy Act record systems and for record quality statistics on all computerized record systems maintained for law enforcement, investigative, and/or intelligence purposes. Only one agency provided any statistics, and very few of the other agencies indicated that such statistics may exist.

With respect to audits of the quality of Privacy Act records, only 16 of 127 (or 13 percent) agencies responding indicated that they conduct such audits; none provided the results.[12] Only one agency provided record quality statistics (for three systems under its jurisdiction) for law enforcement, investigative, and intelligence record systems. No statistics were provided for any of the other 82 systems reported.[13] Subsequent to the data request, the FBI was asked for and did provide the results of partial audits of the National Crime Information Center (see app. A for further discussion).

Should Congress wish to address the record quality problem directly, the appropriate congressional committees could conduct oversight on Federal electronic record quality, and, if satisfied that a significant problem exists, consider amendments to the Privacy Act and/or Paperwork Reduction Act to provide stronger guidance to the executive branch on this topic. Congress could also ask for General Accounting Office and/or Inspector General audits of record quality of selected Federal agency record systems in order to provide additional independent confirmation of Federal record quality. Finally, Congress could direct one or more of the central agencies responsible for information technology management (Office of Information and Regulatory Affairs, OMB; National Bureau of Standards; or Office of Information Resources Management, General Services Administration) to develop audit packages and techniques that could be used by Federal agencies to measure and monitor record quality.

*E. Review issues concerning use of the social security number as a* de facto *national identifier and, if necessary, restrict its use or legislate a new universal identification number.*

The Privacy Act makes it "unlawful for any Federal, State, or local government agency to deny to any individual any right, benefit, or privilege provided by law because of such individual's refusal to disclose his social security account number' unless disclosure is required by law or unless the system of records was in existence prior to January 1, 1975 (the grandfather clause). Although the General Accounting Office, HHS, and numerous task forces all agree that "the social security number is, at best, an imperfect identifier and authenticator, 14 its use has expanded since 1974. The social security number is an impor-

[12] A total of 142 agencies were surveyed; 5 did not respond at all, and 10 others responded that the question was not applicable or that the information was not available, for a net total response of 127 agencies.

[13] Again, 142 agencies were surveyed; a total of 85 computerized law enforcement, investigative, or intelligence record systems were identified. Agencies responded as follows: record quality statistics maintained (3 systems); no record quality statistics (63 systems); no response (17 systems); not applicable or information not available (1 system); and classified (1 system).

[14] Privacy Protection Study Commission, *Personal Privacy in an Information Society*, op. cit., p. 609.

tant component in the matching process, and HHS has developed a software program, which will detect erroneous social security numbers, that is to be used in conjunction with a match.

Contrary to the stated intent of the Privacy Act, the trend in the use of the social security number appears to be towards its adoption as a de *facto* national identifier. Federal, State, and local agencies, as well as the private sector, have increased their requests, as well as their requirements, for disclosing one's social security number (or Taxpayer Identification Number). In hearings on the Privacy Act, concern with the possibility of the adoption of a universal identifier was voiced. Much of the concern focused on the record searches that a universal identifier would allow. Congress considered setting severe restrictions on the use of the social security number, but was dissuaded by testimony that the costs and implications of such restrictions were unknown. Since enactment of the Privacy Act, Congress has passed numerous laws authorizing Federal agencies to collect the social security number and requiring State agencies to collect it in administering Federal programs.

PPSC was asked to study restrictions on the use of the social security number and to make recommendations. The major finding of PPSC was "that restrictions on the collection and use of the social security number to inhibit exchange beyond those already contained in the law would be costly and cumbersome in the short run, ineffectual in the long run, and would also distract public attention from the need to formulate general policies on record exchanges. "[15] PPSC went on to recommend that "the Federal Government not consider taking any action that would foster the development of a standard, universal label for individuals, or a central population register, until such time as significant steps have been taken to implement safeguards and policies regarding permissible uses and disclosures of records about individuals. " Such a comprehensive study has not yet been conducted.

[15]Ibid., p. 614.

If the social security number is being used as a de *facto* standard universal identifier in the United States, both the benefits and hazards of having a national identifier need to be evaluated. The General Accounting Office, PPSC, congressional committees, and the Social Security Administration itself have all discussed parts of these issues. Congress could make a comprehensive review of issues concerning use of the social security number as a *de facto* national identifier and establish a clear policy for the electronic age, with appropriate enforcement mechanisms.

*F. Review policy with regard to access to the Internal Revenue Service's information by Federal and State agencies, and policy with regard to the Internal Revenue Service's access to databases maintained by Federal and State agencies, as well as the private sector. If necessary, legislate a policy that more clearly delineates the circumstances under which such access is permitted.*

IRS files are valuable sources of information for many record searches because of the variety of information on file (e.g., address, earned income, unearned income, social security number, number of dependents) and because the information is relatively up to date. As a general rule, returns and return information are to remain confidential, as provided for in Section 6103 of the Tax Reform Act of 1976. Under this section, information may be disclosed for tax and audit purposes and proceedings, and for use in criminal investigations if certain procedural safeguards are met.

Additionally, Section 6103(1) allows for the disclosure of tax return information for purposes other than tax administration. The list has grown considerably since 1976, and includes: the Social Security Administration and Railroad Retirement Board (Public Law 94-455, 1976); Federal loan agencies regarding tax delinquent accounts (Public Law 97-365, 1982); the Department of Treasury for use in personnel or claimant representative matters (Public Law 98-369, 1984); Federal, State, and local child support enforcement agencies (Public

Law 94-455, 1976); and Federal, State, and local agencies administering certain programs under the Social Security Act or Food Stamp Act of 1977 (Public Law 98-369, 1984). Section 2651 of the Deficit Reduction Act also amends Section 6103(1) of the Tax Reform Act and allows information from W-2 forms and unearned income reported on 1099 forms to be divulged to any Federal, State, or local agency administering one of the following programs: Aid to Families With Dependent Children; medical assistance; supplemental security income; unemployment compensation; food stamps; State-administered supplementary payments; and any benefit provided under a State plan approved under Titles I, X, XIV, or XVI of the Social Security Act. Section 6103(m) of the Tax Reform Act also provides for disclosure of taxpayer identity information to a number of agencies, including the National Institute for Occupational Safety and Health and the Secretary of Education.

In all instances, Sections 6103(1) and (m) specify procedures that other parties are to follow in order to gain access to IRS information. Moreover, Federal, State, and local employees outside of IRS who handle IRS information are subject to the same criminal liabilities as IRS employees for misuse or disclosure of the information. The IRS also puts out a publication, *Tax Information Security Guidelines for Federal, State and Local Agencies* (Publication 1075; Rev. 7-83), that describes the procedures agencies must follow to ensure adequate protection against unauthorized disclosure.

Pressure to extend the list of agencies that can access IRS information has intensified with interest in record searches to detect fraud, waste, and abuse; to register men for the Selective Service; and for any program that requires a current address for an individual. The IRS's position is that its goal is to maintain a voluntary tax system and that the public's perception that tax information should remain confidential is important to maintaining a voluntary system. Thus, the IRS is, in principle, opposed to disclosing tax information.

Technological advances, however, may make voluntary disclosure of tax information by the affected individual less important and thus reduce the IRS's concern for confidentiality. For example, the IRS is moving towards a system where information provided by the individual would be phased out of the tax return process and replaced with information disclosed directly to the IRS by the sources, e.g., employers, banks, credit agencies, investment companies, mortgage companies, etc. If this becomes the case, the IRS will not need to be concerned with maintaining a voluntary tax system or with protecting the confidentiality of tax information.

Congress may wish to legislate a general, but enforceable, policy regarding the circumstances under which tax information may be disclosed and procedures for such disclosure. The ad hoc process of amending Sections 6103(1) and (m) when the political situation allows, as reflected in the long list of congressionally authorized disclosures, may not be the most effective approach to maintaining the confidentiality of tax information.

Congress may also wish to examine IRS access to other agency and private sector databases, and legislate a more clearly delineated policy for such access. This becomes more important as the IRS relies increasingly on sources of information other than the taxpayer. Additionally, IRS access to other databases may result in inaccurate or irrelevant information being included in IRS records.

### Action 3: Institutional Changes

**Congress could initiate a number of institutional adjustments, e.g., strengthening the oversight role of OMB, increasing the Privacy Act staff in agencies, or improving congressional organization and procedures for consideration of information privacy issues. These institutional adjustments could be made individually or in concert. Additionally or separately, Congress could initiate** a major institutional change, such as establishing **a Data Protection or Privacy Board or Commission.**

Strengthening the institutional framework for information privacy policy could achieve

three purposes, either singly or in combination. First, an institution could play the role of an ombudsman in assisting individuals to resolve individual or class grievances with a Federal agency about personal information practices. Second, it could oversee Federal agency compliance with the Privacy Act and related OMB guidelines. Third, an institution could provide a forum in which proposals to alter personal information practices and systems (e.g., to conduct a computer match or to set up a new computerized database) could be discussed in the context of the implications for personal privacy and consistency with the principles of the Privacy Act.

In the increasingly complex, technological, and bureaucratic environment of the late 1980s, the fair information principles of the Privacy Act are even more important, but the Privacy Act scheme of enforcement and oversight appears to be increasingly anachronistic. For instance, it may not be realistic to ask individuals to control information about themselves in view of the cost and time burdens entailed. Also, the number of organizations that retain personal information is large, and the intricacies of their uses and disclosures of information are such that it appears almost impossible for most individuals to monitor how information is being used.

Moreover, the implicit assumption that each individual has a discrete interest in protecting his or her privacy, and that there is no larger societal interest, can be challenged. Many researchers and practitioners believe that there is also a social interest in maintaining certain boundaries of personal information collection and use. As discussed in chapter 2, the results of public opinion polls implicitly support this view.

There are three weaknesses in a personal information policy that provides for enforcement primarily through individual grievances and requires little direct oversight of agency practices.

First, *the policy relies on individuals to protect their interests.* The Privacy Act requires that individuals be aware of their rights, understand the potential threats posed by Federal agency collection and use of personal information, and be willing to invest the time and money necessary to protect their interests. These requirements place a burden on the individual. Every time one comes in contact with an agency seeking personal information, he or she would need to question the purposes for which information is sought and the necessity of each piece of information.

To ensure that information is not misused, the individual would need to follow up to make sure that no new information was added to the file, and that the uses and disclosures of information were in keeping with the agency's stated purposes. If individuals find that files contain inaccurate or irrelevant information, or that information was used for improper purposes, then they would need to know what legal remedies are available and take action against the Federal agency. Such a procedure means that individuals would need to be conscious of their rights at every stage of the information-handling process. Most people are so accustomed to disclosing information that they rarely think through all of the possible consequences. As Michael Baker suggests:

> What we can expect in the way of self-protective action on the part of individual citizens is severely limited by the fact that record-keeping practices are of relatively low visibility to and salience for the individual.[16]

The second weakness in the enforcement scheme of the Privacy Act is that *it only provides remedies once misuses have been identified.* If an individual has the right to correct inaccurate information or make a case for deleting or amending information in his or her record, the right only "rights" a wrong already committed against the individual. It does not protect the record from further errors or misuses, nor does it prevent similar wrongs from being committed against other individuals. It provides no preventive protection unless the granting of new rights to individuals can be

[16] Michael A. Baker, "Record Privacy as a Marginal Problem: The Limits of Consciousness and Concern, *Columbia Human Rights Law Review*, vol. 4, 1972, p. 89,

viewed as a means of deterring agencies from engaging in questionable information practices. But the time and money necessary to take action against a Federal agency make it unlikely that many individuals will take advantage of these rights. Thus, the deterrent effect of such rights on agency information practices is likely to be minimal.

The third weakness is that the personal information policy *is not sensitive to the existing imbalance of power between the individual and Federal agencies.* Under the Privacy Act, the interests of individuals are placed in opposition to the needs of the government for information. In most situations, the individual is dependent on the government for employment, credit, insurance, or some other benefit or service. Therefore, the individual is not likely to "afford" the risk of questioning an agency's information practices. Some view this as the most significant policy weakness and argue that:

> [the] enormous imbalance of power between the isolated individual and the great data collection organizations is perfectly obvious: under these conditions, it is a pure illusion to speak of "control." Indeed, the fact of insisting exclusively on means of individual control can in fact be an alibi on the part of a public power wishing to avoid the new problems brought about by the development of enormous personal data files, seeking refuge in an illusory exaltation of the powers of the individual, who will thus find himself alone to run a game in which he can only be the loser.[17]

Strengthening an existing institution or establishing a new one would bring more visibility to the issue of personal information collection and use; provide a central place for individuals to bring complaints and for agencies to seek advice; and enable Congress, the agencies, and the public to get more complete, accurate, and timely information on agencies' practices. The institution could also place limitations on the initial collection of information; review, and possibly approve, proposals to link record systems; and set standards for and oversee data quality in all systems.

A number of institutional changes available to Congress are discussed below:

*A. Strengthen the role of the Office of Management and Budget in the enforcement and oversight of the Privacy Act.*

Under the Privacy Act, OMB is responsible for providing guidelines and regulations, providing assistance to the agencies, overseeing the procedural mechanisms, and preparing the President Annual Report on Implementation of the Privacy Act. OMB has issued a number of guidelines, most significantly with respect to computer matching and the Debt Collection Act. However, in at least one instance–the guidelines released under the Debt Collection Act–OMB issued its guidelines without time for public comment.[18] In another instance, OMB did not issue guidelines as promised in a judicial action.[19] In addition, OMB has not yet acted on a requirement in the Paperwork Reduction Act to "submit to the President and the Congress legislative proposals to remove inconsistencies in laws and practices involving privacy, confidentiality, and disclosure of information."[20]

From the enactment of the Privacy Act in 1974 until 1980, OMB provided assistance through a separate office with a few staff members within its Information Policy Division. At this time, as the Privacy Protection Study Commission found, "neither OMB nor any of the other agencies with guidance responsibilities have subsequently played an aggressive role in making sure that the agencies are equipped to comply with the act and are, in fact, doing so."[21]

[17] S. Rodota, "Privacy and Data Surveillance: Growing Public Concern," *OECD Information Studies #10–Policy Issues in Data Protection and Privacy* (Paris: OECD, 1976), pp. 139-140.

[18] See comments of Christopher DeMuth, Administrator, Office of Information and Regulatory Affairs (OIRA), Office of Management and Budget (OMB), and Robert Bedell, Deputy Administrator, 01 RA, OMB, in *Oversight of the Privacy Act*, House Committee on Government Operations, Subcommittee on Government Information, Justice, and Agriculture, 1983, pp. 123-124.

[19] See *Bruce v. United States*, 621 F.2d 915 (8th Cir. 1980).

[20] See House Report No. 98-455.

[21] Privacy Protection Study Commission, *Personal Privacy in an Information Society*, op. cit., p. 21.

The Paperwork Reduction Act created the Office of Information and Regulatory Affairs with desk officers to oversee the implementation of information-related policies (including the Privacy Act) within an agency. Although this style of oversight does not necessarily mean that Privacy Act concerns receive less attention, it appears that this has been the practice. Testimony from Christopher DeMuth of OMB at the 1983 hearings on oversight of the Privacy Act[22] indicates (and interviews with OMB confirm) that the desk officers spend little time on Privacy Act matters.

OMB has focused its attention on the review of systems of records, as provided for in the Privacy Act. The act does not offer OMB any other specific guidance and OMB has not taken the initiative—e.g., by reviewing agencies' mechanisms for providing individual access and correction or for maintaining the accuracy of records.

OMB prepares the President's Annual Report on Implementation of the Privacy Act. Annual reports for the years 1975 through 1978 were well-documented studies of agency practices under the Privacy Act, and included descriptions of Federal personal information systems and agency administration, as well as data on use of the access and correction provisions of the act. The information contained in 1980 and 1981 reports was not as complete and focused mainly on systems that agencies designated as exempt from the Privacy Act. In 1982 debates on the Congressional Reports Elimination Act, OMB recommended that the Privacy Act Annual Report be eliminated. Congress rejected this suggestion.[23] The 1982-83 Annual Report on Implementation of the Privacy Act was not delivered to Congress until December 1985. This report synthesized Federal agencies' administration of the act over the past 10 years, and suggested areas for congressional action.

The goal of the Paperwork Reduction Act of 1980 was to reduce paperwork and improve information technology management. The act was designed to coordinate information-related activities of Federal agencies—specifically, automated data processing, telecommunications, office automation, information systems development, data and records management, and, possibly, printing and libraries. The act also acknowledged the importance of information as a resource and made a commitment to the management concept of information resources management, popularly known as IRM.[24]

Concern with protecting the confidentiality and security of personal information and providing individuals access to that information is part of the IRM concept. However, privacy has not been centrally integrated into IRM as presently implemented in Federal agencies. In part, this can be attributed to the fact that the Privacy Act and Paperwork Reduction Act are distinct pieces of legislation, with different public, congressional, and agency constituencies.

Another reason for the lack of integration and coordination is that OMB was somewhat slow to take a lead role in formulating IRM policy. In December 1985, OMB issued Circular A-130, "Management of Federal Information Resources, " which sets basic guidelines for the collection, processing, and dissemination of information by Federal agencies, and for the management of information systems and technology. The circular also revised and coordinated existing directives on privacy and computer security. Although the circular succeeds in centralizing information policy in one document, it does not contain any significant changes from previous congressional and OMB policies, and, in general, does not provide detailed guidance to agencies.

In terms of strengthening OMB'S role, Congress could to do three things. First, it could amend the Privacy Act, giving OMB the authority to issue regulations-not merely guide lines-and the authority to enforce them. Such

[22] "Oversight of the Privacy Act, ibid., pp. 123-124.
[23] See House Report No. 98 455.

[24] For a more complete discussion of IRM, see U.S. Congress, Office of Technology Assessment, *Federal Government Information Technology: Management, Security, and Congressional Oversight*, op. cit.

additional authority would put OMB in the role of policing agency personal information practices. The advantage of strengthening OMB authority is that it could be achieved with minor institutional change and minimal overhead. The major disadvantages are that agencies may resist this expansion in OMB'S authority, and that continued congressional oversight would be required to ensure that OMB was fulfilling its new responsibilities. Given OMB'S prior attention to this area and its other responsibilities, some of which may conflict with data protection/privacy, it may be questionable whether OMB could improve its oversight role even with additional authority.

Second, Congress could enhance OMB'S institutional base for dealing with the Privacy Act. This could be done by setting up a separate office with responsibility for data protection/privacy. In order for this office to be effective, Congress would need to ensure that adequate staff and budget are provided. Alternatively, Congress could increase the staff in the Office of Information and Regulatory Affairs and provide a separate staff person per agency who would be responsible for the privacy issues of that agency. Although the institutional framework is in place to achieve these changes quickly, the problem of ensuring OMB commitment to ensure compliance with the Privacy Act remains.

Third, Congress could upgrade the Office of Information and Regulatory Affairs, possibly by taking it out of OMB and establishing it as anew Office of Federal Management, as provided for in S. 2230, the "Federal Management and Reorganization and Cost Control Act of 1986. " This would have the advantage of removing the conflict that exists within OMB between budgetary constraints and management interests. However, it would be important to ensure that privacy be accorded equal importance with other management interests. The principal disadvantage of such a change is that it would be controversial, as it represents a major institutional reorganization.

*B. Increase the size, stature, and authority of privacy staff in agencies.*

Under the Privacy Act, each agency has designated an official who is responsible for Privacy Act matters. In many agencies, this official is also responsible for the Freedom of Information Act. In most agencies, there is little or no staff support for Privacy Act matters. The OTA survey revealed that 67 percent of agency components responding (67 out of 100) reported one FTE (full-time equivalent) staff person or less assigned to Privacy Act matters. Only 7 percent of agency components (7 out of 100) responding reported having 10 or more FTEs assigned to Privacy Act matters. Five of these components were located in the Department of Justice and included the Drug Enforcement Agency, Immigration and Naturalization Service, Federal Bureau of Investigation, and Criminal Division. The other agencies with more than 10 FTEs assigned to the Privacy Act were the Social Security Administration and the Office of the Secretary in the Department of Commerce.

Congress could amend the Privacy Act to require agencies to provide a certain level of professional and staff support for Privacy Act matters. Such an amendment could provide for adequate training conducted by both related agency staff (e.g., Freedom of Information Act officers, General Counsel staff, staff in the Inspector General's Office, and IRM personnel) and external groups (e.g., OPM'S Government Executive Institute and the American Society of Access Professionals).

In amending the Privacy Act, Congress could also specify the responsibilities and authorities of the Privacy Act officers, e.g., to serve as liaison between individuals and agencies in resolution of problems or grievances; to approve, or be consulted about, new record applications; and to maintain information on agency practices. If Privacy Act staff are to be effective in protecting privacy interests from within the agency, their authority must be stated in the legislation; otherwise it is possible that upper management will thwart their efforts.

The primary problem with this action is that enforcement and oversight responsibilities are

left within the agencies. Therefore, in addition to statutory changes, intensified congressional oversight of each agency may be required.

C. *Improve congressional organization and procedures for consideration of information privacy issues.*

At present, Congress does not have a mechanism for coordinated oversight of public laws and bills having privacy implications. Indeed, almost every committee has responsibility for some aspect of the personal information practices of Federal agencies. For example, issues related to the Privacy Act and privacy in general are of interest to the House Committees on Government Operations and on the Judiciary and the Senate Committees on Governmental Affairs and on the Judiciary; privacy issues involving school records are sent to the House Committee on Education and Labor and the Senate Committee on Labor and Human Resources; issues involving privacy of credit records are sent to the Committees on Banking in each House; privacy issues arising under the Freedom of Information Act are considered by the House Committee on Government Operations and the Senate Committee on the Judiciary; issues involving cable subscriber privacy are sent to the House Committee on Energy and Commerce and the Senate Committee on Commerce, Science, and Transportation; in the House, medical records confidentiality has been discussed by the Committees on Government Operations, Energy and Commerce, and Ways and Means, as well as by the Senate Committee on Energy and Commerce; and tax record confidentiality comes under the purview of the House Committee on Ways and Means and the Senate Committee on Finance.

Because of the fragmentation of the committee system and the primacy of substantive concerns in individual committees, privacy interests are often not given thorough consideration. Moreover, it is difficult for interest groups who define their roles as protecting privacy to keep track of relevant legislation and to monitor all pertinent congressional hearings.

If Committees with crosscutting privacy jurisdiction were established in both Houses, either as permanent committees, new subcommittees, or select committees, and all bills having privacy implications were referred jointly or sequentially to those committees, privacy issues could be debated and resolved in a more deliberate and focused manner. It is theoretically easy for Congress to make a change of this nature, but politically it is likely to be difficult as reform efforts of the past decade indicate.[25]

An easier alternative would be for Congress to retain the existing committee structure, but provide for better monitoring of bills having information privacy implications, and joint referral of such bills to committees with privacy jurisdiction.

D. *Establish a Privacy or Data Protection Board.*[26]

The proposal to establish an entity to oversee the personal information practices of Federal agencies is not new. The original Privacy Act that passed the Senate provided for the establishment of a Privacy Protection Commission with powers to:

- monitor and inspect Federal systems and databanks containing information about individuals;
- compile and publish an annual U.S. Information Directory so that citizens and Members of Congress will have an accurate source of up-to-date information about the personal data-handling practices of Federal agencies and the rights, if any, of citizens to challenge the contents of Federal databanks;
- develop model guidelines for implementation of the Privacy Act and assist agencies and industries in the voluntary development of fair information practices;
- investigate and hold hearings on violations of the act, and recommend corrective action to the agencies, Congress, the

[25] See, for instance, Steven S. Smith and Christopher J. Deering, *Committees in Congress* (Washington, DC: Congressional Quarterly Inc., 1984).

[26] The term "data protection" is a more precise term for the issues that arise from the collection and use of personal information. It is the term adopted by many European countries. However, privacy is the more easily understood term in the United States.

President, the General Accounting Office, and the Office of Management and Budget;
- investigate and hold hearings on proposals by Federal agencies to create new personal information systems or modify existing systems for the purpose of assisting the agencies, Congress, and the President in their effort to assure that the values of privacy, confidentiality, and due process are adequately safeguarded; and
- make a study of the state of the law governing privacy-invading practices in private databanks and in State, local, and multistate data systems.[27]

The Senate's Privacy Protection Commission was to be composed of five persons who were expert in law, social science, computer technology, civil liberties, business, and State and local government.

A professional staff would have been provided for the commission. The Senate Committee on Government Operations concluded:

> There is an urgent need for a permanent staff of experts within the Federal Government to inform Congress and the public of the data-handling practices of major governmental and private personal information systems.[28]

The Senate considered three alternative institutional placements for the commission—in the U.S. General Accounting Office, in OMB, or in an independent commission–and concluded that an independent commission was, on balance, the best solution. The House did not approve the establishment of a Privacy Protection Commission as it did not see the need for outside oversight of agency practices. As a compromise, both Houses approved the establishment of a Privacy Protection Study Commission to study further the personal information systems and practices of government and private organizations, to make recommendations as to whether the principles of the Privacy Act should be extended beyond Federal agencies, and to make other recommendations as the commission deemed necessary.

The Privacy Protection Study Commission released its report in 1977, and also recommended the establishment of a Federal Privacy Board or some other independent entity with responsibilities similar to those approved by the Senate in 1974. These include the responsibility to: monitor and evaluate the implementation of statutes and regulations; participate in agency proceedings; issue interpretative rules; continue to research, study, and investigate areas of privacy concern; and advise the President, Congress, government agencies, and the States on privacy implications of proposed statutes or regulations.[29]

Since 1977, there have been a number of bills creating a Privacy Commission or Data Protection Board, including H.R. 1721, the "Data Protection Act of 1985," introduced in the 99th Congress. None has received serious congressional attention.

Many Western European countries and Canada have established boards or commissions with responsibilities for the protection of personal information. Because these may serve as a model for such an agency in the United States, descriptions of several countries are found in appendix F.

The advantages and disadvantages of a new privacy authority in the United States would be determined by the design of the agency and the powers with which it is vested. In this respect, a number of policy choices are important.

1. **Whether such an agency should have regulatory authority or advisory authority.** The data protection agencies in Sweden and France are regulatory agencies, with power to determine the personal information systems that government and private sector agencies can create, the information that can be retained, and the parties that can have access to the informa-

[27] U.S. Congress, Senate Committee on Government Operations, "Protecting Individual Privacy in Federal Gathering, Use and Disclosure of Information," Report No. 93-1183, 93d Cong., 2d sess., 1974, pp. 23-24.

[28] Ibid, p. 24.

[29] Privacy Protection Study Commission, *Personal Privacy in an Information Society*, op cit, p. 37.

tion. The data protection agencies in West Germany and Canada have advisory authority and act as ombudsmen, serving as intermediaries between individuals and agencies, rendering advisory opinions, and lobbying for protection of personal information across a range of policy areas.

In the United States, it is likely that a regulatory agency would be resisted by existing Federal agencies because it would be perceived as having too much control over internal and day-today agency affairs. A regulatory agency may also become unwieldy and obstructive. An advisory/ombudsman authority may be more compatible with American philosophical and institutional traditions. It also has a precedent at the State level, e.g., New York. Based on the European and Canadian experience, the advisory/ombudsman model appears to have provided effective oversight of agency practices. Another possibility would be to establish an agency that is primarily advisory, but give it some veto power overparticular agency practices.

2. The institutional placement of such an authority. The major choice here is whether to make it independent of the executive branch and responsible to the legislature, or to make it part of the executive branch. If it were to be a new office or domestic council within the Executive Office of the President, it could have a great deal of visibility and stature if the President decided to make protection of personal information a priority. However, the stature of such a new office might well change with changes in administrations. Also, it could be politicized, especially if budgetary interests were given higher priority or if senior White House officials were interested in using personal information for political purposes—e.g., getting access to IRS information on political opponents or political activists.

Another possibility would be to have the authority established as a bureau within an existing executive department. The advantages of this option would be that it probably would be easier to establish and the overhead costs would be minimal. But, there are significant disadvantages. Inevitably, the power of the new authority would be dependent in part on that of the department, and its character shaped by the department. Additionally, any staff or line department, e.g., the Office of Personnel Management or the Department of Health and Human Services, collects and uses personal information, and, therefore, may have a conflict of interest in the resolution of information collection and disclosure policies.

A third possibility would be to have the authority established as an independent agency of the executive branch. While the agency head presumably would still report to the President, top officials could be made subject to Senate confirmation and even given statutory terms of office. These measures would help protect the authority from inappropriate political pressures and strengthen its institutional independence, as discussed later.

Alternatively, the new authority could report to Congress, either directly or through a special joint committee. The advantage of this approach is that an independent, nonoperating authority would have no stake in the existing personal information exchanges of executive agencies and might be more objective in resolving future conflicts. Moreover, an authority reporting to the legislature would increase the means Congress has to directly oversee the activities of executive agencies. Theoretically, a data protection/privacy authority reporting to the legislature, rather than to the executive, would have independence from the day-to-day operating constraints, as well as the political constraints, of executive agencies.

The disadvantage of having the new agency report to the legislature is that it might be subject to competing political interests, especially if there were different partisan majorities in the two Houses or if the executive and legislature were controlled by different parties. But, even if the authority became politicized, the political maneuverings might be more visible to Congress and the public if the authority re-

ported to Congress than if it were part of the executive. This would seem to ensure a certain degree of accountability.

In determining the placement and powers of a new agency, it will be important to consider the Supreme Court's recent decision in *Immigration and Naturalization Service v. Chadha, 103 S.* Ct. 2764 (1983), as well as its pending decision on the constitutionality of the Gramm-Rudman deficit reduction proposal.

3. The scope of issues for which the agency would be responsible. Some have proposed that such an authority should be responsible for all privacy issues, e.g., information privacy, surveillance, autonomy/life choices, and "chilling effects" on first amendment rights. If this were the case, information privacy would receive less sustained attention. Also, the size of the authority would, by necessity, be larger. Others have proposed that such an authority should be responsible for all information technology issues, for example, research and development, security, technology transfer, and industrial competitiveness. The same difficulties of focus and size would also apply to an authority with these responsibilities.

The uniqueness and complexity of problems presented by personal information collection and use argue that if an authority is established, it should be solely responsible for personal information issues—not all privacy issues or all information technology issues. However, the growing interrelationships between Federal and State personal information systems, and between public and private systems, argue that, to be effective, an authority would need the power to address all aspects of personal information exchanges. Limiting its purview to Federal agencies could narrow its effectiveness.

4. Outlining the agency's specific authority and responsibilities. Generally, such an agency is given some authority to require other agencies to register, or list, their personal information systems, with details on the information held, the sources of information, the uses, the period for which information is retained, and the exchanges and disclosures of information. This process of registration is supposed to ensure that there are no secret systems of personal records. Alternatively, the agency could be given the authority not only to register the systems, but also to approve their existence through a process of licensing. Additional responsibilities that could be considered include:

- some role in settling disputes over issues, such as access and accuracy, that develop between individuals and agencies;
- some role in formally making recommendations on proposed systems or new legislation that have implications for personal information;
- establishing guidelines and standards for specific personal information issues, e.g., what is an acceptable "routine use" or what is "accurate, timely, and complete' information;
- compilation and submission of an annual report on present and anticipated trends in personal information practices; and
- monitoring technological developments and assessing their implications for personal information practices.

5. Staffing a new authority. Two models exist for the organization of government agencies. One is to follow the independent regulatory agency model and have multiple commissioners appointed for staggered terms. Another is to have a single head for a fixed term of office. The advantage of the former is that partisan influences are minimized, while the advantage of the latter is that responsibility is clear and visible.

An additional issue is the size of the staff. The maximum number of staff reported for Western European and Canadian counterparts of such an authority is 30. Given the greater population and complexity of Federal/State relations, a somewhat larger staff may be necessary in the United States; however, there are advantages to keeping it small and well organized.

Congress might anticipate two arguments against a proposal to establish a new entity. The first is that it might entail another layer of bureaucracy. However, the purpose of a new

entity is to serve as a check on Federal agencies, not to become a part of the bureaucratic establishment. Additionally, the agency could be kept small and its style and organization nonbureaucratic. The second anticipated argument against a new entity would be that the costs associated with privacy protection may increase. This argument may be somewhat spe cious because, at present, there is no accounting of the costs associated with privacy protection. In calculating these costs, one would need to include agency administrative costs (e.g., the time of Privacy Act Officers, General Counsels, Inspectors General, program managers, and administrative judges); judicial costs (e.g., Department of Justice time and court costs); and the time of individuals.

### Action 4: Consideration of a National Information Policy

Congress could provide for systematic study of the broader social, economic, and political context of information policy, of which privacy is a part.

OTA'S analysis of Federal agency electronic record systems and individual privacy has confirmed once again the complexity and interrelationships of Federal information policy. The broader social, economic, and political context of information policy is in need of systematic policy study. This discussion could occur in existing executive offices or congressional committees. Alternatively, or in concert, a national study commission could also provide a forum for discussion and examination of a national information policy.

A 1981 OTA study[30] found that there were numerous laws and regulations, some overlapping and some potentially or actually conflicting, that directly and indirectly affect the operators and users of information systems, the consumers of information services, and the subjects of personal information databanks. OTA concluded that continuation of this situation could inhibit many socially desirable applications of information systems or could create even more intractable policy problems in the future. At that time, OTA found that few policymakers were interested in a uniform Federal information policy that would encompass the problems that could arise from the many possible uses of data systems.

OTA identified the need for consideration of an "information policy" that would address the confusing array of laws and regulations—and their strengths, overlaps, contradictions, and deficiencies—within some overall policy framework. This need has not yet been met.

There have been numerous proposals for the establishment of new organizations to study information-related policy problems (see table 15 for a summary) .3' Over the last several years, a growing number of Members of Congress and industry leaders, while not necessarily endorsing specific policies, have expressed concern about the lack of coordinated focus on national information policy issues and the absence of adequate institutional mechanisms. For example:

- Representative George Brown (with Representatives Don Fuqua and Doug Walgren) has introduced legislation to establish an Institute for Information Policy and Research and a Special Assistant to the President for Information Technology and Science Information;[32]
- Senator Sam Nunn (with Senator Frank Lautenberg) has introduced legislation to establish an Information Age Commission;[33]
- Representative Cardiss Collins has introduced legislation to establish a new Office of Telecommunications Policy in the Executive Office of the President;[34]

[30]U.S. Congress, Office of Technology Assessment, ***Computer-Based National*** *Information Systems,* **OTA-CIT-146** (Washington, DC: U.S. Government Printing Office, September 1981)

[31]For a more complete discussion of information policy, see U.S. Congress, Office of Technology Assessment, " **Institutional** Options For Addressing Information Policy Issues: A Preliminary Framework for Analyzing the Choices, staff memorandum prepared by the Communication and Information Technologies Program, Nov. 29, 1983.

[32]H.R. 744, "Information Science and Technology Act of 1985", 99th **Cong., 1st sess.**

[33]S. 786, "Information **Age** Commission Act of 1985", 99th Cong., **1st** sess.

[34]H.R. 642, "Telecommunications Policy Coordination Act of 1985", 99th Cong., **1st** sess.

## Table 15.—Selected Institutional Changes for Information Policy Proposed in the 99th Congress

| Proposed institutional change | Problem or issues to which change directed | Organizational form | Functions | Membership | Location | Resources and authority | Duration |
|---|---|---|---|---|---|---|---|
| Information Age Commission, S 786 (Nunn and Lauten berg) | Impact of computer and communication systems on society | Commission | Research, policy formulation and information dissemination | 23 members–6 from Congress 6 from executive branch and 11 from private sector | Independent–reporting to President and Congress | Hold hearings, negotiate and enter into contracts, and secure cooperation and assistance from other executive agencies | 2 years |
| Off Ice of Federal Management, S 2230 (Roth) | Management of the Federal Government | Off Ice | Strengthen overall Federal management and, in particular financial management and Information resources management, and reduce the costs of administration | From OMB Will be transferred to the Off Ice of Federal Procurement Policy, Off Ice of information and Regulatory Affairs, and other appropriate functions of OMB A new Off Ice of Financial Systems Will also be established | Executive Off Ice of the President | Provide central policy direction and leadership in general management maintain oversight of managerial systems and processes, advise President and Congress | Permanent |
| Off Ice of Critical Trends Analysis, S 1031 (Gore) H R 2690 (Gingrich) | Identification and analysis of critical trends and alternate futures | Off Ice | Publish reports, advise President establish advisory commission, and promote public discussion | – | Within Executive Off Ice of the President | Legislation requires President to submit report to Congress and requires Joint Economic Committee to prepare report on similar topic | On-going–prepare report every 4 years beginning in 1990 |
| Institute for information Policy and Research, H R 744 (Brown) | Broad range of information policy concerns | Institute | Research policy formulation Information dissemination and promotion of innovation | 15 member board representing government industry and commerce, and academic and professional organizations | An Independent structure within the executive branch Director to coordinate with other agencies | – | 10 years unless extended by Congress |
| National Technology Foundation, H R 745 (Brown) | High-technology small business, technology transfers, and international activties | Foundation | Analyze and make grants and contracts for development of high-technology small businesses, conduct technology assessments, promote technology transfer and international cooperation | Transfers to the Foundation the following agencies Patent and Trademark Off Ice, NBS, NTIS, parts of NSF, and other specified agency sections | Independent governmental agency | Award grants, loans, and other assistance, conduct assessments, promote technology transfers | Authorizes appropriations for FY 1986 through FY 1988 |
| Data Protection Board, H R 1721 (English) | Personal records held by Federal agencies | Board | Develop guidelines, provide assistance, publish guides Investigate compliance, issue advisory opinions intervene in agency proceedings | Three members appointed by President with advice and consent of Senate for 7-year terms | Independent executive agency | Conduct inspections, hold hearings issue subpoenas | Permanent |
| Department of International Trade and Industry, H R 1928 (Watkins) | International trade and Industry | Department | Full range including advising, negotiating, and regulating | Travel and Tourism Administration Patent and Trademark Off Ice, NBS, NTIS, Off Ice of Telecommunications and Information, Off Ice of Small Business Trade Assistance, and Off Ice of Competitive Analysis | Independent department | Legislation requires President under certain conditions to submit statement on impact on International economic competitiveness of significant domestic product and Service Industries | Permanent |
| Advanced Technology Foundation H R 2374 (LaFalce) | Technology in business, commerce, and Industry | Foundation | Promote the commercial application and diffusion of advanced technology within Industrial sectors | | Within executive branch | Create referral service coordinate programs provide grants, and develop Information management system | Authorizes appropriations through FY 1989 |

SOURCE Off Ice of Technology Assessment

- Representative Glenn English has introduced legislation to establish a Data Protection Board;[35]
- The American Federation of Information Processing Societies has formed a panel of experts on National Information Issues, and the Association of Data Processing Service Organizations has proposed a Temporary National Information Committee.[36]

Most of these proposals view information policy within the context of an information society, i.e., one in which the creation, use, and communication of information will play a central role. There are numerous, interconnected issues arising from the following factors:

- the need to have a greater understanding of the changing role of information and its impact on society;
- the economic and political transition to an information society;
- the effect that the information revolution may have on the governmental process;
- dealing with information as an economic resource, a commodity, and a property;
- the importance of managing information and in trying to assure its accuracy and high quality, especially insofar as it is generated, used, and disseminated by the Federal Government;
- the need to protect individual civil liberties and rights to privacy;
- ensuring access to information and equity that may arise when information is treated more and more as a commodity and less and less as a public good; and
- the enhanced ability of information to travel across nation-al boundaries.

In most discussions of information policy, the relative importance of these issues has not been noted. Indeed, numerous Federal agencies have a role in aspects of information policy, but there is no office or agency providing integration across multiple information policy issue areas. Agencies that might provide such integration, such as the National Telecommunications and Information Administration (in the Department of Commerce) and the Office of Science and Technology Policy (in the Executive Office of the President), have not been provided the necessary mandate and resources, nor do they appear, at least at present, to have the desire to carry out such activities.

Proponents of a national information policy argue that it is just as important as national economic or environmental or defense policy, and deserves a clear focus at the highest levels of government. Beyond this, proponents point to the need for a mechanism to encourage high-level identification and understanding of and leadership on issues arising from the transition to an information society-including issues of protecting individual civil liberties and social equity and the development of information as a valuable economic as well as public good.

Opponents in the past have expressed concern about the dangers of centralizing too much authority over information policy in one place, and have favored continuation of a decentralized policy apparatus with coordination provided through interagency and White House working groups. Some of this concern reflects the experience with the old Office of Telecommunications Policy (created in 1970 in the Executive Office of the President and terminated in 1977). OTP was perceived in part as attempting to influence the content of broadcast news. This raised the specter of a high-level government censorship office.

Realistically, it maybe necessary to divide the information problem into more manageable pieces. Because of the urgency of the emerging privacy-related information problems and because there is no inherent group constituency for privacy rights, it may be timely to establish a study commission with responsibility for examination of these interrelated issues.

Two recent proposals for new study commissions in the information policy area include a "National Commission on Communications Security and Privacy" proposed in 1984 by

[35] H.R. 1721, "Data Protection Act of 1985", 99th Cong., 1st sess.
[36] AFIPS, *Washington Report*, July 1985, p. 5.

Representative Dan Glickman, of the House Committee on Science and Technology Subcommittee, and the "Information Age Commission" noted earlier. Any national commission on information policy would most likely be broad in scope and encompass many of the issue areas previously identified. A commission established along the lines of these proposals would have a finite lifetime, modest budget, and broad composition (e.g., with representatives from industry, labor, academia, State/local government, and Federal Government). Establishing a new commission need not be a substitute for other congressional policy actions. Indeed, a commission could be viewed as complementing related activities by Federal agencies and could help to improve public understanding of and focus on current and emerging information policy issues.

# Appendixes

# Appendix A

# Update on Computerized Criminal History Record Systems*

## Introduction

OTA has carried out an extensive prior study of Federal and State criminal history record systems. The preliminary and final results were published in, respectively, *A Preliminary Assessment of the National Crime Information Center and the Computerized Criminal History System*[1] *(1978)* and *Assessment of Alternatives for a National Computerized Criminal History System*[2] *(1982).*

The 1982 study addressed four major areas:

1. the status of criminal history record systems in the United States;
2. the alternatives for a national computerized criminal history (CCH) system;
3. the possible impacts of such a system on the criminal justice process, Federal-State relations, and civil and constitutional rights; and
4. the relevant policy issues that warranted congressional attention to ensure that the beneficial impacts of a national CCH system are maximized and the possible adverse impacts controlled or minimized.

Since 1982, one particular alternative for a national CCH system, known as the Interstate Identification Index (or Triple I), has been tested and generally accepted by the criminal justice community. Triple I is now one of 12 operational files in the National Crime Information Center (NCIC) operated by the Federal Bureau of Investigation (FBI). Triple I is essentially a national electronic index to persons with Federal and/or State criminal history records. The records themselves are maintained in FBI and State record repositories. Triple I replaced the now defunct Computerized Criminal History file on NCIC, and is the largest file on NCIC, as shown in table A-1.

Also since 1982, the extent of computerization in other criminal history record repositories has continued to increase. The FBI's Automated Identification Division System (a CCH record system separate from the NCIC) included 8,740,908 computerized records as of May 1985, compared to about 5.8 million records in October 1981.[3] At the State level 35 States reported at least a partially computerized criminal history record file as of late 1984, compared to 27 States in August 1982.[4] And 39 States reported, as of late 1984, at least a partially automated name index to persons with criminal history records, as compared with 34 States in August 1982.[5] The fully or partially computerized criminal history files of the States account for an estimated 90 percent of all criminal history record activity.[6]

As discussed in chapter 4 and more extensively in the 1982 OTA report, the Triple I concept evolved after a protracted debate, spanning more than a decade, over the appropriate Federal and State roles in a national CCH system.[7] While the

*Outside reviewers for this appendix included Robert R. Belair, Kirkpatrick & Lockhart; Gary R. Cooper, SEARCH Group, Inc.; David F. Nemecek, Federal Bureau of Investigation; and Fred Wynbrandt, California Department of Justice.

[1] U.S. Congress, Office of Technology Assessment, *A Preliminary Assessment of the National Crime Information Center and the Computerized Criminal History System, OTA-1-80* (Washington, DC: U.S. Government Printing Office, December 1978). Also published as U.S. Congress, Senate Committee on the Judiciary, Subcommittee on Administrative Practice and Procedure and Subcommittee on the Constitution, *Preliminary Report by the Office of Technology Assessment on the Federal Bureau of Investigation National Crime Information Center (NCIC) Accompanied by Letters of Comment on the Draft Report,* 95th Cong., 2d sess., December 1978.

[2] U.S. Congress, Office of Technology Assessment, *An Assessment of Alternatives for a National Computerized Criminal History System, OTA-CIT-161* (Washington, DC: U.S. Government Printing Office, October 1982. Prepared at the request of the House and Senate Committees on the Judiciary, this study was one of four components of the OTA "(Assessment of Societal Impacts of National Information Systems. " The other components included a September 1981 OTA report on *Computer-Based National Information Systems: Technology and Public Policy Issues;* a March 1982 background paper on selected *Electronic Funds Transfer Issues: Privacy, Security, and Equity;* and an August 1982 OTA report on *Implications of Electronic Mail and Message Systems for the U.S. Postal Service.*

[3] Based on Federal Bureau of Investigation data.

[4] Aug. 6, 1982 data from an OTA survey cited in U.S. Congress, Office of Technology Assessment, *Computerized Criminal History System,* op. cit., pp. 46-48; late 1984 data from a SEARCH Group, Inc., survey cited in U.S. Department of Justice, Bureau of Justice Statistics, "State Criminal~ Records Repositories, " technical report, October 1985, pp. 2-3, prepared by SEARCH Group, Inc., for a Jan. 9, 1986, conference cosponsored by SEARCH Group and the Bureau of Justice Statistics.

[5] Ibid.

[6] OTA previously concluded that, for fiscal year 1981, the 27 States with on-line CCH files accounted for about 85 percent of all criminal fingerprint cards submitted to State and Federal criminal record repositories—a valid measure of criminal history record activity. See OTA, *Computerized Criminal History System,* op. cit., pp. 46-48 and table 5. As of late 1984, eight other States (Louisiana, Montana, New Hampshire, Arizona, Connecticut, Wyoming, Idaho, and Pennsylvania) had automated at least partially, accounting collectively for an estimated additional 5 percent of criminal record activity. Actually, based on 1984 data, these eight States together held about 6.5 percent of the total number of State criminal history records. See Bureau of Justice Statistics, "Criminal Records Repositories, " op. cit., p. 2.

[7] Also see U.S. Congress, Senate Committee on the Judiciary, Subcommittee on Patents, Copyrights, and Trademarks, *Computerized Criminal History Records,* hearing, 98th Cong., 1st sess., May 12, 1983; and U.S. General Accounting Office, *Observations on the FBI Zntcr-*

**Table A.1 .–Number of Records Included in NCIC, by File, 1979, 1981, 1985**

| | Number of records as of | | |
|---|---|---|---|
| | June 1981 | October 1981 | May 1985 |
| Interstate identification index | — | — | 9,268,232 |
| Computerized criminal history | 1,482,017 | 1,885,457 | — |
| Stolen securities | 1,998,778 | 2,361,971 | 2,072,785 |
| Stolen guns | 1,337,310 | 1,674,814 | 2,052,018 |
| Stolen vehicles | 970,714 | 1,163,771 | 1,170,613 |
| Stolen articles | 1,091,461 | 1,427,535 | 1,053,415 |
| Stolen license plates | 397,706 | 543,173 | 495,225 |
| Wanted persons | 148,644 | 190,159 | 219,123 |
| Missing persons | 21,535 | 24,610 | 38,374 |
| Stolen boats | 17,615 | 22,807 | 24,370 |
| Unidentified persons | — | — | 1,067 |
| Canadian warrants | n.a. | 183 | 249 |
| U.S. secret service protective | — | — | 91 |
| Total | 7,465,780 | 9,294,327 | 16,395,662 |

NOTES: — =file did not exist.
n.a. = data not available.

SOURCE" Federal Bureau of Investigation.

Triple I now appears to be generally accepted by the criminal justice community, OTA reviewed the results of the 1982 study and found that at least three of the key policy issues previously identified have not yet been resolved: 1) noncriminal justice use of criminal history records; 2) the quality (completeness and accuracy) of such records; and 3) policy oversight of the interstate exchange of criminal history information. The status of each is briefly updated below, along with an overview of policy implications.

## Noncriminal Justice Use

Criminal record checks are increasingly used in screening applicants for a wide range of jobs and licenses. In the 1982 study, OTA found that noncriminal justice use of criminal history records was already substantial (about one-half of all record requests received by the FBI's Identification Division and about one-seventh of all record requests received by State repositories).

Since 1982, the trend toward criminal record checks for employment and licensing has further intensified. For example, Congress included a provision in Public Law 98-473 requiring that States establish procedures to provide for nationwide criminal history checks for all operators and employees of child-care facilities.[8] There has also been growing interest in implementing criminal record checks for teachers, youth group leaders, and elder-care providers. The primary motivation for the increased emphasis on criminal record checks has been the intensified attention and concern about child abuse (and, to a lesser extent, abuse of the elderly) and the perceived need to more carefully screen applicants for positions entrusted with the care of persons who are likely to be especially vulnerable.[9] In addition, there has been increased emphasis on criminal history record checks for current or prospective Federal employees, especially those in sensitive or classified positions.[10]

Absent policy action, this increasing level of record check activity is likely to aggravate access, equity, and due process problems resulting from the inconsistent Federal and State laws and regulations on dissemination of criminal history records for noncriminal justice purposes. These problems were identified in the 1982 OTA report and further amplified in two 1984 studies commissioned by the FBI to study the implications of using Triple I for noncriminal justice record checks.

One study, conducted by former FBI agent Raymond J. Young and reflecting a Federal perspective, concluded that:[11]

> The most obvious impact (of III) would be the total lack of availability of criminal history record information from States for many or all Federal non-criminal uses. The inability to acquire criminal history data would affect many vital uses, including matters involving national security. . . . In

*state Identification Index*. Report to the Chairman, Subcommittee on Civil and Constitutional Rights, House Committee on the Judiciary, Oct. 16, 1984.

[8] U.S. Department of Health and Human Services, *Model Child* Care *Standards Act-Guidance to States To* Prevent *Child* Abuse in Day Care *Facilities*, Washington, DC, January 1985, p. 2.

[9] See, for example, Adrian Higgine, "Day Care Worker Checks Getting Mixed Reviews," *Arlington Journal*, Sept. 6, 1985, p. A7; Linda Lantor, "Fairfax Schools To Tighten Employee Screening," *Arlington Journal*, Sept. 10., 1985, p. A4, and Andee Hochman, "Youth Workers Face Additional Screening; Change Follows Spate of Sex Abuse Cases," *The Washington Post*, Sept. 23, 1985, pp. D1-D2.

[10] See, for example, Mike Causey, "FBI Checks Background of 41,000 at HHS," *The Washington Post*, June 21, 1985, pp. A1-A1 1; S. 274, the Anti-Nuclear Terroriem Act of 1985, 99th Cong., let sees., that would require criminal record checks for nuclear powerplant personnel; S. 1203, 99th Cong., 1st sess., that would allow railroad police and private university or college police access to FBI criminal history records; and S. 1347, the Security Clearance Information Act of 1985, 99th Cong., 1st sees., introduced by Senator Sam Nunn (for himself and Senators William Roth, Lawton Chiles, Albert Gore, and Ted Stevens) and enacted by Congress as Title VIII of Public Law 99-169, that gives the Department of Defense, Office of Personnel Management, and Central Intelligence Agency the statutory authority to access Federal and State criminal history information for national security purposes.

[11] Raymond J. Young, *Federal Non-Criminal Justice Use of the Interstate Identification Index*, prepared for the Federal Bureau of Investigation, Dec. 14, 1984, pp. 5-1, 5-2.

many other instances, Federal agencies would receive only limited amounts of data from States which, while providing some criminal history information from some Federal uses, place restrictions on the type of criminal history records furnished.

A second study, carried out by SEARCH Group, Inc.—a consortium representing State perspectives—found that:[12]

> [T]here is great disparity among present State laws and policies regarding noncriminal justice access and use. Laws and policies on dissemination range from those in a few States that essentially do not permit access to any criminal history records for any noncriminal justice purpose to those of a few "open record" States that permit access to all or most of such records for anyone for any purpose. Between these extremes is an almost bewildering variety of statutory approaches, with access permitted in particular States to specified records for specified purposes and subject to specified conditions, including requirements that access be authorized by separate legal authority or approved by a council, board, or other official.

As a consequence of these and other as yet unresolved problems, noncriminal justice use of Triple I is currently prohibited.

## Record Quality

The importance of accurate records has long been recognized in Federal and State laws and regulations. Since 1970, Congress has explicitly expressed its concern about the completeness and accuracy of criminal history records. Section 524(b) of the Crime Control Act of 1973 required the Law Enforcement Assistance Administration to promulgate regulations that, among other things, were to provide safeguards for the completeness and accuracy of criminal history records. Such regulations were issued in 1975 (as Title 28, Code of Federal Regulations, part 20) and applied to the Federal Government and all States whose criminal history record systems were federally funded in whole or in part.

Federal courts have also ruled on record quality issues. For example, in Tarlton v. Saxbe (1974) the U.S. Court of Appeals for the District of Columbia ruled that the FBI had a duty to prevent dissemination of inaccurate arrest and conviction records, and had to take reasonable precautions to prevent inaccuracy and incompleteness. Most States now have statutes or regulations requiring agencies to ensure reasonably complete and accurate criminal history information, including reporting of court dispositions. The number of States with statutes or regulations on record quality increased from 14 in 1974 to 45 in 1979, and to 49 in 1981.[13]

In spite of legislative and judicial mandates to improve record quality, the 1982 OTA study documented significant record quality problems in Federal and State criminal history record systems. The record quality problem that stands out above all others is the lack of information on dispositions. A long series of record quality audits, including OTA'S, have shown that, on the average, one-third to one-half of the dispositions that occurred were missing from State and Federal criminal history records.[14] OTA'S audits also documented that, for the Federal and State files sampled, roughly one-fifth of criminal history records contained erroneous information.[15]

Since the 1982 OTA report, record quality has received heightened attention. For example, SEARCH Group, Inc.—with Department of Justice (Bureau of Justice Statistics) funding—has held conferences and prepared reports on understanding the problem and on possible solutions, and has developed procedures for conducting record quality audits. 16 The FBI Director has assigned record quality improvement a high priority .17 And the FBI, with the support of the NCIC Advisory Policy Board, has established an audit team to check State compliance with NCIC procedures, including those on record completeness and accuracy. However, as yet, the audit of record quality is limited to the NCIC files on wanted persons and stolen vehicles, and does not include the criminal history records on which the NCIC Triple I is based. "

The FBI has solved part of its record quality problem by terminating the NCIC/CCH file. In effect, it was discontinued as part of the decision to

[12] SEARCH Group, Inc., *A Study To Identify Criminal Justice Information Law, Policy and Management Practices Needed To Accommodate Access to and Use of III for Noncriminal Justice Purposes*, prepared for the Federal Bureau of Investigation. Sept. 18,1984. p. 4.

[13] See, U.S. Congress, Office of Technology Assessment, *Computerized Criminal History System*, op. cit., pp. 71-73 and 94-96.

[14] Ibid., pp. 89-96 and 99-102.

[15] Ibid., pp. 89-96.

[16] See SEARCH Group, Inc., *Audit Manual for Criminal History Records Systems*, Sacramento, CA, December 1982; *Audit Documentation Guide: A Model Study Approach*, Sacramento, CA, January 1984; "SEARCH Audit Clinics Take New Approach" and "National Workshop To Examine Date Quality," *Interface*, summer 1984, pp. 19, 31; U.S. Department of Justice, Bureau of Justice Statistics, *Data Quality of Criminal History Records*, prepared by SEARCH Group, Inc., October 1985; and "National Conference on Data Quality and Criminal History Records," Jan. 9-10, 1986, cosponsored by the Bureau of Justice Statistics and SEARCH Group, Inc.

[17] U.S. Department of Justice, Federal Bureau of Investigation, *Minutes of National Crime Information Center Advisory Policy Board*, Washington, DC, Oct. 17-18, 1984, p. 2.

[18] See U.S. Department of Justice, Federal Bureau of Investigation, *National Crime Information Center Control Terminal Audit Manual*, June 4, 1985.

proceed with the Triple I.[19] The FBI has initiated several actions to improve disposition reporting at the Federal level, such as "computer tape exchange with other Federal agencies, automatic generation of disposition follow-up requests, and field recovery teams to review court and agency records, " and reports some improvement.[20]

However, audits and surveys of State criminal history record files conducted since 1982 have generally confirmed the results of the 1982 OTA study and suggest significant, continuing record quality problems. For example, 1984 audit results from one State–Illinois-indicated that about 20 percent of arrest events audited had erroneous information and about 50 percent of arrest events audited were missing dispositions, a majority of which were included in local police records.[21] Also, a 1984 national survey of criminal history record quality conducted by SEARCH Group, Inc., found wide variability in disposition reporting. Many States were unable to provide estimates of disposition reporting. For those that did, the average disposition reporting by law enforcement, prosecution, and local correctional agencies was estimated to be about 50 percent–a finding generally consistent with results of other, prior audits.[22] On the positive side, disposition reporting by State correctional agencies was estimated to be about 95 percent. About two-thirds of the States believed that disposition reporting and overall record accuracy were increasing, although most States did not provide hard numbers or audit results to support this belief. States cited increased automation as a major reason for improvement. Other reasons cited include, for example, interagency cooperation, periodic audits, training, reporting laws, and tracking systems .23

## Policy Advisory and Oversight Body

The 1982 OTA study documented a long history of debate—at least since 1970—over which organization(s) should have a formal policy advisory and oversight role with regard to a national computerized criminal history system. Policy control over any system for the interstate exchange of criminal history information is complicated by several factors:

- the involvement of a wide range of criminal justice agencies—from law enforcement and prosecutorial to judicial and correctional-as providers and users of criminal history information,
- the frequently conflicting Federal and State laws on noncriminal justice access and use,
- the trend towards increasing use of criminal history record checks for employee screening and other noncriminal justice purposes,
- the inevitable tension between Federal and at least some State governments in a sensitive area of interstate activity, and
- the implications of record use for privacy and constitutional rights.

Current policy control over the Triple I is vested in the Attorney General of the United States who has delegated authority to the FBI with a strong advisory role assigned to the NCIC Advisory Policy Board (APB). APB is comprised of 30 representatives:[24]

- 20 law enforcement members elected from the States and localities;
- 6 members appointed by the FBI Director (2 each from the judiciary, prosecutor agencies, and correctional institutions); and
- 4 members appointed by criminal justice associations (1 each by the International Association of Chiefs of Police, National Sheriff's Association, National District Attorney's Association, and National Probation and Parole Association).

However, now that the NCIC/CCH file has been terminated, APB has not defined a clear role for itself with respect to criminal history records beyond the pilot testing and operation of Triple I. The FBI's Identification Division still maintains a large, increasingly computerized criminal history record system, but has no advisory board or council similar to APB. Should an advisory or oversight board be created for criminal history record exchange, either a new board or a modification of APB, membership could encompass groups not currently represented on APB. These could include representatives of, among others, defense attorneys, civil liberties groups, research criminologists (from government or academia), and social scientists concerned with the effects of criminal records on rehabilitation.

[19] U.S. Department of Justice, Federal Bureau of Investigation, *NCIC 2000 Project Statement of Work*, Washington, DC, January 1985, p. A-9.

[20] U.S. Department of Justice, *Minutes*, op. cit., p. 226.

[21] Illinois Criminal Justice Information Authority, "Many 'Rap Sheets' Not Automated, Audit Finds, " *The Compiler*, vol. 6, No. 2, summer 1985, pp. 3, 8. Also see Bureau of Justice Statistics, *Data Quality*, op. cit. The State of Illinois now has a uniform disposition reporting law and the Criminal Justice Information Authority has prepared an advisory for criminal justice agencies.

[22] Bureau of Justice Statistics, "State Criminal Records, " op. cit., p. 4.

[23] Ibid. Also see, for example, improvements in disposition reporting cited in the State of California, per Nov. 18, 1985 memo from Roy T. Iwata, Manager, Disposition Update Section, Record Analysis and Processing Program, Bureau of Criminal Identification.

[24] U.S. Department of Justice, *NCIC 2000*, op. cit., p. A-10.

SEARCH Group, Inc., has, for example, repeatedly taken the position that an advisory body for interstate criminal history record exchange should be more broadly constituted than the present APB. SEARCH Group has stated that the board "be predominantly representative of the States" and that "its representation should ensure that it is responsive to all components of the criminal justice community, not just law enforcement. " SEARCH Group also believes that "public interest positions, representing the public at large as well as components of the criminal justice community, must be appropriately represented on the board to ensure that policy decisions are consistent with broad, national considerations. "[25]

As long as there is no clear advisory or oversight body for criminal history records exchange, whether APB or some other group, the policy control issue is further complicated by FBI proposals for new intelligence applications of NCIC, for example to include files on white-collar crime and organized crime suspects and associates-as contrasted with the existing wanted persons file, which is limited to persons who have been charged with a crime. These kinds of proposals pose difficult questions. On the one hand, intelligence applications aggravate already existing concerns about record quality and raise new concerns about possible abuse or misuse." On the other hand, the one intelligence file now on NCIC (the Secret Service file) apparently has proved useful, and similar applications may be helpful in other areas.[27]

[25]SEARCH Group, Inc., policy statement as reprinted in Federal Bureau of Investigation, National Crime Information Center, agenda materials for NCIC Advisory Policy Board meeting, Oct. 17-18, 1984, p. 63.

[26]See, for example, *Privacy Journal*, November 1984, p. 2, and August 1985, pp. 1, 3; Faye A. Silas, "A Bad Rap; Snafus in Computer Warrants, " *ABA Journal*, January 1985, pp. 24-25; "Jailing the Wrong Man, " *Time*, Feb. 25, 1985, p. 25; Donna Raimondi, "False Arrests Require Police To Monitor Systems Closely, " *Computerworld*, Feb. 25, 1985, p. 23; Charles Babcock, "On-line Crime Suspect System Implicated in False Arrest, " *Computerworld*, Aug. 19, 1985, p. 12; and John Bennett, "White-Collar Crime File Draws Ire of Left, Right, " *Arlington Journal*, Oct. 23, 1985, p. 2. Also see U.S. Congress, House Committee on the Judiciary, Subcommittee on Civil and Constitutional Rights, *Proposed Contract To Study and Redesign the National Crime Information Center*, Oversight Hearing, 98th Cong., 2d sess., Aug. 1, 1984.

[27]For further discussion, see U.S. Congress, Office of Technology Assessment, *Federal Government Information Technology: Electronic Surveillance and Civil Liberties*, OTA-CIT-293 (Washington, DC: U.S. Government Printing Office, October 1985), esp. ch. 5 section on "Data Base Surveillance. "

## Policy Implications

The issues discussed above raise the following policy questions:

First, how should differences between and among State and Federal laws on noncriminal justice criminal history record checks be reconciled? Presumably, this should be done in a way that reasonably ensures that, for record checks deemed to be lawful and in the public interest, criminal history information will be complete, accurate, and timely. Differences could be reconciled by Federal law, interstate compact, or a set of uniform State laws.[28] Failing any of these, an option would be to use a national full-record file for noncriminal justice purposes, while retaining the Triple I for criminal justice purposes only. A national file maintained by a Federal agency, such as the FBI, would be governed by Federal, not State, laws on record access and dissemination. 'g

Second, how can record quality be improved? Independent audits of Federal and State criminal history record files could be required. The existing FBI audit function could be extended to include State and local criminal history records that support Triple I index entries (and related Automated Identification Division System records). An audit function could be assigned to APB or some other advisory body. Congress could enact legislation, along the lines previously proposed by Representative Charles Schumer, that would establish and fund a record quality audit program.'" Whatever the mechanism, the audits could be conducted so as to produce quantitative estimates of record completeness and accuracy to provide a firm basis for measuring record quality improvement (or lack thereof).

Actually, the current FBI audit process provides a good prototype. As part of the audit function, the FBI audit team selects a statistically valid sample of NCIC entries from the NCIC wanted persons and stolen vehicles files and compares the record contents with State and local source information (e.g., from courts and prosecutors) to determine whether the records are accurate and valid. This FBI record quality audit procedure is similar to that used by OTA as reported in the 1982 study. Indeed, the results of FBI audits of five States in-

[28]See Young, Federal *Non-criminal* Justice Use, op. cit.; and SEARCH Group, Inc., *Use of III for Noncriminal Justice Purposes*, op. cit.

[29]SEARCH Group, Inc., Ibid., p. 20.

[30]See H.R. 896, Jan. 31, 1985, H. R. 2129, Apr. 18, 1985, and an amendment in the nature of a substitute to H.R. 2129 (discussion draft), Nov. 12, 1985, all entitled the "Criminal Justice Information Improvement Act of 1985, " 99th Cong., 1st sess.

dicated that an average of 5.5 percent of the NCIC wanted persons entries were invalid,[31] almost identical to the 5.8 percent result obtained by OTA.[32] The FBI found comparable error rates in the NCIC stolen vehicles files from the same five States.[33] Overall, the FBI audit process appears to be successfully identifying record problems and possible solutions with respect to these two files, and could be extended to include criminal history record files that are relevant to Triple I.

Third, what kind of national policy council or board should oversee the interstate exchange of criminal history records? Policy oversight issues include, for example: 1) should an advisory policy board have more than advisory power? 2) should the board report to the Attorney General or the FBI Director? 3) should the board have a broader composition when compared to the present APB to reflect the growing noncriminal justice use of criminal history records? 4) should the board include State representatives appointed by the respective Governors rather than, or as a complement to, those elected by law enforcement practitioners? and 5) should a separate board be established with respect to noncriminal justice uses and concerns, while retaining the current APB for criminal justice applications?[34]

One option is to establish statutory guidelines for the role and composition of an advisory body." Another option, not necessarily mutually exclusive, is to assign some oversight responsibilities to any independent Federal data or privacy protection board that might be established (as discussed in ch. 6). One reason that law enforcement and *criminal* justice record systems were exempted from key provisions of the Privacy Act of 1974 was the expectation at that time that separate criminal justice record privacy legislation would be enacted shortly. One of the legislative proposals at that time, introduced by the late Senator Sam Ervin, Jr., would have established a Federal Information Systems Board. While congressional hearings were held, neither this nor related proposals ever were reported out of committee or voted on by the House or Senate.[36]

[31] See Federal Bureau of Investigation, National Crime Information Center Audit Reports for Wisconsin (September 1984), Oregon (October 1984), Arizona (December 1984), Alabama (March 1985), and South Carolina (April 1985).

[32] See U.S. Congress, Office of Technology Assessment, *Computerized Criminal History System*, op. cit., pp. 191-192; also see Kenneth C. Laudon, "Data Quality and Due Process in Large Interorganizational Record Systems," *Communications of the ACM, vol. 29, No. 1*, January 1986, pp. 4-11; David Burnham, "FBI Says 12,000 Faulty Reports On Suspects Are Issued Each Day," *The New York Times*, Aug. 25, 1985; and David Burnham, "Computer Data Faulted in Suit Over Wrongful Arrest," New *York Times*, Jan. 19, 1986.

[33] See FBI NCIC Audit Reports, op. cit.

[34] See OTA, *Computerized Criminal History System*, op. cit., pp. 169-172.

[35] This approach was taken in the original version of H.R. 2129, the Criminal Justice Information Improvement Act of 1985, 99th Cong., 1st sess. A later draft version, dated Nov. 12, 1985, in the nature of a substitute, was limited to record quality matters.

[36] See OTA, Computerized *Criminal History System*, op. cit., PP. 73-74, and S. 2963, the Criminal Justice Information Control and Protection of Privacy Act of 1974, 93d Cong., 2d sess. Also see U.S. Congress, Senate Cornmittee on the Judiciary, Subcommittee on Constitutional Rights, *Criminal Justice Data Banks, Hearings*, 93d Cong., 2d sess., March 1974; *Criminal Justice Information and Protection Privacy Act of 1975, Hearings*, 94th Cong., 1st sess., July 15 and 16, 1975; U.S. Congress, House Coremittee on the Judiciary, Subcommittee on Civil and Constitutional Rights, *Criminal Justice Information Control and Protection of Privacy Act of 1975; Hearings*, 94th Cong., 1st sess., July 14, 17, and Sept. 5, 1975; and Donald A. Marchand, *The Politics of Privacy, Computers, and Criminal Justice Records* (Arlington, VA: Information Resources Press, 1980).

## Appendix B

# OTA Federal Agency Data Request

After reviewing all available sources of information on Federal use of information technology, OTA determined that important information was not available in certain areas critical to the OTA assessment. To meet the need for additional information, OTA drafted a request for current agency data covering the areas in which information was lacking or incomplete. The draft request was reviewed by congressional staff of interested committees, and then pretested in four agencies—the Energy Information Administration (Department of Energy), the Food and Nutrition Service (Department of Agriculture), the Office of the Assistant Secretary for Postsecondary Education (Department of Education), and the Veterans Administration. Based on the results of the pretest, the data request was revised. (See attachment 1 for portions of the final, revised data request relevant to this report.)

In April 1985, the data request was sent to the 13 cabinet-level agencies and 20 selected subcabinet agencies (see attachment 2) with a turnaround time of 5 weeks. Sufficient copies were provided for each of the subcomponents of the cabinet agencies. Agencies were informed that no new data collection was to be conducted. An OTA staff member was identified who could be contacted to provide clarification where necessary.

All agencies that were sent the request provided a response, although the responses varied in completeness and quality. A total of 142 agency components provided information. While many of the agencies provided responses well within the time allotted, the completion time for the entire request (142 agency components) was approximately 2 months. The data provided were compiled by OTA staff and appear as appropriate throughout the report.

A draft copy of the OTA report was provided to each of the participating agencies for review and comment.

ATTACHMENT 1

III. Privacy Act (General)

A. Please provide the following data on Privacy Act Implementation in your agency:

1. Position and GS level of the Privacy Act Officer or agency official with day-to-day operating authority
2. Position and level of agency official. with policy authority
3. Total number of agency staff (In full-time equivalents) assigned to Privacy Act matters
4. Role and responsibility of your agency's Office of Inspector General (e.g., in developing internal agency procedures, responding to Privacy Act requests, preparing Privacy Act materials for OMB).

B. Please specify the procedures your agency follows to ensure Privacy Act record quality, e.g., complete and accurate records. Attach a copy of agency regulations or procedures.

C. Does your agency conduct record quality audits? Yes No . If yes, please provide the results of such audits, including copies of any written audit reports.

D. Has your agency developed agency-specific guidelines or procedures for determining what is "relevant" and "timely" information within your agency? Yes ____ No ____. If yes, please provide a copy of such guidelines.

E. Has your agency been a defendant in Privacy Act suits at any time since 1980? Yes ____ No ____. If yes, please list or describe the legal action(s) and basic issue(s) and provide citations

F. Has your agency revised or updated Privacy Act guidelines with respect to microcomputers? Yes ____ No ____. If yes, please provide a copy of such revised or updated guidelines.

Name _____ Agency/Unit _____

Title _____ Telephone No. _____

## Iv. Privacy Act/Computer Matching and Front-End Verification

A. Has your agency Participated in computer matching activities* as a matching agency (the agency performing the match) or as a source agency (the agency disclosing records to the matching Echlng agency for use in the match) at any time since 1980? Yes No Please provide a copy of any reports on your matching activities including the information listed below, to the extent available- Please give priority to information on matches conducted in 1984, with complete quantitative data provided where possible.

1. Date of match
2. Participating parties (indicate source and matching agencies):
   Federal agencies
   State agencies
   Private sector organizations
3. Location of match
4. Frequency of match: one time or ongoing
5. Files matched
6. Method(s) used to exchange records (e.g., direct electronic~ computer tape, computer disk)
7. Purpose of match
8. Number of records involved
90 Number of hits
100 Percentage of hits verified

B. Are cost-benefit analyses done prior to- computer matching? Yes No • If yes, what are the quantitative and qualitative categories used for assessing costs and benefits? How are the cost-benefit analyses used within the agency? Please provide a copy of your agency's three most recent cost benefit analyses.

C. Do the indivldual subjects of the match provide written consent prior to a match? Yes ___ No ___. If yes, please attach a copy of the consent form.

D. Are your matches explicitly required or authorized by legislation? Yes No If yes, please list matches required or authorized and cite public law section for each type of match.

E. Are procedures used to ensure that the subject record files contain accurate information? Yes ___ No ___. If yes, please specify the procedures used.

---

*Defined as the computerized comparison of two or more automated systems of records to identify individuals common to two or more of the record systems or unique to one of the record systems.

F. What is the process once a hit has occurred? What are the standards , procedures, and costs (estimate if necessary) for verification?
What *is* the appeal process, within the agency and outside, for an individual to respond to a "hit"? Have there been any court challenges to the matches? Yes ___ No ___. If yes, what were the results? Please attach case numbers.

G. Are cost-benefit analyses done after matches? Yes No If yes, please provide a copy of your agency's three most recent post-match cost-benefit analyses.

H. Has your agency used computerized front-end verification (i.e., certification of the accuracy and authenticity of information supplied by an applicant by checking against similar information from another agency or source) at any time since 1980 as part of the application process for participation in Federal programs or benefits? Yes No If yes, please provide a copy of any agency reports on your use of front-end verification and describe the process, including use of computers, notice to applicants, and costs. If no, please describe any agency plans for use of front-end verification.

1. What have been the average results of front-end verification as measured by hits (i.e., applicant's eligibility for Federal program or benefit not verified) overall and by Federal program or benefit category. If available, please break down by computerized and manual verifications.

J. Has your agency conducted any cost-benefit studies of front-end verification? Yes ___ No ___. If yes, please provide copies of the three most recent studies.

Name _____ Agency/Unit _____

Title _____ Telephone No. _____

## V. Privacy Act/Third Party Information and Profiling

A. Does your agency collect any personally-identifiable information in electronic form from third party sources (i.e., from sources other than the subject individual)? Yes ____ No ____. If yes, please provide information on third party collection, including nature of information sources, authority for collection, agency use, procedures to assure accuracy, subject individual's rights to access, *review*, and challenge the information, and secondary dissemination of third party information outside the agency (specify to whom and for what purpose). If no, please describe any agency plans for collecting third party information.

B. Does your agency use computer-assisted statistical programs and/or related software co develop generic profiles of types or categories of individuals and/or probabilities of such categories of Individuals engaging in activities or behavior of interest to the agency (e.g., with respect to misrepresentation of eligibility to receive Federal aid or benefits, non-compliance with or violation of agency regulations, violation of civil or criminal statutes)? Yes No If yes, please provide further details below. If no, please describe any agency plans for the use of such profiling.

c* For each specific use of profiling, please provide the following information, to the extent available:

1. Description of profiling (categories and numbers of individuals, types of behavior)
2. Types of programs and/or software used
3. Development and testing of programs and/or software (please be specific; provide a copy of any written research reports)
4* Source(s) of input data
5. Authority for the profiling (cite specific statute or regulation where applicable)
6. Agency use of the profiling
7. Results of agency use of the profiling (e.g., percentage of hits on targeted individuals, civil and/or criminal penalties imposed). Please provide a copy of any profiling evaluation reports.

Name _____ Agency/Unit _____

Title _____ Telephone No. _____

## VI* Privacy Act/Debt Collection Act

A. Does your agency report or refer delinquent and/or nondelinquent commercial and/or consumer (individual) debts to private sector credit bureaus? Yes No . If yes, please provide further details below. If no, please describe any agency plans for the use of private sector credit agencies.

B. For each specific type of debt referred to private sector credit bureaus, please provide the following Information, to the extent available:

1. Description of type of debt referred
2. Format of referral (e.g., paper, microfiche, computer tape, direct electronic)
3. Procedures/agreements between the agency and credit bureau with regard to:
   - o security
   - o record quality (completeness and accuracy)
   - o secondary dissemination
   - o subject individual's or organization's access, review, and challenge rights
4. Number and type of complaints received from debtors referred to private sector credit bureaus, and resolution of those complaints
5. Results of debt referrals by type of debt (e.g., dollars recovered and as percentage of debt referred)

C. Does your agency use private sector credit reports in making agency decisions about eligibility for Federal programs and benefits? Yes No . If yes, please provide details on the specific purposes of such use (e.g.,when awarding loans, contracts, grants).

Name _____ Agency/Unit _____

Title _____ Te eph e N _____

VII. privacy Act/Electronic Records Management and Electronic Mail

A. Please estimate, to the extent possible, the number and percentage of manual versus computerized records maintained by your agency in the following categories for fiscal years 1975 and 1984:

| | | Manual | | Computerized | | Total | |
|---|---|---|---|---|---|---|---|
| | | No. | % | No. | z | No. | % |
| Records subject to Privacy Act | 1975 | | | | | | |
| | 1984 | | | | | | |
| Other records maintained subject to public law or agency regulation | 1975 | | | | | | |
| | 1984 | | | | | | |

B. If your agency maintains one or more record systems subject to the privacy Act, please list the 10 largest Privacy Act record systems, the total number of persona and records in **each system and the percentage** of manual versus computerized records for each system.

| Record System | No. Persons | No. Records | %Manual | %Computerized |
|---|---|---|---|---|
| 1. | | | % | % |
| 2. | | | | |
| 3. | | | | |
| 4. | | | | |
| 5. | | | | |
| 6. | | | | |
| 7. | | | | |
| 8. | | | | |
| 9. | | | | |
| 10. | | | | |

c* For your agency's computerized records (e.g., records stored in electronic form on computer tape or disk), please provide the following information, to the extent available:

1. Procedures for backup copies (please estimate percentage of records backed up by each of the following: paper copy, microfiche or microform, duplicate computer tape or disk, no backup, more than one backup)

2. Procedures for storage and maintenance of electronic records (please specify how long such records are stored) what protections are used to protect against alteration, and when and how electronic records are archived, i.e. , moved off premises to a remote storage location)

3. Procedures for purging of electronic records (under what conditions and when are records purged, i.e., eliminated or destroyed)

4. Procedures for verification of signatures on or authenticity of electronic records

5. Procedures for duplication or copying of electronic records (e.g., what is the agency definition of "record copy" of an electronic record)

D. Does your agency use electronic mail? Yes No If yes, please provide further details below. If no, please decribe any agency plans for use of electronic mail not otherwise described in response to Section I.

E. Please provide the following Information, to the extent available, on your agency's use of electronic mail.

1. Total volume in number of messages sent (I.e., pieces of electronic mail) per year for fiscal year 1984
2. Type of electronic mail system used (e.g., in-house, outside contractor, commercial)
3. Total volume in number of messages received per year for 1984
4. Content of messages sent (in percentage of 1984 total):

| Purpose | Percentage |
|---|---|
| Intra-agency correspondence/memos | ____% |
| Intra-agency records/reports | ____ |
| Interagency correspondence/memos | ____ |
| Interagency records/reports | ____ |
| External correspondence/memos | ____ |
| External records/reports | ____ |

5. How long are backup message copies retained in electronic and/or paper form?

6. Who participates in electronic mall? (Specify type of agency staff, e.g., administrative, secretarial, technical, research)

F. Does your agency have a set of privacy/confidentiality/security practices or policies developed specifically for electronic mail? Yes ___ No ___. If yes, please provide a copy or describe in detail.

Name _____ Agency/Unit _____

Title _____ Telephone No. _____

## VIII. Investigative, Law Enforcement, and Intelligence Applications

A. Does your agency maintain computerized record *systems* for investigative, law enforcement, and/or intelligence purposes? Yes . No . If yes, please provide the detailed information below.

B. For each such computerized record system, please provide the following information, to the extent available:

1. Name of record system
2. Purpose of record system
3. Number of records
4. Number of persons
5. Types of record information (e-g-, individual names, social security number, address)
6. Sources of record information
7. Users of record systems and rules on access
8. Statistics on quality of records and procedures for maintaining record completeness and accuracy

C. Does your agency use computer-assisted statistical programs and software to develop profiles of types or categories of individuals engaging or likely co engage in activities of investigative, law enforcement, and/or intelligence interest to your agency? Yes . No If *yes,* please provide further details below. If no, please describe any agency plans for the use of such profiling.

D. For each specific use of computer-based profiling, please provide the following information, to the extent available (and not otherwise provided in Section V):

1. Description of profiling (categories and number of individuals, types of behavior)
2. Types of programs and/or software used
3. Development and testing of programs and/or software (Please be specific; provide a copy of any written research reports)
4. Sources(s) of input data
5. Authority for the profiling (cite specific statute or regulation where applicable)
60 Agency use of the profiling
7. Results of agency use of the profiling (e.g., percentage of hits on targeted individuals, civil and/or criminal penalties imposed). Please provide a copy of any profiling evaluation reports.

## Attachment 2.—Federal Departments and Agencies Responding to OTA Data Request

| *Cabjnet department* | *Number of components agency responding* |
|---|---|
| Agriculture | 25 |
| Commerce | 17 |
| Defense | 14 |
| Education (agencywide) | 2 |
| Energy (EIA, FERC, and rest of agency) | 3 |
| Health and Human Services | 9 |
| Housing and Urban Development (agencywide) | 1 |
| Interior | 9 |
| Justice | 13 |
| Labor | 8 |
| State (agencyWide) | 1 |
| Transportation | 11 |
| Treasury | 9 |
| Subtotal | 122 |
| ***Independent agencies*** | |
| Commission on Civil Rights | 1 |
| Consumer Product Safety Commission | 1 |
| Environmental Protection Agency | 1 |
| Equal Employment Opportunity commission | 1 |
| Federal Communications Commission | 1 |
| Federal Elections Commission | 1 |
| Federal Emergency Management Agency | 1 |
| Federal Reserve System | 1 |
| Federal Trade Commission | 1 |
| General Services Administration | 1 |
| National Aeronautics and Space Administration | 1 |
| National Archives and Records Administration | 1 |
| Nuclear Regulatory Commission | 1 |
| Securities and Exchange Commission | 1 |
| Selective Service System | 1 |
| Small Business Administration | 1 |
| Arms Control and Disarmament Agency | 1 |
| U.S. Information Agency | 1 |
| Agency for international Development | 1 |
| Veterans Administration | 1 |
| Subtotal | 20 |
| Total | 142 |

## Appendix C

# List of Contractor Reports

Copies of the following contractor reports completed in support of this assessment will be available in late 1986 from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161, (703) 487-4650.

1. William H. Dutton and Robert G. Meadow, *Public Perspectives on Government Information Technology: A Review of Survey Research on Privacy, Civil Liberties, and the Democratic Process,* Annenberg School of Communications, University of Southern California, prepared for OTA, January 1985.
2. David Flaherty, *Data Protection and Privacy: Comparative Policies,* prepared for OTA by The Privacy Project, University of Western Ontario, Jan. 8, 1985.
3. Karen B. Levitan, Patricia D. Barth, and Diane Griffin Shook, *Agency Profiles of Civil Liberties Practices,* prepared for OTA by The KBL Group, Inc., Dec. 28, 1984.
4. Robert Ellis Smith, *Report on Data Protection and Privacy in Seven Selected States,* prepared for OTA, Feb. 15, 1985.

Appendix D

# Other Reviewers and Contributors

Ralph W. Adams
National Security Agency

Patricia Aronsson
National Archives and Records Administration

William L. Ball
U.S. Department of State

Robert P. Bedell
Office of Management and Budget

Jane Bortnick
Congressional Research Service

Frank G. Burke
Acting Archivist of the United States

Richard Ehlke
Congressional Research Service

Kenneth R. Erney
U.S. Department of State

Liz Handley
U.S. Department of Health and Human Services

Mary C. Lawton
U.S. Department of Justice

Fred Lothrop
PSC, Inc.

Gary Marx
Massachusetts Institute of Technology

Francis A. McDonough
U.S. General Services Administration

Sandra Milevski
Congressional Research Service

Oscar W. Mueller, Jr.
U.S. Department of the Interior

David Mullins
U.S. General Services Administration

Dale Nesbary
National Conference of State Legislatures

Hugh O'Neill
Formerly U.S. Department of Health and Human Services

Ronald S. Plesser
Blum, Nash & Railsback

Edward J. Regan
Manufacturers Hanover Trust Co.

Nancy Reichman
University of Denver

Harold Relyea
Congressional Research Service

David N. Richardson
Yankelovich, Skelly & White, Inc.

Alice Robbin
University of Wisconsin, Madison

Roger K. Salaman
National Telecommunications and Information Administration
U.S. Department of Commerce

Gail Shelton
U.S. Department of Health and Human Services

Ollie R. Smoot
Computer & Business Equipment Manufacturers Association

Appendix E

# Summary of Final Rules for Income and Eligibility Verification Required Under the Deficit Reduction Act of 1984*

The Departments of Agriculture, Labor, and Health and Human Services issued final rules in the Federal *Register* on February 28, 1986, to implement Section 2651 of the Deficit Reduction Act of 1984 (DEFRA). Section 2651 amended the Social Security Act, the Food Stamp Act, and the Internal Revenue Code to require federally funded public assistance and unemployment agencies to improve the accuracy of eligibility determinations and benefit programs by exchanging information with each other and by obtaining unearned income data from the Internal Revenue Service (IRS) and other income and wage data from the Social Security Administration (SSA) and from State wage and Unemployment Insurance Benefit (UIB) data files. The rules require State agencies to develop an Income and Eligibility Verification System (IEVS) for administering the following programs:

1. The Food Stamp Program under the Food Stamp Act of 1977, as amended.
2. The Aid to Families With Dependent Children (AFDC) Program under Title IV-A of the Social Security Act; the Adult Assistance Programs under Titles I, X, XIV, and XVI of the Social Security Act.
3. The Medicaid Program under Title XIX of the Social Security Act.
4. The Unemployment Compensation Program under Title III of the Social Security Act.

*Use of IEVS Data.*- *IEVS* data can be used to obtain information for prosecutions, i.e., as the basis for investigations in the same way as it is used as a basis of inquiry about household circumstances.

*Oversight and Coordination of IEVS.*-*No* specified type of oversight requirement on States; no statutory requirement on States to organize implementation of IEVS in any special or uniform way; no plan to add to existing Federal oversight mechanisms; not feasible, within established timeframes, to establish uniform guidelines and programming specifications for the required matches.

*Access and Use of* Information.-Data must be requested from all of the required sources on applicants for Medicaid, AFDC, adult assistance, and food stamp programs at the first available opportunity, which would be the next scheduled match for each source. The State Wage Information Collection Agency (SWICA) and the State Unemployment Compensation Agency must accept and process requests for wage information at least twice a month. Requests for IRS data for applicants must be made at the first available monthly IRS match date. With regard to requesting data from SSA, at the first available opportunity, the applicant should be processed in the next cycle of the Beneficiary and Earnings Data Exchange (BENDEX) System or queried through the Third Party Query (TPQY) System.

*Timeframes.*-Proposed rules required that IEVS information be used to determine eligibility within 20 calendar days of receipt. Final rules extended this to **30** days because of the need to verify IEVS information.

*Cost Effectiveness.*-". . . *all* of the required information sources have been demonstrated to be useful in preventing incorrect eligibility y and benefit amounts, either by directly offsetting costs or by helping deter nonreporting by applicants and recipients" (p. 7183).

*Automation.*— "We encourage States to develop on-line systems and other methods for rapid turnaround of State agency requests so that wage and UIB data can be used to determine eligibility and benefits of applicants" (p. 7180). "We encourage the use of on-line systems for front-end verification, but our rules do not require States to have this capability" (p. 7181). "SSA and IRS have not found it cost effective to make the wage and self-employment (SSA) and unearned income (IRS) information accessible on-line for their own agency purposes. Therefore, it would not be feasible to allow States on-line access to these files. SSA has the capability of providing on-line access to bene-

*The final rules appeared in the ***Federal Register*** on Feb. 28, 1986 (vol. 51, No. 40, pp. 7178-721 7). The proposed rules were published in the *Federal* ***Register*** on Mar. 14, 1985 (50 FR 10450). Comments on the proposed rules were received from 53 parties: 38 States, 6 client advocate groups, 4 local or county welfare agencies, 4 Federal agencies, and 1 private citizen.

fit data. A pilot project is being conducted with Tennessee to provide wire-to-wire exchange of benefit data" (p. 7184).

In the proposed rules, it was stated that "the statutory requirements for IEVS mandated a logical process and not necessarily a physical or automated system to assure the timely and efficient exchange of information among the various programs. " It was recognized that "an increasing number of States are operating automated on-line systems to exchange, maintain and make data available to workers, but this level of automation was not required. Many commenters suggested that automation would be required to meet IEVS requirements fully. The Federal agencies agreed that "automating the required IEVS functions would enhance a State agency ability to respond in a timely fashion to the substantial amount of information made available to the State agencies as a consequence of the data exchange requirements, " but did not believe that such automation should be required in the rules (p. 7194).

*State Wage Information Collection Agencies.* –Final rules retain requirement for quarterly wage matching. Employers in each State are required to report wages quarterly.

*Unemployment Insurance Benefits.* –Agencies are required to do data matches for UIB information at application and for 3 months following application or loss of employment. For the Food Stamp Program, in addition to wage and UIB information, State agencies are required to request and utilize any information available from Unemployment Compensation (UC) agencies to the extent permitted.

*Internal Revenue Service.* –An annual match of recipients against IRS data on unearned income is required. IRS has scheduled 11 monthly runs of State tapes against its national file of unearned income information. IRS will only process one tape per month per State.

Social *Security Administration.* –State agencies are required to access all available SSA data on applicants by using the TPQY system (for SSA benefit data) or the BENDEX System (for pension, earnings, and self-employment information). If TPQY is used, when the applicant becomes a recipient the State agency must add the name to BE NDEX. Regarding data quality, the final rules emphasize two factors: 1) except for UC and SSA benefit data, the information obtained through IEVS will be generally treated as a lead for further verification activity, for example, SSA earnings will almost always need to be verified; and 2) "if a State receives what they believe [sic] is incorrect information, no adverse action should be initiated until the discrepancy is resolved" (p. 7186).

*Interprogram and Interstate Exchange. -All* programs in IEVS are required to exchange income and eligibility information with each other in accordance with interstate and intrastate agreements in effect and as appropriate to the requesting program's verification and eligibility determination needs. State agencies are encouraged to request data from adjacent jurisdictions and other States where experience indicates the data would be useful. States may also access the State Employment Security Internet System for IEVS matches, although this is not a requirement. The Internet System is still under development and its potential uses are still being evaluated by the Department of Labor.

*Alternate Sources.* —A State agency may obtain data from sources other than those specified in the regulations (from banks, for example) if it can demonstrate to the respective Secretaries that the alternate source furnishes data as timely, complete, and useful as data from the source specified in the regulations.

*Independent* Verification. -Independent verification is an inquiry about a possible discrepancy in the information reported by the individual and information reported from other sources. Information can be independently verified by contacting the applicant or a third-party source (for example, the employer or bank that reported the information). "The option of contacting a third party is necessary in cases where the recipient fails or refuses to cooperate, the State agency believes it to be in the interest of the investigation of potential fraud or when other factors indicate that a third party contact is preferable" (p. 7188).

DEFRA requires independent verification of IRS unearned income. With respect to other information obtained through IEVS, the food stamp program set explicit guidelines for verification, while the AFDC, adult assistance, and Medicaid programs require independent verification of IEVS information if determined appropriate based on agency experience. "The State agencies remain responsible for ensuring that any information they use in determining eligibility and payment amounts is correct" (p. 7196).

*Social Security Numbers: Furnishing, Using, and Verifying.-DEFRA* requires each applicant for, and each recipient of, AFDC, adult assistance in the territories, food stamps, unemployment compensation, and Medicaid to furnish his or her so-

cial security number in order to associate information on applicants and recipients for the required matches. Existing AFDC and food stamp program rules already require the furnishing of social security numbers. All State agencies implementing Medicaid, AFDC, food stamp, and adult assistance programs must verify applicant and recipient social security numbers to ensure efficient administration of the matching programs and to prevent improper disclosure of information. However, eligibility determinations cannot be delayed pending social security number verification.

Social security numbers can be verified through the BENDEX, State Data Exchange (SDX), TPQY, and social security number verification systems. There is no required order for using these systems. SSA generally verifies the social security numbers of recipients of title II or title XVI benefits. Therefore, a social security number for such an individual received through BENDEX can be considered verified. However, not all social security numbers in BENDEX are verified. At present, the social security number verification system is being redesigned, and when completed, verification of social security numbers should be completed within 3 weeks. On-line access to the social security number verification system is not feasible at this time. SSA is working on a pilot project with Tennessee to provide wire-to-wire exchange of benefit data, including verification of social security numbers. SSA expects to offer the same service to other States.

*Routine Notice to* Individuals. -DEFRA requires that all applicants and recipients be notified that information available through IEVS will be requested and utilized. Notification is to be given at application and periodically thereafter, i.e., based on existing program case-processing cycles. Notice must be written and must inform the individual that income and eligibility y information may be obtained using his or her social security number and will be used in determining eligibility. The notice must include the types of agencies that will be contacted, for example, unemployment compensation agencies.

The Departments of Labor, Agriculture, and Health and Human Services "believe that State agencies should obtain assurances from provider agencies that their automatic data processing methods prevent providers from recording what recipient names and/or social security numbers are processed and that individuals having access to such information are bound by the disclosure rules of the various programs" (p. 7191).

*Notice of Expiration or Adverse Action.* –Under the proposed rules, the applicant or recipient had to be notified of any planned adverse action and had to be given the opportunity for a fair hearing. The food stamp program proposed rules also included a provision under which households that failed to respond in a timely fashion to State agency requests for information would be sent a notice of expiration of their certification period. The final rule replaces the proposed use of the notice of expiration with a notice of adverse actions when a household does not respond in a timely fashion to a State agency inquiry about IEVS information.

*Safeguards for Confidentiality. -DEFRA* requires each State agency to institute adequate safeguards to assure: "(l) that information is made available only to the extent necessary to assist in the valid administrative needs of the program receiving the information and that unearned income data from IRS is exchanged only with those agencies authorized to receive it; and (2) the information is adequately protected against unauthorized disclosure for other purposes" (p. 7192).

Oversight.–DEFRA did not mandate any reporting system to gather information on actions taken and savings realized. The proposed rules asked for comments on such a system. In the final rules, the Departments of Agriculture, Labor, and Health and Human Services stated that reporting "is necessary to help ensure the proper and efficient administration of the programs, " and that they were "developing uniform, annual reporting requirements intended to minimize the recordkeeping and reporting costs and burden on States, while enabling the Federal Government to monitor compliance with the requirements for accessing and using information" (p. 7197).

Appendix F

# Privacy and Data Protection Policy in Selected Foreign Countries[1]

Many Western European countries and Canada have also established policy to protect the collection and use of personal information. Many of these countries have created boards or commissions with responsibilities for overseeing government and private sector information practices, and acting as ombudsmen for individuals. Because the policies of these countries may serve as a model for policy actions in the United States, descriptions of the policies of several countries follow.

## The Federal Republic of Germany

The Federal Data Protection Act became law on January 27, 1977. Its provisions apply to both computerized and manual personal information systems in both the public and private sectors. Registration of all private and computerized public information systems is required under the act. Although the general principles regarding rights of individuals and restrictions on the collection and use of personal information are the same for public and private organizations, the methods of regulating the two sectors differ.

The act provides for the appointment of a Federal Commissioner for Data Protection to supervise public sector information systems. This position was added to the draft legislation at the insistence of the West German legislature; the original government bill did not call for such an official. The Commissioner, who serves for a 5-year term and may be reappointed once, has the authority to investigate complaints, inspect information systems, require information from agencies, and make recommendations. The Commissioner does not have licensing power. Nor does the office have enforcement powers; rather, the head of each public agency is responsible for ensuring compliance by the agency. The Commissioner serves, therefore, in an advisory capacity rather than a regulatory one. Up to now, the advice of the Commissioner has been taken seriously by the Federal agencies, including the national security agencies and the Federal police. In essence, it has not been politically viable for the heads of Federal agencies to ignore the Commissioner's advice, which is nor-really given privately at first and later as part of a process of negotiation over competing interests in the use of information. The Federal Commissioner for Data Protection is subject to supervision by the government and reports to both the Minister of the Interior and to Parliament.

Private organizations maintaining personal information systems are supervised by the Land (State) authorities to which the organization belongs. For example, the Land authority that regulates banking activity is now responsible for ensuring that the banks also comply with data protection rules.

## Sweden

Sweden was the first country to pass national legislation regarding the collection and use of personal information. The purpose of the 1973 Data Act was to protect the confidentiality of records, to rationalize the personal information policies of organizations, and to expand individual rights and state protection to private information systems. The Data Act covers all computerized personal information systems in the public and private sectors. It established a regulatory agency, the Data Inspection Board (DIB), which is independent of the government and has the responsibility for licensing all automated personal information systems in both the public and private sectors. The 1973 statute mandated DIB licensing in advance, but a more permissive and somewhat less bureaucratic system, focusing more on sensitive uses of personal information, was introduced in the 1982 revision. The revised law was designed to reduce the bureaucratic burden of data protection and to make the system of selective licensing of personal information systems self-supporting. These revisions occurred in response to DIB's own internal assessment of what changes were necessary and the government general desire to reduce the costs of government. It is noteworthy that, because of Opposition fears of appearing to weaken data protection, the 1982 amendments passed by only one vote.

The Data Inspection Board has a Board of Directors, appointed for fixed terms, representing various political parties and interest groups, and a staff of less than 30. DIB exercises a great deal

[1] Material for this section was derived from David H. Flaherty, "Data Protection and Privacy: Comparative Policies," OTA contractor report, January 1985.

of power. It has the authority to control the collection and dissemination of personal data, to regulate the usages of the resulting register, and to set up a system of responsible keepers for computerized databanks. DIB also has the powers to investigate complaints, to inspect information systems, and to require information from organizations. The power of the cabinet or legislature to create a personal file outside the jurisdiction of DIB is an example of several safety valves in the legislation that prevent DIB from acting in a discretionary fashion on any specific measure.

The Data Act contains a few general data protection rules, for example, the data subject right of access and right to challenge are guaranteed in the act. But, DIB is responsible for designing detailed rules for particular systems and users, including what information may be collected, and the uses and disclosures of this information.

## France

The 1978 Law on Informatics, Data Banks, and Freedoms is an expansive and innovative statute. Article 1 well illustrates this point:

> Informatics ought to be at the service of each citizen. Its development should occur in the context of international cooperation. It ought not to threaten human identity, the rights of man, private life, nor individual or public freedoms.

The 1978 law created an independent administrative agency with regulatory power, the National Commission on Informatics and Freedoms (CNIL). It is the first administrative agency in France with statutory independence from the government. CNIL is obliged to ensure the observance of the 1978 law and to make decisions on the authorization of particular information systems in response to requests. The Commission has 17 part-time members chosen for 5-year terms by various official government bodies, including the Senate, the National Assembly, the Council of State, the Court of Cessation, and the Court of Financial Accounts. There are also data protection officials in each government agency.

Critics argue that CNIL has never taken a tough decision against the government with respect to a proposed new personal information system. CNIL has rarely turned down a government proposal; it tends to negotiate changes during the process of application for approval. Because of the way it works in responding to specific requests for advice or licenses, CNIL has not yet reviewed in detail all of the databanks that existed prior to the enactment of the 1978 law.

## United Kingdom

The Data Protection Act became law on July 12, 1984, and will gradually become fully operative over the next 3 years. The act established an independent Data Protection Registrar with a staff of 20 to 30 members who are not civil servants. They are to maintain a register of personal data users and computer bureaus in the public and private sectors. Although the Home Office emphasizes that the law requires simple registration of automated systems rather than licensing, as in Sweden and France, the act requires quite complete information on each system and the users of the system. It remains to be seen whether there are any practical differences in terms of the amount of paperwork required.

## Canada

Part IV of the Canadian Human Rights Act of 1977 introduced principles of fair information practice for the Federal public sector and created the position of Privacy Commissioner. The powers of the Commissioner consisted primarily in responding to complaints from individuals about denials of individual access to government personal data. The current Privacy Commissioner was a member of the Canadian Human Rights Commission.

In 1982, the Federal Privacy Act supplanted and significantly strengthened the privacy provisions of the Human Rights Act. Sections 4 to 10 of the 1982 act regulate the collection, retention, disposal, protection, and disclosure of personal information held by the Federal Government by means of a code of fair information practices. Its provisions are similar to the American Privacy Act. The Canadian law also specifies a list of 13 purposes for which a government institution may disclose personal information.

The Treasury Board is responsible for publishing an annual index of all the personal information systems maintained by the Federal Government in both manual and automated form, including the fewer than 25 systems that are exempt from access by individuals. The current edition runs to about 300 pages. Copies are available in post offices and libraries across Canada, but it is unusual to find persons who have consulted them.

The 1982 Privacy Act considerably strengthened the general powers of investigation and monitoring, and set up a separate Office of the Privacy Commissioner. The Privacy Commissioner holds office for 7 years, and is eligible for reappointment

once. His independence is assured, in theory, by the fact that he is an officer of Parliament and is appointed by resolution of the Senate and House of Commons. In practice, the initial selection is in the hands of the government of the day; thereafter, the Privacy Commissioner has to retain the confidence of these two legislative bodies. Presently, the Information Commissioner, who is responsible for the law on access to government information, and the Privacy Commissioner share some administrative staff in the same office. The Privacy Commissioner has a legal advisor, a director of complaints and 5 investigators, and a director of compliance and 3 investigators, for a total of 15 direct staff and a share of 18 others.

The Privacy Commissioner has the overall responsibility to monitor the implementation of the Privacy Act. His recommendations to government departments are likely to carry a considerable amount of weight, although he does not have regulatory power, because he is an independent officer of Parliament. He can request a response from a department to one of his recommendations. He prepares an Annual Report to Parliament and may make special reports at his discretion. The act directs that a permanent committee of Parliament should review the administration of the statute. An individual may complain to the Privacy Commissioner about any alleged form of personal information misuse by the Federal Government. Moreover, the Commissioner has the power and resources to initiate and investigate a complaint himself.

## Australia

In April 1976, the Australian Law Reform Commission was given a broad mandate to consider a variety of privacy issues, including data protection. After an exhaustive inquiry and the publication of a number of specialized reports, a comprehensive three-volume report was released at the end of 1983. With respect to its recommendations for data protection legislation, the Commission formulated 10 general principles for data protection modeled on the Organization for Economic Cooperation and Development's Guidelines. The Commission concluded that the private sector, as well as the public sector, should come within the ambit of legislation. It rejected the licensing model for data protection, but recommended the creation of a "statutory guardian" or "administrative body with the specific function of advocating privacy interests. " Such a Privacy Commissioner would function primarily as an ombudsman, but would have regulatory power in one specific area-the handling of individual requests to obtain access to their own data and to amend incorrect records. In general, the basic functions of the Australian Privacy Commissioner would be similar to those of his or her counterpart in Canada and data protection officials in Western Europe.

O

58-924 0 - 86

*Accessibility and Integrity of Networked Information Collections*

August 1993

OTA-BP-TCT-109
NTIS order #PB93-218923

Office of Technology Assessment

Accessibility
and
Integrity of
Networked
Information
Collections

BACKGROUND PAPER

Congress of the
United States

# Foreword

Technological advances in networking--ranging from widespread use of Internet to development of the National Research and Education Network and National Information Infrastructure--are ushering in a new era of information systems and online information resources. Networked information collections, or "digital libraries, " allowing online access to books, journals, music, images, databases, and multimedia works will be an important part of this new infrastructure. The extent to which these resources' potential is realized will depend, in part, on their accessibility and "user-friendliness. " But measures to increase user friendliness have an equally important counterpart in measures to ensure the integrity and proper use of network information resources. Otherwise, potential problems like plagiarism, corruption of databases, and improper use of copyrighted materials could impede development of networked information collections or limit their public accessibility.

Clifford Lynch's contractor report was prepared as part of an Office of Technology Assessment planning activity on the accessibility and integrity of digital libraries, in partial fulfillment of Section 2385 of the Rural Economic Development Act of 1990 (Public Law 101 -624).

In this contractor report, Lynch takes the perspective of the library community and its users and patrons in examining issues and potential problems facing digital libraries, with emphasis on accessibility, integrity, and the interrelationships between them. Lynch discusses the technological and institutional measures that can be used to address access and integrity issues, identifies problems that cannot be appropriately resolved by current technologies and/or institutions, and offers his views concerning actions by Government and others that will be needed to address them.

Roger C. Herdman, Director

## REVIEWERS

Robert J. Aiken
U.S. Department of Energy

Pat Battin
Commission on Preservation and Access

Ann Bishop
University of Illinois

Bonnie C. Carroll
Information International Associates Inc.

Steve Cisler
Apple Computer, Inc.

Nancy Eaton
Iowa State University

Jane C. Ginsburg
Columbia Law School

Jane Bortnick Griffith
Congressional Research Service
Library of Congress

Lee A. Hollaar
University of Utah

Karen Hunter
Elsevier Science Publishing Company, Inc.

Brian Kahin
Harvard University

Brewster Kahle
WAIS, Inc.

Tim Lance
State University of New York at Albany

Jean Pony
NYSERNet

Maria Zemankova
National Science Foundation

*OTA appreciates and is grateful for the valuable assistance and thoughtful critiques provided by these reviewers. They do not, however, necessarily approve, disapprove, or endorse this contractor report.*

# ACCESSIBILITY AND INTEGRITY

# OF

# NETWORKED INFORMATION COLLECTIONS

by

CLIFFORD A. LYNCH

Contractor Report prepared for the
Office of Technology Assessment
Telecommunication and Computing Technologies Program
Joan D. Winston, Contract Coordinator

The views expressed in this contractor report are those
of the author and do not necessarily reflect the analytical
findings of OTA, the Technology Assessment Board,
or individual members thereof.

Accessibility and Integrity of Networked Information Collections
Clifford A. Lynch
July 5, 1993

Table of Contents


1. Introduction . . . . . 1
    Objectives and Purpose of This Paper . . . . . .
    Overview of the Paper. . . . . 3
    Acknowledgments and Disclaimers . . . . . .6
2. The Networked Information Contetex . . . . . .....7
3. An Overview of Access and Integrity Issues in the Networked Information Environment..... . . . . .1 4
4. Access to Electronic information: the Changing Legal Framework . . . . . 17
    From Copyright and Sale to Contract Law . . . . . 17
    Licensing and the Interlibrary Loan System in the United States . . . . . .1 9
    Other implications of the Contract Law Frameworkfor Libraries . . . . . .2 2
    Libraries and Multimedia: New Legal issues . . . . . 2 5
5. The Role of Secondary Information Sources and Automatic Indexing in Access to Electronic Information. . . . . .. .. . . . ..... ...28
6. Access to and integrity of the Historical and Scholarly Record . . . . . 36
7. The Impact of Micropublishing, Narrowcasting, and Information Feeds.....4l
8. Privacy issues in access in the networked information environment . . . . . .4 6
    Confidentiality and Anonymity of Access . . . . . .....46
    Who Owns Access Histories? : Privacy and Market Research . . . . . .5 2
    Privacy, Intellectual Property and Electronic Mail Enabled Communication . . . . . .5 5
9. The Internet Tradition of"Free"Information: Quality, integrity and Liability Issues .....57
10. Access to Electronic Artifacts: The Problems of Versions, Presentation and Standards .....60
    The Nature of Electronic Works . . . . . .. 60
    Presentation and Content Standards .....61
    The Problems of Structured Infomation . . . .. . . . . 6 5
11. Digital images and the integrity of the Visual Record .....66
12. Authenticating versions and sources . . . . . 6 7
13. Citing, identifying and Describing Networked Information Resources .....73
14. Directories and Catalogs of Networked information Resources.....78
15. Conclusions . . . . . 8 1
    Integrity and Access Issues in the Broader Networked Information Context.....81
    Ensuring Access to Information in the Networked Information Environment .....84
    Privacy, Confidentiality and Anonymity in Access to Electronic Information .....86
    Infrastructure and Standards..... . . . . 87
16. Recommendations for Possible Action . . . . . 9 1
    Legislative, Government and Public Policy Actions.....91

Actions by the Stakeholders: Authors, Publishers, Libraries, Information Technology Providers and the Education Community . . . . . 93

Glossay . . . . . .....94
Suggested Background Readings . . . . . .99
Networking Technology, Networks, and the Internet.... . . . . .9 9
Networked Information Services and Networked information . . . . . .9 9
Surveys of Topics Related to this Paper . . . . . 99
References . . . . . .....101

# Accessibility and Integrity of Networked Information Collections

**July 5, 1993**

**Clifford A. Lynch**

This is an Office of Technology Assessment outside contractor report. It does not necessarily reflect the views or findings of the Office of Technology Assessment or the Technology Assessment Board.

## *1.* Introduction

### *Objectives and Purpose of This Paper*

We are entering an age when a great deal of information is available in electronic formats and can be obtained through computer communications networks. Sometimes, such collections of network-accessible electronic information are referred to as "digital libraries" (although, as discussed later, I view this terminology as somewhat misleading; indeed, one of the issues explored here is the developing roles of such electronic information collections and their relationships to institutions such as libraries). These new electronic information resources are still being defined and almost everything about them is poorly understood; conventions, expectations, social customs, institutional roles in society, business relationships, legal precedents and public policy choices based on print and mass media broadcast environments may no longer be directly applicable in the networked information environment. This new environment challenges us to reconsider assumptions and to rethink many aspects of the current systems of information creation, management, organization, preservation and use, as well as the broader public policy context of these systems.

This paper, prepared under contract to the United States Congress Office of Technology Assessment, is intended to consider questions related to the integrity and accessibility of these new electronic information resources. These are key areas that pervade many of our assumptions about all forms of information; the resolution of integrity and access questions in the electronic information environment will play a major role in defining how we create, use and manage these new forms of information and the economic system that will support these activities, the purposes it will serve within our society, and the functions that various institutions will need to carry out in this new environment. Both integrity and access are complex issues, and have intricate

relationships not only to each other but to legal and public policy considerations such as privacy and intellectual property. Further, integrity and access cannot be considered outside of the context of the communities and institutions that create, distribute, organize, manage and use information; this leads to a discussion of the evolving roles of libraries, publishers, and the authors and readers of various types of material. Finally, because the growing flood of available information is increasingly overwhelming and unmanageable without mechanisms and conventions to allow us to identify, select, and cite particular works, I have devoted considerable space to exploring how these mechanisms are developing in the world of electronic information resources.

While there is some emphasis here on printed information and its electronic descendants, and particularly scholarly communication and publication, i have tried to at least begin an examination of the broader issues related to mass communications, newspapers and news broadcasts, government information, and new multimedia information resources that might be of interest to the general public rather than just the research and higher education communities. The reader should recognize that my comments in these areas are at times rather speculative; the research and higher education communities have been early adopters of networked information, and while even within those communities the issues are far from resolved, we have much more experience with scholarly information than we do with mass market network-accessible information.

To a great extent, this paper presents a personal view of the developing issues. My background is large research libraries and the higher education community. Other stakeholders in the areas discussed here, such as publishers or public librarians, may well have other views, and certainly even within my own communities there are many different opinions not only on what should be done but even the significance of many of the trends and developments discussed here. While I have attempted in various places in this paper to at least indicate the existence of other viewpoints and sometimes to sketch the rationale behind them, the reader should not view this paper as a fully balanced survey and synthesis of the various viewpoints on the issues. Finally, I must stress that I am not an attorney, and thus my comments on legal issues should be viewed as probably better informed about the implications of current legislation and case law on systems of information creation, dissemination, preservation and use than on the legal specifics themselves. While this paper has benefited greatly from reviewers more learned in the law, I may well have misunderstood or overlooked specific legal issues despite their efforts to help.

I have not provided an explicit executive summary of this paper. Those interested in obtaining a quick overview of the paper's coverage and conclusions should read the subsection directly following this one which gives an overview of the paper, and then proceed to Sections 15 and 16, the conclusions and recommendations for possible follow-on actions. These proposals are somewhat limited in nature; my purpose in this paper is primarily to illuminate the relationships among technological developments, electronic information, legislation and public policy, and various institutions such as libraries, and to identify and define the issues that new technologies are bringing to the fore. In a number of areas I have suggested consideration of a review and reexamination of current public policies, legislative positions, and government and private sector investment decisions in light of the trends described here; but in cases of public policy and legislation I have focused on providing information that could help to

inform such a review and reexamination rather than attempting to suggest specific outcomes. I have mentioned a few specific areas where I felt that there was a need for funding, to move the implementation of existing policy directions.

### *Overview of the Paper*

The paper begins with a survey of recent developments in networked information resources and tools to identify, navigate, and use such resources (Section 2). This is a vital context and source of examples for the remainder of the report. As these new information resources are surveyed, this section also examines the idea of "digital libraries" and the relationship between today's libraries as organizations and collections of electronic information. Those readers unfamiliar with the rather extensive and quickly evolving developments in networked information resources may find this section of the paper particularly difficult reading, heavy with acronyms and references to a bewildering array of organizations and systems. To help those readers, I have provided a brief glossary of terms that I felt might be unfamiliar and a short annotated bibliography of useful background readings (as distinct from specific references cited in the text).

Building on the context established in Section 2, Section 3 provides an overview of the issues involved in access and integrity questions and the relationships among them. The boundaries of the paper are also largely defined here: for example, issues of network access (as opposed to access to information on the network) are excluded, as are most of the public policy issues specific to government mandates to provide access to government information. If access is defined somewhat narrowly, integrity is treated quite broadly and encompasses not only conventions and audit trails needed to ensure consistency and accountability in the scholarly, historical and cultural records, but also questions of content bias and accuracy. The links between access and integrity are stressed: for example, ensuring the integrity of a corpus of publication is meaningless if there is no access to this body of information.

The paper then discusses the changing legal framework that governs use of electronic information as contract law rather than simple sale within the context of copyright law becomes the dominant model for acquiring access to electronic information (Section 4). This shift is shown to have profound implications for access to information and also for the integrity of the historical, scholarly and cultural record that "published" information has traditionally represented. The effects of this shift on libraries and the interlibrary loan system, which has historically been a key part of the library community's strategy for providing their patrons with very broad access to information, is examined in depth. This is followed by an exploration of the new and even more complex questions raised by image and multimedia content in electronic networks; here we start from an environment of much greater ambiguity when applying copyright law, and find both libraries and rightsholders facing substantial problems in understanding how to manage these materials. Because of this, we find that the shift to contract **law** offers stronger operational benefits for those institutions (including libraries) that want to acquire access to image and multimedia information, although this shift again raises serious public policy issues.

The role of secondary information sources in providing access to, and evaluation of, electronic information is examined from several perspectives in Section **5;** these include the role of the extensive and well-developed marketplace that exists today in databases (for example, so-called abstracting and indexing databases) and other tools that provide access to the print literature (and in future to electronic source material) and the potential impact of new tools derived from technologies such as automatic indexing. Appropriate application scenarios for the different access tools are discussed, as is the growing power of these access tools in defining the part of the "published literature" that will be read and evaluated by various communities. This power places great responsibility for quality and integrity on the secondary information providers and access tools and thus plays a significant role in establishing the overall accessibility and integrity of the published literature. Considerable attention is given to the implications and impact of quality and integrity problems in this area.

The paper then turns to the central issues of the historical and scholarly record and the access and integrity questions that surround it as much of the information that comprises this record shifts to electronic forms (Section 6). Much of the theme here revolves around the multiple uses to which this record is put, the different expectations and requirements of the various communities that rely on this record, and the social, business, legal and technical issues involved in trying to address these somewhat conflicting needs. The implications of the shift from sale under copyright to contract law based licensing emerge clearly here as a potentially serious threat to both access and integrity for this record, and help to frame the public policy issues involved in ensuring its preservation.

A series of technology developments, economic factors and market demands have led to the creation of ever more specialized publications; in the print world this is often termed "micropublishing" while in the broadcast industries it is sometimes referred to as "narrowcasting. " The trend towards increasingly personalized information products, both in print and digital forms, combined with the appearance of new electronic information sources such as news feeds or sensor data streams that can essentially be "personalized" by the recipient (or the recipient's organization) using digital technology again raise serious challenges to the accessibility and integrity of the electronic record. They create enormous problems for libraries as the traditional managers of this record. And they raise new, complex public policy problems related to equality of access to information sources, and the implications of inequitable access by various sectors of the population. Section 7 explores this area.

The relationships between privacy and access in the electronic environment are very complex. Do users have a right to anonymous access to information resources, or at least a reasonable expectation that records of their information seeking and use should be kept confidential? Complicating these questions are conflicting requirements for accounting, cost recovery, system management and tuning, ensuring the security and integrity of the underlying computer networks, and the economic motivations of various parties to collect usage and demographic data for potential resale or reuse that run counter to the long standing policies of confidentiality and anonymity that have been established within the library community. At the same time, technological developments are in some cases preempting (at least in the near term) the as yet unresolved policy debates. Section 8, on privacy and access attempts to summarize the issues in these areas and to reflect some of the competing pressures.

Much of the information currently available through computer networks is "free"; that is, the user is not charged for using it. Section 9 examines some of the implications of free information, such as expectations about accuracy, timeliness, and value. In a real sense, the extensive reliance upon free information sources is shown to add a certain instability to the networked information environment because of the ease with which inaccurate information can quickly spread with little accountability. While the public policy questions here seem to be rather limited, this is important to developing a full picture of the networked information environment.

The paper then considers the nature of electronic works and how access to such works can be provided in Section 10. Here, one major theme is the current tendency to intertwine content and access mechanism. This has serious implications for long-term access and preservation of these works as the information technology environment continues to change. There are also subtle integrity issues that arise as we attempt to define the distinctions between content and presentation of that content. This section also emphasizes the importance of establishing and encouraging the widespread adoption of appropriate standards that allow information providers and users to separate content, presentation, and access or viewing tools for electronic works.

While we realize intellectually that photographs can be altered, the visual evidence provided by photography has provided a very important part of the historical record in our society. We have made very strong intuitive assumptions about the integrity of photography. The section on digital imaging and the integrity of the visual record (Section 11) summarizes how the development of new digital imaging technologies calls this role into question and places much greater emphasis not on the image as artifact, but on the use of verified true versions of images combined with formal, auditable links from that image to the individual or organization that has created it. This serves as motivation for the Section 12, which deals with the authentication or verification of electronic works and their creators. This requirement is a cornerstone of the integrity of electronic information; while perhaps most visible in the context of images, it pervades the use of all types of electronic information. A discussion of the issues involved in making such authentication possible leads directly to a series of issues concerning cryptographic technology, including standards, intellectual property rights and export controls (which in turn are related to the search for an appropriate balance between national security concerns and the needs for privacy and authentication in the networked environment).

The final sections of the paper consider two related issues that are again central to both access and integrity in the electronic information environment. The first, covered in Section 13, has to do with identifying and citing electronic works, and summarizes requirements, architectural approaches and standards developments to address these needs. The second issue is the intellectual identification of networked information resources; here the development of catalogs and directories for these resources is considered, with some emphasis on the role of libraries on the one hand in creating these directories and catalogs and, conversely, the way in which such directories and catalogs will integrate with existing tools used by libraries to provide access to the print literature during the long transitional period where both electronic and print information are essential parts of the scholarly communication system and co-exist. Section 14 addresses these questions.

The paper divides its conclusions into two parts The first concluding section (Section 15) tries to summarize and tie together the various developments, trends and issues that have been surveyed, and also to set the conclusions of this paper in a broader context. One point that is emphasized in this section is that two of the key areas slowing progress in infrastructure development for networked information—standards development and the deployment of integrity and authentication support services based on cryptographic technologies—call for issues well outside the primary scope of this report to be addressed, but until these issues are addressed, will continue to cause problems in our ability to resolve access and integrity questions related to networked information. Section 16 builds on and Section 15 and enumerates specific issues and projects where actions—by Congress, by various government agencies, or by various groups within the stakeholder community—should, in my view, be considered.

### *Acknowledgments and Disclaimers*

This paper has benefited greatly from the careful work of a large number of reviewers some known to me and others anonymous. Their comments greatly improved the quality and accuracy of the work. I thank them for their efforts and their thoughts, while taking all responsibility for errors and for opinions with which they may well disagree. I owe particular thanks to Joan Winston of the Office of Technology Assessment for multiple reasons: for inviting me to explore these extremely interesting issues in the first place; for a very careful and thoughtful review of a draft of the paper; and for her patience as deadlines slipped and the paper continued to grow and evolve at its own pace towards completion. Conversations with any number of colleagues, both in person and electronically, have been of immeasurable help in understanding these issues. Particular thanks are due Steve Cisler, Michael Buckland and Howard Besser. Various electronic newsgroups and discussion lists have also been very valuable to me in preparing this paper, both in terms of discussions (where typically I listened much more than I contributed) and by providing direct object lessons about several of the topics discussed in the paper. I would also like to thank Cecilia Preston both for a careful review of a draft of this paper and for her help with the citations and references.

References and citations for a paper of this type are a huge problem. I have not even tried to be comprehensive in the citations; while a comprehensive bibliography of developments in networked information would be extremely useful, compiling it would be a truly enormous effort, and, given the rate of developments in this area, it would need continual updating. Worse, a number of the projects have not yet been much described in the published literature, and what material exists is scattered across obscure conference and workshop proceedings and technical reports. Thus, I have had to satisfy myself with the inclusion of a limited number of references suggestive of the type of work going on today, particularly in Section 2 on the networked information context, and I would beg the indulgence of those authors who have made important contributions to the literature in this area but do not find their efforts reflected in the citations here. Some readers may find the bibliography in [Lynch & Preston, 1990] a helpful supplementary source. In sections this paper touches on other large and complex areas, such as the technology of cryptography; for basic citations in this area (as well as a very readable overview of some of the technology) the reader should see [U.S. Congress Office of Technology Assessment, 1987]. Similarly, I have not provided

much coverage of the literature on intellectual property and copyright issues; an excellent source for these is [U.S. Congress Office of Technology Assessment, 1986]

The opinions expressed here are mine and do not necessarily reflect the views of any other organization or individual. A number of products and projects are mentioned here as examples; these mentions should not be interpreted as endorsements of products. Finally, a large number of trademarks and service marks are mentioned in the text. In particular: UNIX is a registered trademark of Bell Labs; Macintosh is registered to Apple Computers, as is Quicktime. MELVYL is registered to the Regents of the University of California. PostScript, SuperATM and Acrobat are trademarks of Adobe. I apologize in advance for those that I have failed to mention here.

## 2. The Networked Information Context

As use of the Internet becomes more widespread, it is becoming clear that access to information resources through the network will become one of the central applications of the network. The term "access," here and *throughout this* paper, is *used in the broadest sense: not simply electronic connectivity to information resource providers through the network, but the ability for people to successfully locate, retrieve and use the information contained within various computer systems.* For a large segment of the Internet user community, such access to information and network-based communications tools such as electronic mail will increasingly be the applications which initially attract users to the Internet and subsequently are most heavily used. Further, we are already seeing both the development of a series of information discovery and navigation tools aimed at the end user and the evolution of a set of services that combine the access and communications capabilities offered by the network into new hybrid information services that are more interactive and immediate than traditional print or broadcast media. Libraries, publishers, and government organizations that, create, collect and provide access to information, are all striving to define and understand their possible new roles in this developing networked information environment.

Discussions about the development of the Internet and the evolution of the National Research and Education Network (NREN) program are increasingly taking a broader view that emphasizes not only the continued expansion and upgrading of the technical communications infrastructure but also the need for policies and planning to foster the development of networked information resources and the tools and educational programs needed to allow network users to create, navigate, and utilize such resources. Another component of this shift of focus is the recognition of the need to transform existing printed and electronic information resources into network-accessible forms. As we look beyond the NREN program, which is targeted to serve the research, education, library and government communities (broadly defined) towards the discussions about a full National Information Infrastructure (Nil) it is clear that electronic information content on future networks will play an increasingly dominant and driving role in network evolution.

This shifting focus is evident, for example, in a wide range of bills before the 1993 Congress, including S4, The National Competitiveness Act of 1993 (Hollings), S626, The Electronic Library Act of 1993 (Kerrey), HR1757, The High Performance

Computing and High Speed Networking Applications Act of 1993 (Boucher) and HR1328, The Government Printing Office Electronic Information Access Enhancement Act of 1993,[1] to name some of the most major. Various federal agency based programs (for example, at NASA, the National Agriculture Library, the Library of Congress, and the National Library of Medicine) are also underway to foster the availability of networked information resources.[2] The recent revision of the Office of Management and Budget Circular A-130 and its associated guidance to government organizations also speaks to the need to make information available in electronic form. Paralleling these activities at the federal level is a growing interest on the part of some state governments in the potential of the network to provide improved access to state, regional and local information (for example, in Texas [Stout, 1992]. Colorado, Utah,[3] California and North Carolina). State libraries are using the Internet as a focus for statewide multitype library planning in several states, including Colorado [Mitchell & Saunders, 1991] and Maryland.[5]

Broader based initiatives that span the government, research and education, and commercial sectors recognize networked access to information resources as a key element. For example, the National Science Foundation (NSF) sponsored Synthesis Coalition, which is focused on improving engineering education at all levels, includes a component called NEEDS (the National Engineering Education Delivery System) which focuses specifically on the creation of networked information resources in support of elements of the Synthesis program [Saylor, 1992], The Council on Library Resources is examining how to improve access to engineering information [Council on Library Resources, 1990; Council on Library Resources, 1992]; here, again, network-based access to information is viewed as playing a key role. Major scientific programs in areas ranging from Molecular Biology to Global Climate Change all view the development and operation of networked databases as essential program components [Olson, 1993; Stonebraker, 1992]. The higher education and research library communities have created the Coalition for Networked Information (CNI), a joint project of CAUSE, EDUCOM and the Association of Research Libraries (ARL) to specifically address

---

[1]This was signed into law by President Clinton on June 8, 1993 as Public Law 103-40

[2]While agencies such as NASA seem to have made a commitment to the **networked** information model of access to agency information, other groups—for example, the **Bureau of the Census—have addressed the distribution of government information through the publication** of CD-ROM databases, leaving it to the user communities (i.e. the universities or the Federal depository libraries) to determine how the information they publish on CD-ROM should be made generally accessible through the network. At present most CD-ROM databases are unsuitable for use in the networked information environment, despite efforts by group such as the Air Transport Association and the Intelligence Community Data Handling Committee to define network interfaces to CD-ROM databases [Bowers & Shapiro, 1992]. In still other cases federal agencies such as the SEC have formed alliances with the private section (Mead Data Corp., in the case of the SEC) to offer access to federal information through the networks [Love, 1993]. There is a great need for more consistent policies for access to government information through the networks.

[3]Utah makes legislative information available through the Internet.

[4]In California, Assembly Bill 1624, currently under consideration in the state legislature, would make legislative information available through network access.

[5]The Maryland State Librarian is leading a major effort to link libraries throughout the state using the Internet. Similar projects are under discussion in Virginia and other states, [Library of Congress Network Advisory Committee, 1992].

networked information issues [Peters, 1992a], as well as devoting substantial attention to broader networking initiatives within the programs of the parent organizations (for example, EDUCOM'S Networking and Telecommunications Task Force); the CNI initiative reaches out beyond the research and education community to reach providers of networks and information technology, publishers, and even, to a limited extend information users in the private sector.

There is a great deal of talk about "digital libraries", "electronic library collections", "electronic libraries", "electronic journals" and "network-accessible library collections"; such visions have captured the imagination of many scholars, educators, librarians, and policy-makers [Lynch, 1991a], and are increasingly attracting the interests of the commercial sector—particularly publishers, mass media, and other information providers—as potentially lucrative new markets. Indeed, the upcoming transition to electronic information resources is viewed hopefully by some as a means of relieving increasingly intolerable financial pressures on the entire system of scholarly publishing.[6] Yet the definition of a digital library remains controversial. Personally, I prefer to consider the viewpoint that stresses electronic library collections; a library is an organizaiion that acquires, structures, provides access to, and preserves information of all kinds, and within this context network-based electronic information is just another type of material that the library can acquire and manage (albeit one with many unique, intriguing and novel properties). Additionally, the availability of digital content and sophisticated information technology of course permit the library as an organization to offer new organizational and access services, and to move from a primarily "passive" organization to one that actively provides information to its users through interactive services and automated, network-based, content-driven information delivery.

When we speak of digital libraries, however, we invoke not only this concept of electronic library collections but also visions of the integration of networked information resources of all kinds (including, for example, numeric databases that libraries have typically neglected and remote sensing databases that are collected as part of various scientific programs outside of the library context) into new collaborative environments (co-laboratories) [Lederberg & Uncapher, 1989]; some term such collections of databases and related networked information resources to be digital libraries. There is discussion of coupling information technology with a new partnership among researchers, information management specialists and information technology experts

6 The primary source of these pressures is that libraries can afford to purchase a smaller and smaller part of the annual output of scholarly books and journals worldwide. The roots of this crisis are complex and form the subject of extensive debate between librarians, publishers, and academics. Many librarians argue that the prices for these materials are rising far in excess of the rate of inflation, in part due to price gouging by the publishers. The publishers submit that the size of the body of annual research is growing rapidly, and that library funding has not kept up with this rate of growth; they also identify other factors such as international currency fluctuations that have contributed to the problem in recent years. For a survey of some of these issues see [Grycz 1992]. A discussion of these issues is outside the scope of this paper; however, it is important to recognize that conversion of scholarly materials to electronic formats may reduce printing, distribution, handling and storage costs for these materials somewhat, but will generate new costs in retooling editorial and production processes and in investments in information technology and infrastructure. Overall, it is not clear that conversion to electronic formats will substantially reduce costs for scholarly materials, though it will undoubtedly redistribute these costs. Further, if, as some librarians argue, much of the problem is publisher profiteering, a shift to electronic scholarly publications will only alter the economics if it causes substantial changes in the system and the role of publishers—particularly commercial publishers—within that system.

under a model called "knowledge management" that relies heavily on networked information resources to directly support and integrate with the research process itself as well as the diffusion of new knowledge to the research community and the dissemination of research results [Lucier, 1990; Lucier, 1992]. I view these new networked information resources to be something fundamentally new, and different from library collections (though they might, in some cases, be part of a library's collection, or a part of the services offered by a library); certainly they are different from libraries (as organizations), though in some cases libraries may be the organizations that create, manage or fund these new networked information resources. We will need a new terminology and taxonomy for the networked information age. But, in any event, the focus of this paper will be collections of network-accessible information resources and the roles of libraries in maintaining and providing access to them.

There are a vast number of experiments underway at present to use the network to deliver or provide access to bitmapped images of print publications, including document delivery services being offered by CARL, Engineering Index, University Microfilms, Faxon and others, often in complex business partnerships with traditional secondary (i.e. bibliographic) database access providers such as OCLC, RLG, or Dialog. Primary scientific journal publishers such as Elsevier and Springer-Verlag are conducting experiments with various universities under which they are making the contents of some of their journals available electronically to institutions either under site license or pay per view arrangements. In addition, various third-party aggregators and relicenses such as UMI and Information Access Corporation are now licensing full text or page images of journal contents for sizable collections of journals in specific areas directly to institutions, and a number of publishers are making the text of their newspapers, magazines, or other publications available for searching through database access providers such as Dialog or BRS on a transactional basis.

Indigenous network-based electronic journals are now well established, and their number continues to grow (though it is important to recognize that they are still a very minor force, compared to the existing print journal system, in most disciplines). The vast majority of these are made available for free on the Internet, and they include both journals structured in analogy to peer-reviewed print journals, such as Postmoderm *Culture* (which Oxford University Press has recently agreed to market in a parallel print version), *Psycology, Public Access Computer Systems Review* and many others (see Michael Strangelove's bibliography [Strangelove, 1993] and other regular publications that are similar to edited newsletters *(Newsletter on* Setials *Pricing Issues, TidBits,* etc.). The edited newsletters are part of a continuum that fades off into "moderated discussion lists" implemented through LISTSERVERs or other mail reflectors, which might be viewed as continuously-published electronic newsletters that exploit the electronic medium of the network to avoid the need to gather submissions and commentary together into discrete issues. There are also thousands of unmoderated discussion groups which provide additional forums for discussion and information interchange; these have some elements in common with newsletters or other publications, but are really a new and unique artifact of the networked environment. Recently, a few for-fee journals have begun to publish either solely in electronic form (i.e. the OCLC/AAAS *Current Clinical* Trials experiment, a fully peer-reviewed journal [Keyhani, 1993; Palca, 1991], or the *Computist's Communiquè* by Ken Laws, more of a newsletter), or in parallel print and electronic editions (e.g. John Quarterman's Matrix

News). It seems probable that the development of for-fee journals on the network has been inhibited by publisher concerns about the acceptable use policies that govern traffic on much of the Internet; as it appears that the acceptable use policy may well either be abandoned or interpreted liberally enough to comfortably accommodate this type of network based publication and publishers can find other publishers distributing journals in the network environment without problems, the number of for-fee journals will grow rapidly.

Paralleling these initiatives in the creation of content, a great deal of effort is being devoted to the development of tools for network navigation and information retrieval. The development of standards for resource description, location, and access in a distributed environment are also a key part of the development of the tools themselves. Major efforts in this area include the Gopher project at the University of Minnesota [Alberti, Anklesaria, Linder, MaCahill, & Torrey, 1992; Wiggens, 1993], the World Wide Web system from CERN [Berners-Lee, Cailliau, Groff, & Pollermann, 1992], the WAIS system that was originally developed as joint project of Thinking Machines, Apple, Dow Jones and KPMG which is now being carried forward by a number of organizations, including the new NSF-funded Clearinghouse for Networked Information Discovery and Retrieval in North Carolina [Kahle, Morris, Davis, & Tiene, 1992a], the archie system developed at McGill University [Emtage & Deutsch, 1991; Simmonds, 1993], the resource discovery work carried out by Mike Schwartz and his colleagues at the University of Colorado, and others [Schwartz, Emtage, Kahle, & Neuman, 1992; Schwartz, 1989; Schwartz, Hardy, Heinzman, & Hirschowitz, 1991]. Recently, the National Science Foundation awarded a sizable contract to AT&T for the development of directories for the Internet; while this contract is primarily to compile and operate such a directory using existing technologies and standards, and the resource directory being developed does not seem to incorporate any sophisticated classification or retrieval approaches, the AT&T effect is likely to focus and stimulate further effort in this area. More research-oriented work is also underway in developing cataloging and directory tools through the CNI Topnode project,[7] the Department of Education funded project for cataloging Internet resources at OCLC [Dillon, 1993], the work of the Library of Congress on cataloging practices for electronic information resources [Library of Congress, 1991a; Library of Congress, 1991b; Library of Congress, 1993], and the efforts of various working groups within the Internet Engineering Task Force on document location and identification.[8] Other important standards work includes activities such as the Z39.50 Implementor's Group, which is addressing both functional extensions to the Z39.50 computer-to-computer information retrieval protocol, a basic tool for searching information resources on the network, and also attempting to resolve interoperability issues as the Z39.50 protocol moves toward widespread implementation [Lynch, 1991b; Lynch, Hinnebusch, Peters, & McCallum, 1990]. In addition, of course, standards developed within broader communities to describe various types of record and document interchange formats are of critical importance to the development of networked information retrieval tools.

---

[7] Information on the current status of this project can be obtained from the Coalition for Networked Information, or by searching CNI's file server at ftp. cni.org.

[8] While not much has been published on this yet, the IETF should be issuing a series of draft RFCS within the next six months. The general approach that is being pursued is described in Section 13 of this paper.

Institutionally based projects at universities such as Carnegie-Mellon (Project Mercury) [Arms, 1992], Cornell (various projects) [Lesk, 1991], the University of California (various projects) [Lynch, 1989; Lynch, 1992]. The University of Southern California Watkins, 1991] have focused on developing systems for the storage and delivery of electronic information resources to institutional user communities, in some cases integrating and building upon tools and standards developed on a national and international level. Some other projects, notably at Yale, Cornell, Barry Shein's Online Book Initiative (hosted at world. std.com) and Michael Hart's Project Gutenberg are working with public domain collections (either out of copyright materials or government material not subject to copyright) as prototypical electronic library collections. In some ways, these out of copyright collections are very attractive test cases as they permit the institution to focus on the technology and end user requirements of the application without becoming enmeshed in economic and legal (intellectual property) concerns.

Not all work on networked information access is rooted in the higher education and library communities or the efforts of commercial firms that primarily serve these communities. For example, a number of communities have deployed versions of the Freenet system developed by Tom Grunder in Cleveland, Ohio [Grunder, 1992; Watkins, 1992]. This is a system which provides the electronic analog of a town, complete with a town hall, libraries, schools, medical facilities, discussion groups, and other areas. While some implementations have been supported in part by libraries and/or universities, Freenets may equally well be sponsored by municipal governments or private citizen groups outside of the higher education and research communities. In addition, commercial services such as CompuServe and America Online are now well established and offer access to sizable collections of information; their primary user communities are outside of the academic or library worlds.

Recently, Carl Malamud established a project called Internet Talk Radio which offers a mixture of interviews and live coverage of speeches and other newsworthy events; the content is somewhat similar to that of the C-SPAN cable network, although it includes announcements from commercial underwriters similar to those found on Public Television (not really full scale advertising by sponsors) [Markoff, 1993]. Internet Talk Radio has coverage into the National Press Club in Washington, DC and is scheduled to "broadcast" its first coverage of a congressional hearing later this summer, Internet

---

[9]It should be noted that, while out of copyright material will be a very important resource for libraries that wish to explore the electronic storage and dissemination of material exactly because this material is not subject to copyright constraints, such material is substantially difficult to identify; worse, the identification of such material is growing more complex as the issues are explored in more depth. Consider first simple textual materials. In the US, currently, the period during which a work is subject to copyright is defined by the life of the author plus a certain number of years, rather than the old definition which was just a fixed number of years from the date of copyright. This means that a library that wants to determine whether a work is still under copyright protection must attempt to determine whether the work's author is still living or when he or she died. This is a major research problem. Further, international copyright issues have become extremely complex. For example, Project Gutenberg recently made a copy of Peter Pan available, since the work appeared to be out of copyright in the US, only to subsequently discover that there is apparently a special exemption for this work under UK copyright law that permanently assigns copyright protection to this work and donates the proceeds to a hospital in the UK, and thus the electronic text could not be distributed in the UK legally. How this strange exemption in UK copyright law relates to the Berne Convention and the internationally reciprocal copyright agreements that the US has agreed to honor is a subject for legal scholars that I will not speculate upon here; however, it is a good illustration of the problems of identifying material that is no longer subject to copyright.

Talk Radio captures audio from these events and distributes it over the Internet in real time using multicast technology to interested listeners; in addition, digital files containing the audio for the broadcast events are archived on the network and can be retrieved and played by individuals with appropriate audio boards in their machines at any time on demand. Even for simple audio, these files are quite large and stretch the capabilities of many machines on the net. The Internet Engineering Task Force has been experimenting with digital audio and video distribution of key sessions at its meetings using similar multicasting technology, though this is considerably more taxing for the network due to the data volumes involved. As the network capacity expands and the supporting capture, playback and multicasting technologies become more mature and more widely available it seems likely that this type of audio and video coverage of events of interest, both multicast real time and stored in archives for later demand playback, will become more commonplace.

Yet, despite this rich and wide-reaching series of projects (and what has been described here is only a sampling intended to give the reader a sense of current developments) which we hope will yield knowledge and insight that will inform future efforts,[10]" there is little consensus about the future structure of electronic libraries, digital libraries, network-accessible libraries or whatever one wants to call them-or even if these terms refer to the same things. Some people refer to collections of network-accessible files as a digital library; this is common in some parts of the computer science community, for example. Some from the publishing community speak of digital libraries when a perhaps more accurate term might be a digital bookstore or information store. Those viewing the evolution of electronic information resources from the library tradition tend to think of networked information as simply one more component of a traditional library's collection, subject to the basic library functions of selection, organization, provision of access, and preservation, suitably adapted for the unique characteristics of the network environment (for example, you can select a network resource that you provide access to without physically making it a part of a given library's collection—in other words, performing acquisitions without taking physical possession, as distinct to providing some form of subsidized access to a resource that the library continues to regard as "external" and available through mechanisms such as interlibray loan or short term contract to subsidize access in the networked environment). Indeed, with the network's ability to dissolve geographic boundaries and unify access to autonomous, physically and organizationally distinct resources, fundamental questions are being raised about the nature of these future electronic information collections—for example, might there just be one logical "library" of information for each discipline [Loken, 1990], perhaps managed by a professional society, in the sense that the user would communicate with only a single access-providing organization for the discipline's literature?

---

xtent to which current prototypes will in fact help to resolve the open questions is problematic. Many of the prototypes are being rushed into production use, without any systematic evaluation of the human or economic outcomes. Too often there is funding to build prototypes, but no funding to evaluate them rigorously. In some cases the economic viability of projects beyond the prototype state is unclear, and there is a real lack of common economic models that permit comparisons to be drawn between projects. The definitional difficulties surrounding the concept of "digital libraries" are indicative of the severity of this problem.

## 3. An Overview of Access and Integrity Issues in the Networked Information Environment

The institutions that are libraries—be they public libraries or research libraries—have addressed a number of concerns about the accessibility and integrity of printed information that arise from diverse quarters ranging from the needs of the academic community to manage and provide access to the scholarly record through the needs of the government to ensure the existence of an informed citizenry with access to vital government information resources. Libraries ensure a certain base level of access to information irrelevant of the financial status of the information seeker. Many of these concerns do not have well-established, clearly defined constituencies or clearly stated requirements. But the concerns are nonetheless real, and of vital importance to our nation and our society. As the nature of the information changes from printed pages to network-accessible information resources, we can no longer assume that old mechanisms will continue to work, or that they will be adequate to address the full range of new issues that are raised by electronic information resources; indeed, we do not yet fully understand the implications of a large scale shift to electronic information or the new roles that we will expect libraries to undertake in this context.

This paper examines a series of specific issues related to the access and integrity of electronic information in a networked environment, and current and potential roles that libraries and other institutions may play in addressing these issues. It also explores the ways in which the transition to the networked information environment may call existing library practices and roles into question.

Access to information in a networked environment is an area that is often treated very superficially. There is a tendency to incorrectly equate access to the network with access to information; part of this is a legacy from the early focus on communications infrastructure rather than network content. Another part is the fact that traditionally the vast bulk of information on the Internet has been publicly accessible if one could simply obtain access to the Internet itself, figure out how to use it, and figure out where to locate the information you wanted. As proprietary information becomes accessible on the Internet on a large scale, this will change drastically. In my view, access to the network will become commonplace over the next decade or so, much as access to the public switched telephone network is relatively ubiquitous today. But in the new "information age" information will not necessarily be readily accessible or affordable; indeed, if information is to become the new coin of the realm, there is no doubt in my mind that there will still be the rich and the impoverished-though even the impoverished may have a relatively high standard of access to information, compared to today's information poor in the US, or tomorrow's information poor globally .11 This

---

11 The information **poor** are not the same as the illiterate. The illiterate are a group that lack specific training and skills increasingly essential in modern society; the information poor may not necessarily lack the skills to find or comprehend the information they need, but rather may simply not be able to afford to pay for access to information. While illiteracy is a problem that is often the result *of* poverty, it is really the lack of a specific skill. Lack of access to information is a condition that is created by at least in part by **poverty** (and really more generally by a gap between the price of information access and the economic conditions of the person who needs to obtain access to information), but which is at least to some extent rectified by subsidizing information access, as opposed to illiteracy, which usually is not the result of inability of the illiterate to obtain access to printed material. Information poverty is a mix of two factors:

paper will focus on information access issues and largely omit issues related to base network access. Access here will also be viewed in the broad sense; not only considering who can have access and how much they must pay, but when they can have access, and who knows what they are accessing.

Integrity of electronic information is another problematic area. Many people have a bias that leads them to view electronic information as less stable than printed information-electronic information is subject to unannounced revision by insidious parties, corruption by viruses unleashed on the network or by individuals breaking into computer systems. In fact, the issues here are extremely complex, ranging from the balancing of the ability of the network to support access to the most current information against the need to maintain a trail of citeable versions linked to specific points in time, through questions of long term preservation of digital information. It is interesting to note in this connection that many of our expectations about networked electronic information are derived from our experience with, and expectations about, print information, and that in fact we regularly accept completely different rules for broadcast mass media information than we do for print; similarly, much of the legal framework for electronic information (with the exception of some very specific counterexamples, such as integrated circuit masks) also has its basis in practices related to print materials. (It is also worth noting that most libraries have tended to avoid becoming much involved with providing access to the contents of broadcast mass media). Other issues in this area include the problems of intellectual property, hidden bias of many different types, and loss of information diversity available to the public. Our expectations about the integrity of electronic information are unclear; in fact, these expectations vary depending on the use we wish to make of a given electronic information resource.

Integrity and access are interrelated in complex ways; in the evolving context of networked information, the relationship is far more complex than in the existing print-based environment. In the electronic environment the balance of relationships between the creators of information, the distributors and rights holders (publishers), the stewards (libraries) and the consumers of information seem to be changing radically. Within the print literature framework each of these stakeholders had well-established roles in ensuring integrity and providing access; with the shift in relationships and responsibilities, these roles will also change. Access to electronic information is of questionable value if the integrity of that information is seriously compromised; indeed, access to inaccurate information, or even deliberate misinformation, may be worse than no access at all, particularly for the naive user who is not inclined to question the information that the new electronic infrastructure is offering. Further, certain characteristics of the mechanics of accessing electronic information in the networked environment may lead to new means of comprising the integrity of that information.

---

inability to afford access to information and lack of skills to obtain, navigate, and evaluate information. One might argue, for example, that many scientists and engineers are in fact information illiterate, although they can certainly afford access to substantial bodies of information; they lack the skills to utilize this body of information. Further, literacy, in an world that is increasingly full of multimedia information, may not be always be a prerequisite to being able to understand information once one obtains access to it. Indeed, in the future, the relationships between literacy, having the skills necessary to locate needed information and/or access to trained intermediaries such as librarians who can help to locate information, and having the ability to afford access to information (either by paying for it directly or thorough organizations like libraries that have historically subsidized access to information) are going to become much more complex, and deserve new attention in the context of the coming age of electronic information.

Conversely, even if the integrity of the scholarly or cultural record is guaranteed, such a guarantee is of limited value without a corresponding guarantee of continued access to such a record.

In discussing issues of access and integrity in networked information today, there is a very strong bias towards issues specific to scholarly information; this is to be expected, given that the academic and research communities have up until now been the primary constituencies on the Internet. These are relatively sophisticated communities, and communities with an ethos that is strongly oriented towards citation, attribution, and preservation of the scholarly record. Indeed, as one reviewer noted, this ethos ties these communities to the system of print publication, and emphasizes that networked information must offer the same integrity and access if it is to become an acceptable replacement for print. Scholars must be certain that their work is accessible to their peers and that the integrity of their works is maintained. Yet if one examines the current growth of the Internet, the fastest growing sector is commercial organizations. Primary and secondary education and public libraries are one of the major potential growth sectors (if funding strategies can be found to pay for their connection and for the associated information technology and staff training investments that will have to be made within the institutions themselves). There is now discussion about the role of the Internet as a precursor to a National Information Infrastructure (which really might be more appropriately called a National Networking Infrastructure) which would connect an even broader constituency. As the networked community expands, we will see a continued shift in expectations and values about information access, integrity and management, and the appearance of new types of information that have much more in common with the traditional contents of print and electronic mass media today than the bulk of the information that populates the current Internet. This paper thus attempts to take a broader view of the access and integrity issues, and to view them in terms of the expanding constituencies and types of information on the network.

To help to further clarify the scope and focus of this paper, let me emphasize that the paper devotes very little attention to the important and currently vigorously debated questions about government information, and in particular what government information should be available to the public, under what terms, and at what costs. This is a public policy issue of considerable complexity in its own right, and has a number of specific implications for libraries, particularly in their roles as depositories of and access points to government information. If anything, the emphasis here is more on information created and owned by other institutions, such as publishers and the academic community. An aggressive government program which expands the base of publicly owned information that is then made available to the public widely, at little or no additional cost, could well begin to alter some of the trends and evolving balances that are discussed throughout this paper. These issues merit considerably more exploration and discussion. Hopefully, however, this paper will provide a basis for such discussion, since, to a great extent, issues of access and integrity of collections of information in digital formats, and the roles of libraries in organizing, preserving and providing access to these collections of information are independent of the information's source.

## 4. Access to Electronic Information: the Changing Legal Framework

### *From Copyright and Sale to Contract Law*

In order to understand the massive shift that is taking place as libraries begin to acquire electronic information to supplement and perhaps ultimately replace much of the printed information they have traditionally purchased, it is first necessary to review the evolution of the system through which libraries support access to printed material to provide a basis for comparison.

To a great extent, the cornerstone for the success of libraries as institutions in the United States has relied on the fact that printed information is virtually always purchased by libraries, and thus its subsequent use is controlled by the copyright laws of the US and the doctrine of first sale. Without going into great depth (and the reader should note that the author of this paper is not an attorney, and is not attempting to give legal advice, but only his own understanding of the way that the involved institutions have typically interpreted the relevant laws), what this legal framework for print publications means is that a library, having purchased a book or journal is free to loan it to patrons or other libraries, or to re-sell the material.[12] Moreover, the use of this copyrighted material by the libray or its patrons is governed by the copyright law, which represents a public policy established by the Congress through legislation which is based on the constitutional responsibility delegated to the Congress to promote the useful arts and sciences. The promotion of these useful arts and sciences has historically required the Congress to achieve a balance between compensating the creators of intellectual property and recognizing the needs and interests of the public to have access to such intellectual property as a basis for further progress in these arts and sciences. (Note that in this section I will use the term "copyright law" to refer rather broadly to the actual legislation itself, and also the body of legal interpretation and legislative history and intent that surrounds and supports the actual legislation.). Under copyright law, while a library, for example, is free to loan the physical artifact it has purchased, it is restricted, in most cases, from duplicating the material. Certain provisions of the current copyright law such as the Fair Use provisions explicitly recognize, for example, the importance of permitting a scholar to quote from a copyrighted work (within limits) for the purpose of criticism or further scholarly analysis; the copyright law also recognizes the importance of permitting the use of copyrighted material in certain educational settings. The importance to society of preserving the historical record is recognized in the permission that is granted to libraries to make a single copy of a physically deteriorating out of print but still copyrighted work for preservation purposes if no reasonable substitute is available. Patent law, to cite another example of the balance between encouraging creators of intellectual property and the needs of the public to be able to make use of that property once it is created, can be viewed as a compact that trades protection of an inventor's intellectual property

12 There is no international consensus on the details of the doctrine of first Sale; In some European countries, for example, I believe that the right to loan purchased materials can in fact be excluded from the set of rights passed to a purchaser (such as the right to resell the copy of the work that he or she has purchased) as part of the initial sale.

(during a limited period of time) for the disclosure of that property to the public as a means of moving the state of the art forward.

It is important to note that there is nothing, as far as I know, in the current copyright law that requires printed material protected by copyright to be sold to purchasers; there is no a priori reason why a contract could not be written by a rights holder which allows an organization to license copyrighted printed material for a limited period of time and for specific uses outside of the framework of the doctrine of first sale, and such a license might grant the licensee much more limited rights than the licensee would have had under copyright law if the material had been actually sold. In fact, certain print publishers (for example, publishers of some types of atlases, directories and other compilations such as Dun and Bradstreet) have attempted to use licensing as a means of controlling their products for years, and recently the library community has encountered a number of other cases where print publishers have attempted to restrict the use of their publications through license agreements.[13] In many cases, the validity and enforceability of these terms imposed by the publishers have been of ambiguous as the publishers seem to be following the "shrink-wrap" license model of the software industry (that is, a buyer "purchases" something and by opening the package, or at least by using the "purchased" product the publisher states that the buyer is implicitly entering into a legal agreement to adhere to the license terms stated in the license agreement enclosed with the product). In most cases the library community has resisted (and occasionally prominently ignored license terms) products that come with such license constraints. I am unaware of significant test cases that have come to litigation to clarify these situations; further, I would speculate that in many cases where publishers have a strong motivation to attempt to protect print publications with license terms that limit, for example, who can see the printed material, are for very high priced, specialized, highly time sensitive information which is more typically purchased by commercial corporations rather than the general library community (for example, information related to the financial or securities industries).

An interesting related issue is whether publishers are required to make available their publications to libraries,[14] and if so under what terms and constraints. It is a well established practice today for publishers to use differential pricing for libraries and individuals, often with a very wide price spread: a scholarly journal might cost a library $1000/year for a subscription, but the publisher may also offer individuals at a subscribing institution (such as a university) the option of purchasing supplemental individual subscriptions for only $150/year.[15] Another common variation is a journal

---

13 The a~emp t t. contro[ access to print material through license agreement is not new, although it is relatively unusual in the United States. See, for example, Library Journal Volume 1 Number 2 (1877) in which the practices of publishers are deplored.

14 From a strictly legal perspective, I understand that publishers are under no compulsion to offer their works to any specific individual or institution.

15 While Commercial publishers often offer discounted subscriptions to *individuals* as a benefit for belonging to institutions where a library subscription to a journal is held, some professional societies go even farther and offer discounted subscriptions not only to individuals but to academic *departments* (for example, for departmental libraries, which are typically funded at the department level) for additional copies of journals. Today, subscription costs for scholarly journals have escalated to the level where individuals seldom subscribe even at the discounted prices, but the departmental level offerings are very attractive to some academic departments. The American Mathematical Society, for example, has used a

that comes to each individual that is a member of a professional society as a benefit of membership, but that is only available for library subscriptions (since membership in the organization is on an individual, rather than institutional basis, or institutional membership is many times as expensive as individual membership).[16] In response to such pricing schemes a librarian at an institution is often "encouraged" to join the professional society to get the journal at a reduced rate to the institution, with the membership fee reimbursed to the librarian. The publishers and professional societies suggest that it is at the least unethical for an individual to act as a "front" for an institution by ordering an individual subscription that is really going to be placed in the institution's library for general use,[17] and sometimes ask individuals ordering under the individual pricing scheme to sign statements asserting that they will not place the copies they are ordering in a library (indeed, some publishers have gone so far as to affix stickers indicating "individual use only" to publications shipped under individual subscription rates), but the legal enforceability of this is unclear given that once the individual obtains the material it is subject to the doctrine of first sale. Yet another interesting variation on this theme **arises** with publishers of very costly, time sensitive material-for example, a market research report at $5000 a copy. Such a publisher might well choose not to market to libraries, and perhaps not even to sell to a library that placed an order (or to process such an order very slowly). It is unclear whether such a publisher could actually be compelled to make a sale to a library.[18]

### *Licensing and the Interlibrary Loan System in the United States*

The doctrine of first sale has also had another critically important implication for the library community in the United States; it has allowed the development of the interlibrary loan (ILL) system which permits one library to borrow materials from another. Historically, interlibrary loan was implemented by the physical shipment of materials from one library to another, which is clearly permitted under the doctrine of

---

central library subscription to subsidize departmental copies of some of their publications for a long time. It should also be noted that differential pricing is not clearly a completely evil thing; in the case of the American Mathematical Society, for example, one can view the library as subsidizing increased access to the Society's publications (in the pre-computer networking era) by allowing the Mathematics department to obtain a second, easily accessible copy of the material for a very small co-payment, Of course, in an age of pervasive networks, where institutional copies of material should be as readily accessible to the institutional community as departmental or personal copies, the entire concept of differential pricing begins to break down quickly.

[16] It should be emphasized that what might appear to be "discriminatory" pricing that Costs libraries much more than individual is really much more complex than it appears on the surface. One view of this situation is that libraries are partially subsidizing individual or departmental access to the literature for the cost of a small "co-payment". Another point to be considered is that publishers are making material available to individuals for the marginal costs of additional copies in print runs, discounted for the advertising that they can sell because of the individual readers of the journal. Differential pricing for libraries actually offers some benefits to libraries, institutions, and individual subscribers at those institutions.

[17] Arguably this could be regarded as fraud; while the argument that it Is unethical **IS Clear, it** is **unclear** whether it is' really illegal.

[18] Ironically, many of these publishers of very expensive reports file copies of their material with the Library of Congress for copyright reasons, and this material is often publicly available there (though not always on a very timely basis).

first sale (though recently technological innovations, ranging from inexpensive xerography to facsimile transmission and more recently image transfer across the Internet have greatly complicated the picture and moved libraries into areas that are of much less clear legality). Libraries—and particularly research libraries—in the US are linked by an elaborate, complex set of multilateral interlibrary loan agreements. In many cases libraries have traditionally simply agreed to reciprocal sharing without charge to either library; in other cases (which are becoming more frequent as the size of the total body of material grows and the ability of individual libraries to locally acquire even a significant portion of this body of published material diminishes, leading to a massive explosion in interlibrary loan) the supplying library charges the borrowing library a fee, which may or may not be passed on to the patron at the borrowing library that initiated the interlibrary loan request. (Currently, even many public libraries assess patrons a nuisance charge—perhaps a dollar per book—for interlibrary loan, more as a means of controlling frivolous requests than anything else.) In cases where one library recharged another, there were often resource sharing funds established through agencies such as state libraries which helped to subsidize the costs of interlibrary loan (either by directly compensating the lending library for its costs in servicing interlibrary loan requests, or by compensating borrowing libraries for interlibrary loan charges that they incurred), at least in the case of public libraries as borrowers or lenders, leading to patterns where research libraries (for example, at universities) acted as the providers of last resort to public libraries within a state or even nationally.

To provide some quantitative sense of the size, cost and importance of the interlibrary loan system in the United States, consider the following figures from a recent Association of Research Libraries study on interlibrary loan [Baker & Jackson, 1992; Roche, 1993]. ARL's figures indicate that among the 119 ARL libraries in the US and Canada, interlibraryborrowing has increased I08% between 1981 and 1992; lending has grown 52% in the same time period. Recently, the growth rate in these areas has accelerated: lending has grown 45% from 1985-6 to 1991-2, and borrowing 47% from 1985-6 to 1991-2. As indicated, much of the driving force for this growth has been the increasing inability of libraries budgets to acquire published materials: since 1981, ARL library materials budgets have increased 224% while collections grew only by 12%; during this period ARL tracks the average cost of a book as rising 49%, and the cost of a journal subscription I09%. The average cost among ARL libraries for an ILL lending transaction is now about $11; for an ILL borrowing transaction, the cost is about $19 (note that these are costs, and not necessarily what one library charges another for such a transaction). Current estimates suggest that the US library community spends around $70 million per year performing ILL transactions.

The interlibrary loan system essentially allows libraries, as a community, to purchase expensive and/or lightly used printed materials and share them within the community, though of course the specific library that has actually purchased the material can offer better access to it than other libraries which may have to obtain it on behalf of their patrons through interlibrary loan. Interlibrary loan has historically been a rather slow, expensive, and inefficient process. As use of ILL has increased due to the diminishing ability of any given library, even a world-class research library, to hold the majority of the published literature, considerable attention has been focused on speeding up and

improving the cost efficiency of the interlibrary loan process.[19] Work in this area has ranged from the use of electronic ILL systems linked to large national databases of holdings (such as OCLC) which allow a requesting library to quickly identify other libraries that probably hold material and dispatch loan requests to them through the exploitation of technology to reduce the cost of the actual shipment of material. The first step in this latter area was for the lending institution to send a Xerox of a journal article rather than the actual journal copy, so that the borrowing library did not have to return the material and the lending library did not lose use of it while it was out on interlibrary loan. This use of photocopying technology was controversial, and libraries ultimately agreed to limit it through the CONTU guidelines, which define limits on the number of articles that a borrowing library can request (from whatever source) from a given journal per year without additional payment to the rights holder (either through article-based fees assessed through organizations such as the Copyright Clearance Center or by subscribing to the journal directly); while the CONTU guidelines are generally accepted practice, they do not have any legal standing that I am aware of, and have not been subjected to any test cases—they merely represent an ad hoc, informal agreement between the library and publisher communities as to the bounds of acceptable behavior by libraries. More recently, libraries have employed both fax and Internet-based transmission systems such as the RLG Ariel product to further speed up the transfer of copies of material from one library to another in the ILL context, and with each additional application of technology the publishers have become more uncomfortable, and more resistant (with some legal grounds for doing so, though again this has not been subject to test). Interestingly, over the past two years we have seen the deployment of a number of commercial document delivery services (the fees for which cover not only the delivery of the document to the requesting library but also copyright fees to the publisher) offering rates that are competitive—indeed, perhaps substantially better-than the costs that a borrowing library would incur for obtaining material such as journal articles through traditional interlibrary loan processe.[20] At the same time, the research library community is mounting a major effort under the

19 It is important to recognize that in the United States the interlibrary loan system is a distributed sYstem in most cases; that is, individual libraries make their own decisions about which libraries they will use as potential interlibrary loan providers. Libraries set their own rates for interlibrary loans typically, though in some cases there are consortium arrangements that provide fixed rates for libraries that are members of the consortium. This is in contrast to the situation in many other countries, where there is a national library that acts both as the coordinator of the country's ILL system and usually as the lending library of last resort for the national ILL system. In the US, such arrangements only occur in the biomedical and health sciences, where the National Library of Medicine is designated to act as a national library, and to a lesser extent in the agricultural disciplines, where the National Agricultural Library often acts in this role, although its primary mandate is to function as a library supporting the US Department of Agriculture rather than the agricultural community as a whole. The Library of Congress is not a national library and does not serve this function for interlibrary loan. In some states, such as California, the state research university (the University of California, the case of California) serves as the library of last resort for borrowing libraries within the state, though often coordination of in-state interlibrary borrowing patterns is done by the State Library rather than the libraries of last resort for interlibrary loan.

20 The economics of interlibrary loan are actually quite complex, in the sense that lenders of material through interlibrary loan frequently do not recover their full costs for providing material, and, in fact, it is a matter of considerable debate whether they can even identify what these costs are; although the ARL study provides an important step in this direction, it focuses on one specific type of library. One can look simply at how a library trying to obtain material for a patron can do this most cheaply, but perhaps a better perspective would be to consider the overall cost to the library community as a whole in supporting interlibrary loan as opposed to using commercial document delivery services.

auspices of the Association for Research Libraries to improve interlibrary access to materials. It remains to be seen to what extent the new commercial document delivery services supplant traditional ILL in the 1990s.

Now, consider a library acquiring information in an electronic format. Such information is almost never, today, so/cl to a library (under the doctrine of first sale); rather, it is licensed to the library that acquires it, with the terms under which the acquiring library can utilize the information defined by a contract typically far more restrictive than copyright law. The licensing contract typically includes statements that define the user community permitted to utilize the electronic information as well as terms that define the specific uses that this user community may make of the licensed electronic information. These terms typically do not reflect any consideration of public policy decisions such as fair use, and in fact the licensing organization may well be liable for what its patrons do with the licensed information. Of equal importance, the contracts typically do not recognize activities such as interlibrary loan, and prohibit the library licensing the information from making it available outside of that library's immediate user community. This destroys the current cost-sharing structure that has been put in place among libraries through the existing interlibrary loan system, and makes each library (or, perhaps, the patrons of that library) responsible for the acquisitions cost of any material that is to be supplied to those patrons in electronic form.[21] The implications of this shift from copyright law and the doctrine of first sale to contract law (and very restrictive contract terms) is potentially devastating to the library community and to the ability of library patrons to obtain access to electronic information—in particular, it dissolves the historical linkage by which public libraries can provide access to information that is primarily held by research libraries to individuals desiring access to this information. There is also a great irony in the move to licensing in the context of computer communications networks-while these networks promise to largely eliminate the accidents of geography as an organizing principle for inter-institutional cooperation and to usher in a new era of cooperation among geographically dispersed organizations, the shift to licensing essentially means that each library contracting with a publisher or other information provider becomes as isolated, insular organization that cannot share its resources with any other organization on the network.

### *Other Implications of the Contract Law Framework for Libraries*

The shift from copyright law to license agreements has a number of other implications, all of them troublesome. At a public policy level, the ability of the Congress to manage the balance between the creators of intellectual property and the public has been undermined since copyright law no longer defines this balance; rather it is defined by specific contracts between rights holders and libraries. From the legal perspective, there is a very complex and ambiguous area having to do with the preemption of contract law (defined, at least in part, at the State level) of provisions defined by Federal (contract) law. At the level of the individual library writing contracts for

---

21 It is important to recognize that there are several conflicting and perhaps equally legitimate viewpoints here. Some publishers are viewing the transition to contract law as an opportunity to address what they view as an interlibrary loan system that has been pushed to the limit, well beyond where they are comfortable: they view some of the fax-based interlibrary loan resource sharing arrangements in force today as going well beyond the legislative mandate for shared access to information.

information in electronic form, the implications are even worse. Very few contracts with publishers today are perpetual licenses; rather, they are licenses for a fixed period of time, with terms subject to renegotiation when that time period expires.[22] Libraries typically have no controls on price increase when the license is renewed; thus, rather than considering a traditional collection development decision about whether to renew a given subscription in light of recent price increases, they face the decision as to whether to lose all existing material that is part of the subscription as well as future material if they choose not to commit funds to cover the publisher's price increase at renewal time. (In this context, it is important to recognize that price increases of 50°/0 or more at renewal time for electronic information are not uncommon, and that publishers are offering libraries various types of one-year free trials or other special introductory offers; given that all evidence is that electronic information, from a patron perspective, is among the most attractive and heavily used offerings of many libraries, and that in a large number of areas a given provider has what is essentially a monopoly position with no effective competition, no substitutable alternative available to the library-leaving the library with little choice but to seek the funds to pay for very large, unexpected, and, from the library's perspective sometimes extortionate cost increases). In a bad budget year a library might cut back on its purchases and subscriptions in the print environment, relying on its existing collection and ILL for new material that it cannot afford to purchase in the bad year; in the new electronic environment, a bad budget year might well cause the disappearance of much of the existing collection as well as affecting patron access to newly-published information.

The most common licensing situation today is for information that is either stored on some physical medium (such as a CD-ROM) which is then housed in the library or information provided on tape which is then mounted on computer systems belonging to the library. But, in fact, various types of usage restrictions defined by license agreements apply equally to remote databases where the library has contracted for access through the network. Indeed, some of the traditional database producers who offer access to their databases through online services like Dialog have a long-standing tradition of introducing complex and odious contractual terms[23] that predate any

---

22 One should not make too much of the problem of limited term as opposed to perpetual licenses. Some electronic information vendors are already incorporating perpetual license terms in their offerings, and several of the major publishers of primary journals have indicated a willingness to at least discuss a permanent licensing framework that parallels print practice. But this is an issue that libraries need to be aware of, and which does definitely have budgetary implications.

23 T. provide a sense of the contractual restrictions that one might encounter, consider the following terms excerpted from a recent set of contracts drafted by Dun and Bradstreet Inc. (D&B) and distributed to members of the Association of Independent Information Professionals, an association that represents free-lance information searchers that often work under contract to large companies. These terms are part of a rather complex legal agreement involving an independent information professional (IIP), and an end user employing the IIP, as well as D&B itself, and govern use of D&B information retrieved from Dialog. The terms include the following provisions:

- The information may only be provided by the IIP to a single end user.
- The I I P may not overlay additional data elements to information retrieved from the D&B files, or merge this information with information from other sources (thus adding value). The information cannot be used to create a whole or any portion of a mailing or telemarketing list, or any other marketing or research aid or data compilation sold by the IIP to a third party.
- Only certain specific D&B files mounted on Dialog can be searched.
- I I P may not provide information to any person the I I P has reason to believe will use it to engage in unfair or deceptive practices, or that will republish, subsequently license or resell the information.

significant use of these databases by research or public (as opposed to corporate) libraries.[24]

Practically speaking, no library is going to negotiate thousands of contracts, and no publisher wants to maintain contracts with thousands of libraries. This means that an industry of rights brokers will come into being. These may be aggregators such as University Microfilms (UMI) or Information Access Company; they may be clearinghouses similar to the Copyright Clearing Center (CCC) in the print world. Utilities may come into being that provide aggregated access to material provided by publishers, just as companies such as Dialog provided access to databases in the 1970s and 1980s; OCLC, among others, is already positioning for such a role through its mounting of various electronic journals such as *Current Clinical Trials.*

The uncertainty and restrictions surrounding contracts for access to electronic information are not the only problems that libraries will face in this transition from copyright law and purchase to contract law. Today, at least, there is no standardization of contracts, and efforts such as the CNI READI project [Peters, 1992b] that have sought to explore the potential of such standardization have been discouraging. Given that a large research library may well deal with several thousands of publishers in a given year, one can quickly see that in the electronic information environment, such a research library will be in no position to negotiate several thousand unique contracts with publishers for electronic information resources. Further, it is not just setting the contracts in place, which, as discussed, can be addressed to some extent through rights brokerage organizations (though this may impose a new cost on the overall system that is largely absent in print publication). Imagine the plight of a library that is attempting to support its host university's decision to enter into a cooperative agreement with a small not-for-profit research center which includes access to the university's library: there are now potentially thousands of separate contracts that need to be reviewed before the library can understand access constraints on this new cooperative venture; even in the case of rights brokerage organizations, unless there is a high degree of uniformity in the set of rights that the broker is prepared to license from various rightsholders and also in the terms of these licenses. In the most absurd case, a request to have access to a specific electronic information resource at the university library might well become a matter for the university's general counsel to

---

- The IIP acknowledges that D&B may introduce identifiable erroneous names that permit D&B to audit information use compliance, and the IIP agrees not to remove any such erroneous names.
- The end user agrees not to use the information as a factor in establishing an individual's eligibility for credit or insurance to be used primarily for personal, family or household purposes, or for employment.

24 One particularly sore point has been prohibitions on independent researchers who act as contractors for companies. The Association of Independent Information Professionals (AIIP) has spent a great deal of time over the past few years attempting to negotiate permissions to allow its members to legally use these databases on behalf of their clients. AIIP members, as independent small business people, tend to be meticulous about the legality of their database usage; the complexities that such usage restrictions would create for members of the academic community are so profound that in the academic environment it is likely that the restrictions on use would either be largely ignored, or the restricted use databases simply wouldn't be used. In an competitive environment this would be to some extent self-correcting; the restrictions would reduce revenue, and the competing product with the most liberal usage restrictions would be likely to gain market share. But, in an environment where the business community is the predominant revenue producer and the database is a relatively unique resource that is also needed by the research community, such restrictions quickly begin to function as a major barrier to access.

consider, and might take weeks or even months to resolve. This is a great distance from the old world in which a library could make its own decisions about who had access to its collection, and could readjust these policies in response to changes in the agreements that its host institution entered into.

The ability of publishers to specify usage conditions in a contract can also create other complex liabilities and administrative problems for libraries. For example, a publisher can specify that licensed material is being provided only for research and teaching applications. This can create legal ambiguities (and thus potential liabilities) if, for example, a faculty member uses some of this material in a book that is being commercially published, even if the use of this material might normally have been covered under the fair use exemptions for criticism within the copyright framework. Other information providers restrict printing of their material, downloading into personal databases, or the sharing of information with third parties; many of these restrictions are hard to define precisely, impossible for the library to enforce, and probably unrealistic. Given that the complexity of the usage restrictions is defined only by the publisher's ingenuity, we should not overlook the possibility of very odious usage restrictions being incorporated in contracts, or the liability that such contracts may create for libraries.

In providing a balanced view of the transition from copyright law to contract law it is important to recognize that institutions acquiring intellectual property are not simply being bullied into accepting contracts. While librarians may feel some discomfort about the transition, senior executives in most large organizations (for profit corporations or educational institutions) are relatively conservative and risk-adverse. The definition of rights under copyright law is interpreted through court decisions (case law) and thus appears to these decision-makers as fraught with ambiguity and uncertainty. Events such as the recent Texaco decision highlight the risks that institutions operating under copyright law can face. Contract law is less ambiguous and less risky as a framework for acquiring intellectual property, assuming that one has the money to pay the rights holders; this is a real attraction to management at many institutions.

The shift from copyright law to contract law seems inexorable, barring major restructuring of the copyright law. While consideration of such restructuring is outside the scope of this paper, which is focused on extrapolating and illuminating the implications of current trends within the current legal framework, it should be noted that a number of proposals have been developed for alternative legal frameworks. One example is compulsory licensing of materials in electronic formats. Depending on how fees were set in such an environment, some of the problems with current trends could be addressed, although such a shift would give rise to a number of new problems and issues, both at the public policy and practical operational levels.

### *Libraries and Multimedia: New Legal Issues*

An increasing amount of electronic information will be in various multimedia formats. Here, a new and complex problem arises that is already causing great concern in the library community; the initial source of that concern is videotapes, videodisks and related materials. When dealing with printed material, the typical concern is over the right to make copies for various purposes. Loaning of physical artifacts is not much of a

question. Certainly, the ability of a library patron to view a book owned by a library is not an issue. But, if one studies the copyright law, one quickly finds that it addresses not only the right to copy, but also rights to display and perform works. While these are not much of an issue in respect to printed materials circulated by libraries, they represent real and complex problems for multimedia works. Without reviewing the legal details (which appear to be highly controversial at the present time) it does seem clear that the role of libraries in providing access to multimedia works is at best ambiguous. Even when operating under the general framework of copyright, as opposed to contract law, the worst case may be that libraries can acquire material that they cannot legally provide their patrons with the facilities to view within the library. (For a good summary of these issues from a legal perspective, see [Cochran].) Given the growing complexity of the technology base need to view (and interact with) certain types of multimedia products, this represents a very real barrier to access.

The issues related to traditional audiovisual materials are already serious, and have been a source of major problems for libraries. Early experiences with the lending of software has also revealed numerous issues. But perhaps even more important is the unresolved extent to which rights of performance and display will be attributed to the viewing of electronic information of all types, ranging from the browsing of bitmapped images of print pages through interaction with a digital movie driven by a program.

The widespread development of multimedia authoring tools will raise other issues as well, perhaps less for libraries than for users of digital information on the network. Multimedia integrates film clips, images, music and sound along with other content, and most developers of multimedia are not simultaneously artists, composers and musical performers. There will be a great demand for copyright-free (public domain) materials that can be included in multimedia works.[25] Here again one encounters a large number of ambiguous questions related to copyright law. One can find numerous opinions on these questions, but only very limited consensus and even less certainty. The questions include:

- Who owns the rights to an image? This includes photographs, images of classic paintings, and other materials? This is a particularly vexing question in regards to paintings owned by museums, for example. It's important to recognize that this is not a new problem that has been created by the digital environment; ownership of images is a very complex (and sometimes ambiguous) area in copyright law even in the older print and photographic environments. Digital imaging technologies and network distribution simply underscore existing uncertainties.

- If an image is digitized, and then perhaps subsequently enhanced, is this protected under copyright?

---

25 We are already seeing the beginning of this process in the very complex restrictions that accompany demo disks for products such as Apple's Quicktime and the various "clip art" and "stock photo" products being offered to multimedia developers. People want to incorporate bits of this material into all sorts of new multimedia: sales presentations, educational materials, demonstrations, and training materials. In many cases it is unclear when rights must be cleared. It seems likely that a few well-publicized legal actions could lead to an atmosphere of pervasive paranoia that might quickly retard the use of multimedia technologies, particularly by the business and educational communities.

•To what extent is the linkage of a series of media (for example, images and a sound track) copyright able separately from the images themselves and the sound track itself?

•If an out of copyright text is scanned or keyboarded and then edited, to what extent is this protectable under copyright?

•How does the developer of a multimedia product, attempting to comply with the law and to behave responsibly, determine whether component works that he or she wishes to incorporate into a multimedia product are protected by copyright?

•To what extent are libraries (or other networked information providers) liable for contributing to copyright infringement in an electronic information environment. To give only one example, a number of libraries are currently considering how to upgrade their facilities to accommodate users of portable notebook computers in conjunction with the overall move to electronic information. If a library permits patrons to connect to a library network to download images from the library's collection, to what extent is the library liable if these images prove to be copyrighted?

The problem here is again in large part the uncertainty. As matters stand today, many of these questions will have to be decided, at least as I understand it, through test cases in court. Most libraries (and their parent organizations, such as universities) are not eager to serve as such test **cases.** It is quite possible that attempts by libraries to limit the potential legal liabilities of the current uncertain copyright framework may also contribute to the destruction of the interlibrary loan system through a migration to acquiring material under license (contract law); understandably, most organizations will place a greater priority on managing their own legal exposures than they will on the ability to share their material with other organizations. Of course, there are alternatives to defining the specifics of copyright in the electronic environment through case law: these include both specific legislative actions that clarify and perhaps further define the law in the electronic information world, or joint agreements between information providers and information purchasers similar to the CONTU efforts which establish community guidelines without having the actual force of law.

As one reviewer of an draft of this paper reminded me, it is also important to consider the perspective of the individual developer (perhaps in a very modest, non commercial sense) or user of multimedia—the teacher in the classroom, the marketing representative preparing a sales presentation, or the individual citizen amusing him or herself. As has been illustrated by such issues as the transcription of phonograph records or more recently CD-Audio recordings to audio tape, or the taping of shows from the broadcast media for later viewing, many such users are literally unwilling to recognize or worry about legal restrictions to actions that seem to them to be reasonable. Such individuals—the majority of the citizens of the US—are likely to become outlaws (most often unwittingly, but occasionally as a matter of deliberate choice) outside of the framework of institutions and institutional liability rather than abide by complex, hard to understand legal restrictions that seem intuitively senseless. Laws that are strongly at odds with social norms are frequently bad laws, as they undermine people's acceptance of the overall law of the land. In areas such as personal taping (so that one can play the Audio CD one just purchased on one's car stereo, or watch a program that was broadcast again from one's VCR) my sense is that

the courts have largely upheld community consensus about what is reasonable, and sided with the consumer; any other choice would be impractical, to say the least. We will face many of the same issues in digital multimedia.

## 5. The Role of Secondary Information Sources and Automatic Indexing in Access to Electronic Information

Today, universities and other organizations are licensing and mounting abstracting and indexing (A&1) databases as adjuncts to their online catalogs; these databases provide library patrons with logical access to the journal literature (and sometimes also book chapters, books, technical reports, and other material) in a given discipline. A&1 databases contain citation information—authors, titles, journals of publication, page numbers—for material such as journal articles; in addition, they often include subject headings or other access terms assigned by indexers, and sometimes also abstracts of the contents of the articles, book chapters or other materials. Abstracting and indexing database records fill a role similar to that of cataloging records for books in a library catalog, but often provide more information about the work than a library catalog will. In general, library catalogs focus on monographic material-books, maps, sound recordings, films-or entire periodicals (for example, recording the fact that the library had a subscription to a given journal); abstracting and indexing databases typically focus on articles in journals or chapters in books.

Currently one of the major challenges for libraries is bridging the gap between the intellectual access offered by abstracting and indexing databases and access to their physical journal collections (as described in their catalogs thorough records of which journals they hold); in future the abstracting and indexing database providers will also offer links to electronic publications directly. The compilers of these databases wield great power that is just now being fully recognized. The experience of libraries in mounting online catalog databases (which typically cover only the monographic literature held by a given library) has been that when only the online catalog database was available some patrons tended to use monographic material almost exclusively; other (arguably more sophisticated) patrons who recognized that the journal literature was vital to their discipline tended to reject the online catalog as irrelevant. Indeed, this reaction to online catalogs was one of the primary forces that motivated libraries to license abstracting and indexing databases to attempt to bring access to the journal literature into balance with the access that they already offered to the monographic literature. Now that A&I databases in various disciplines are readily available to library patrons[26] these effectively define the relevant literature in these disciplines both in their

26 A few Points **should** be made about the origins and development of abstracting and indexing databases, and the impact of their conversion to electronic formats. In the mid 1800s various individuals and organizations began to compile indexes to parts of the journal literature and market these to libraries; however, the size of the journal literature was sufficiently small until the early 20th century so that at least large research libraries could actually create article-level card catalog entries for articles in journals to which they subscribed. Thus, up until the early 20th century, the library catalog served as a record of material that the library held, and specialized indices **served** as a means of providing access to the entire published literature in an area (whether the library owning the index owned the material or not). With the explosion of publication during the later part of the 20th century, economic considerations forced libraries to abandon the cataloging of articles in their journals, and they began to rely exclusively on subject bibliographies of the journal literature to provide patrons with access to journal articles. Thus, the print analogs of abstracting and indexing databases are nothing new. However, these printed tools were

selection of journals to index and in their chronological span. For all intents and purposes, if material in a given journal (or even a given issue of a given journal) isn't covered in the abstracting and indexing database, it might as well not exist from the patron's perspective.

Thus, the processes through which the compilers of these A&1 databases select which journals to index, and which articles within these journals should be indexed, are effectively defining the literature in various disciplines. Most library users are unaware of the precise chronological or literature coverage of these databases, or the differences from one database to another (and note that a library generally selects only one database in a given discipline, typically based on a mixture of quality and cost considerations, due to the very high cost of licensing and mounting such a database); indeed most database providers are very vague about even stating their coverage and selection policies, which can be substantially complex. This confusion is compounded by the fact that these A&1 databases evolve over time, and revisit their selection of journals to index, and the indexing policies (i.e. cover to cover indexing, which creates a record for every item that appears in the journal, or selective indexing, which only

---

generally hard to use and were seldom consulted except by librarians and by scholars familiar with their organization.

In the 1960s the organizations that prepared these bibliographies of the journal literature began to employ computers to manage citation databases than were then formatted for print; as the cost of computers began to drop, they made the databases available for online access, either directly or through service bureaus like Dialog or BRS. The first such databases supported relatively well-funded disciplines like the biomedical and health sciences (for example, the MEDLINE database), general science (the Current Contents and Science Citation Index databases), engineering (the INSPEC database), or the business and financial communities (ABI Inform); access to these files was very expensive (sometimes hundreds of dollars per hour) and because of the high costs use of these databases was largely limited to researchers in commercial corporations or occasionally academics with grant support. Universities sometimes offered a very limited amount of subsidized searching (for example, a few searches per year for faculty, or a search or two for doctoral candidates working on their dissertations). Also, because the search systems were not only very costly but also very difficult to use, most searching was performed by trained intermediaries (typically librarians with special training). As a consequence, while these databases were important resources for researchers in commercial settings, they had an extremely limited impact within the academic community.

In the 1980s computing costs dropped to the point where universities could begin to license these databases at flat fees and mount them on local computers for unlimited use by their academic user communities, typically using software that was designed to support access by end users rather than by trained search intermediaries. Usage grew by orders of magnitude; for example, at the University of California, popular databases such as **MEDLINE** now support in excess of 100,000 searches ***per*** week by the UC academic community, and the availability of such databases began to have a major impact on university-based research and instructional programs.

The other point that should be emphasized is the very powerful impact of computer-based information retrieval tools in academic libraries. The experience with online **catalogs was that most users of the library found these automated information systems so much more convenient than the card catalogs they replaced that they would typically use the online catalog even if its coverage was less complete than the older card catalog because some material in the card catalog did not yet have machine-readable records that allowed this material to be represented in the online catalog. Similarly, while the printed abstracting and indexing tools were very difficult to use, the online versions of these tools (at least in conjunction with end-user oriented retrieval software typically** used when they are mounted at university libraries) make *the* electronic databases very easy to use, and these databases consequently gain very high user acceptance and quickly begin to **serve** as the primary-indeed often nearly the sole-means of access to the journal literature.

creates records for certain material in the journal, based on type of material or article content) for selected journals from year to year; just because an A&I database currently covers a given journal at a given level of detail does not mean that it provides historical coverage of that journal, or that it has always covered the journal at the same level of detail.[27] Yet users—at least those in disciplines which still take the published literature seriously, as opposed to disciplines that view the key literature as preprints, technical reports and other electronic publications—tend to regard the coverage of the A&I databases available to them as effectively defining relevant literature in a discipline.

In a very real sense, the challenge facing an author of a scholarly article under the "publish or perish" regime still commonplace in academia for print publication is to get *published;* whether anyone reads the publication is a secondary issue. In the evolving networked information environment, all evidence suggests that it is all too easy for anyone to share their thoughts with the networked community through self-publication. The challenge in the networked environment will not be to make one's writings available, but rather to get people to read them. This will assign an ever greater emphasis on the selection and coverage choices made by abstracting and indexing services, particularly those that are explicitly recognized by scholarly communities because (for example) they are provided by various scholarly societies.

On one hand it seems that this trend is encouraging. Greater importance will be assigned to reviewers and bibliographers of all types. A researcher in a given area may well be willing to pay for the bibliographies of important recent articles provided by major figures in his or her field. Reviewers for journals-currently normally largely

---

[27] Close examination of editorial policies for abstracting and indexing databases indicate that they are very complex and have considerable impact on what information the user locates and how they can locate it. Consider, as one example, a popular database that offers coverage of the parts of the computing literature. Basic records in this database include the author, title, date of publication, subject headings describing the contents of an article and related material. Some, but not all, database records also include abstracts. Some of the journals in the database are indexed "cover to cover", which means that descriptive records for all material *of certain types* appearing in the journals are included in the **database—but** this may only include news announcements and articles, and not letters to the editor, errata announcements for articles in previous issues, conference announcements and calls for papers, or other materials. Advertising is almost always omitted, even lengthy special advertising sections and the sort of quasi-editorial material like new product announcements that are often found in trade journals. For other types of journals only articles related to computing are included in the database; thus a paper in a journal like *Scientific American* would be included only if it dealt with computing. Since the database vendor incurs a significant additional cost for each abstract that is included in the database, abstracts are only prepared for some of the material, most commonly longer articles. The vendor also offers a supplementary extra cost product that provides full text for some of the material in some of the journals that are covered by database; journals are included primarily based on the ability of the database provider to negotiate an acceptable agreement with the journal publisher for the remarketing of the text of the material in electronic form. Within the journals that are supplied in full text form, the database provider again employs editorial policies to select only specific types of material for inclusion as **fulltext,** since for most journals the database provider must pay for scanning or rekey boarding of the material and thus again incurs substantial costs for material included. For some types of material there may be **fulltext** but no abstract. Now, add to this rather complex set of criteria for what is placed in the database the additional complexity that all of the editorial policies just described are subject to continual revision and fine-tuning.

The user of such a database is typically unaware of all of these subtleties. However, searching by subject terms will actually search a different, larger set of articles than those accessible when searching by full text or keywords in the abstracts, and the table of contents of a given journal issue as derived from this database are likely to be somewhat different than the contents of the printed journal.

unrecognized and uncompensated for their labors—may find their evaluations recorded in databases and assigned great importance. Those who edit, filter, and select may play a much more important role in the networked information world. But, at the same time, established arbiters of taste within a given discipline, such as the compilers of abstracting and indexing databases, may have a much greater role in describing the relevant literature of a discipline.[28]

A key question here will be the amount of diversity available. One perspective on the matter extrapolates from the existing compilers of abstracting and indexing databases: these are organizations that attempt to provide systematic and comprehensive coverage of the literature in a discipline. Developing these databases is a costly proposition; the creation of such a database is a major investment by a corporation or other institution. The other perspective uses the network to expand the reach of what has traditionally been interpersonal communication—someone passes an interesting article to a colleague. The individual-based filtering and selection services serve different purposes and in some ways are more valuable to information seekers increasingly pressed for time as they help such information seekers to locate key publications quickly. Here the model is more one of bibliographies and reader's guides, which can be produced for limited areas by a single specialist or a small cadre of experts with a fairly limited investment Of course, individual-based services are more subjective. One of the most attractive points about individually produced bibliographies and reader's guides is that it gives wider voice to major thinkers in a given scholarly discipline-the "geniuses", to use one reviewer's term, can reach beyond their immediate circle of students and colleagues to highlight what they believe to be particularly important works for the broader scholarly community. Both approaches will have their roles.

The entire issue of evaluation of literatures is controversial [White, 1989]. Some librarians and researchers (such as F. W. Lancaster) argue that this is one of the key contributions of librarians and of various reviewing services. Certainly, every library makes evaluations daily as part of its acquisitions decisions, but the often it avoids suggesting that one item it its collection is "better" than another once the evaluation decision leading to acquisition has been made. The argument has also been made that the standard review sources in many disciplines are at best very conservative: they only tend to cover material from certain mainstream publishers (and, indeed, in some cases they are owned by one of the major publishers in the field) and as such tend to reduce diversity and the introduction of innovative new material, in part because librarians at

28 Occasionally, one reads visions of future electronic libraries that include a very intensive reader commentary component. The idea is that readers will attach their reactions and comments to material placed in an electronic library by the primary authors. Effective realizations of such a framework have proved elusive in practice. There are too many readers, with greatly varying levels of expertise and objectivity. While broad-based reader commentary may be a useful thing to incorporate in future electronic libraries, I do not believe that it will replace the role of expert selectors and commentators. It is also worth noting that there are subtle intellectual property problems here. Will the general public be willing to contribute their comments on material for public access? Certainly, some experts will try to make income by providing such commentary; if the public at large emulates this, one has an administrative, legal and accounting nightmare. If the public does not, then one must ask why certain commentators are willing to share their thoughts on a work freely while other commentators are not. Some projects, such as Ted Nelson's XANADU {Nelson, 1988], have attempted to explore the compensation and intellectual property issues implied by a move from published works to a rich web of commentary that surrounds these works.

many institutions, overworked and/or lacking the necessary expertise to make an independent evaluation, will simply use the review sources as purchasing guides. It seems to be that the networked environment will increase diversity in reviewing sources, thought it is not clear to me that many librarians (as opposed to subject matter experts) will step up to the challenge of providing these new bibliographies, abstracting and indexing tools, and reader's guides.

The trend towards having large, costly abstracting and indexing databases define the "core" of a disciplinary literature is of particular concern in conjunction with visions of the future which place professional societies in charge of the canonical literature in a given discipline (see, for example American Physical Society document on the development of a future international physics electronic library [Loken, 1990]); the problem here is that while a given researcher who is out of step with the conventional wisdom in a given field may be able to make his or her thoughts available on the network, it is unlikely that anyone will find them. One can all too easily envision the "establishment" in a given discipline taking control of the definition of the literature in that discipline through the compilation of the de facto standard abstracting and indexing databases in that discipline. To a certain extent, the easy self-publishing that is possible in the networked information environment addresses these concerns, but as indicated earlier the challenge is not to be published but to be read, In cases when tenure and promotion are at issue, there is likely to be no near-term substitute for publication on a prestigious journal; but, when the objective is more communication with one's peers, the question is whether the developing tools for identification and discovery of networked information resources will provide an adequate "safety net" to allow self-published materials to be located and read by those peers.

Another aspect of the role of secondary information services is their role in author evaluation decisions—for example, tenure and promotion decisions for academic authors. Some of the more sophisticated universities are recognizing the potential for subjective bias that may be present in the traditional abstracting and indexing services, and prefer what are allegedly more quantitatively objective secondary services such as the various citation indexing offerings from the Institute for Scientific Information (ISI) such as Science Citation Index [Garfield, 1979]. Citation indices count the number of times that a given publication is cited in the published literature; it is only a short step from these to even more "objective" measures of quality based on the number of times that a given article is accessed (in electronic form, where this number of accesses can easily be computed); this raises fundamental questions about the privacy of searches and the uses to which searches can be put that are discussed in a later section.[29]

The growing power of abstracting and indexing services raises many questions that need to be explored, and places at least a moral responsibility on the abstracting and indexing services to exercise a very high degree of quality control (though the legal liability of such services, as far as I know, has yet to be defined; the general issue of legal liability of information providers is discussed in a later section of this paper).

[29] It should be noted here that citation rates are a somewhat controversial measure of the impact of publications. They are subject to "gaming" in various forms: repeated and extensive self-citation, or the development of tight circles of authors who continually cite each other's works [Pertiz, 1992]. Similar questions will undoubtedly apply to the use of measures based on the number of accesses to articles in the networked environment.

Consider the possible impact of an service that abstracts and indexes only selectively: for all practical purposes, by not including a given article, the service excludes that article from the literature of a discipline and makes it unlikely that researchers in that discipline will subsequently find the article in question. This at least is an editorial judgment,[30] consider the case where by some error an abstracting and indexing service "misses" an issue of a journal and all its contents (perhaps the issue was lost in shipment to the service, or lost by the service during its processing stream), or makes an indexing error which causes a publication to become unretrievable. Such omissions evidently do occur today in some of the major services that are used in contexts such as tenure and promotion decisions.[31] The entire issue of the quality of abstracting and indexing databases is quite complex and subtle; the interested reader might wish to examine the recent series of articles by Peter Jasco on aspects of this topic [Jacso, 1992a; Jacso, 1992b; Jacso, 1993a; Jacso, 1993b; Jacso, 1993c].

As we begin to transition from printed materials to electronic materials we tend to think of abstracting and indexing services (first in print formats, and now as electronic databases) as perhaps the primary means of identifying source materials. In fact, as more and more primary (e.g. full text, or source) material is available in electronic form, new methods of identifying relevant material will come into play based on various forms of automated indexing and full text searching,[32] [Salton, 1988]. This is inevitable for three reasons.

First, the human intellectual effort for abstracting and indexing is costly and the user community cannot afford or is unwilling to pay for people (particularly expensive people with subject expertise) to index everything, particularly in great depth. Even if an abstracting and indexing database is available which covers a given set of material, a library might offer that source material in electronic format but may have chosen not to license the abstracting and indexing database for any number of reasons (for example, because the library only holds a very small proportion of the material that is covered in

30 One of the reviewers of the initial draft of this paper raised a very interesting issues about editorial selectivity: based on the Feist decision, one might argue that a comprehensive, cover-to-cover indexing and abstracting service would have very limited protection under copyright, while a service that was more selective would find that their selectivity would justify stronger copyright protection. Copyright protections may well encourage greater selectivity.

31 In many disciplines there are multiple competing abstracting and indexing services. Publishers have less to fear than individual institutions from errors that are made by a single service; over the broad subscriber base, the multiplicity of services will make it probable that at least one service provides proper access to the publisher's materials. However, given the very high cost of acquiring an abstracting and indexing database, a given library or university will probably select a single supplier from the various alternatives available on the market; thus, for a given university community (within which, for example, a tenure decision is made) a single abstracting and indexing database will dominate.

32 Note that all of the issues raised already about the power of editorial decision making In the compilation of abstracting and indexing databases also apply to fulltext databases, whether created independently of A&I databases or constructed as extensions of these databases. Choices about what to include in fulltext will have to be made; some database producers may not choose to include full text of ail articles that they abstract and index, or more generally of all articles that appear in a given issue of a journal. And new opportunities for editorial bias appear: for example, a given service might exclude full text of articles that are critical of that service's performance or practices. Given that many users will be satisfied with the part of the published literature that is immediately available in full text electronic format, such editorial decisions can have a powerful impact.

the A&I database); in these cases there is no choice but to use information derived from the source material to provide access to it.

Second, while mechanisms based on full text may or may not offer "better" access to material [Blair & Maron, 1985; Tenopir & Ro, 1990], they certainly offer different access which is at least a useful complement to human intellectual indexing,[33] Access based on full text can be an excellent supplement to shallow abstracting and indexing databases (for example, those that provide little or no subject access). Full text access can help to identify documents that mention people, places or things that may not have been sufficiently central to the theme of the work to be recognized by an indexer or abstracter; in this sense they provide much greater depth of access. Further, there is a sense that full text access is less "biased" than abstracting and indexing services in the sense that human judgment does not come into play. Full text access can help users who are having difficulty with specialized controlled vocabularies typically used in subject classification in A&1 databases.[34] In situations where large textual documents are available online, the two techniques may be used together: first, search an abstracting and indexing database to identify relevant documents, then use full text based access techniques to identify relevant parts within these documents.

A third reason why full text based access is coming into wide use is because of the delay inherent in human intellectual indexing. Today, major (and expensive) abstracting and indexing services often run as much as four to six months behind the appearance of source material in print (and recall that the print material itself may be months or even years behind the distribution of preprints or manuscript versions of material with the "invisible college" community). As electronic dissemination of information increases the speed with which material is made accessible, these lags in abstracting and indexing will become increasingly unacceptable to some users of the material-particularly those interested in the most recent material rather than those performing retrospective literature searches. Full text indexing allows access through

---

33 The definition of the quality of a method of providing access to documents is a very complex and somewhat subjective area. In the information retrieval research community measures such as precision, relevance and recall are used+ssentially measuring how many of the relevant documents are retrieved by the access method, how many irrelevant documents are returned along with the relevant ones, and how many relevant documents are missed by the access method. Clearly, performing large scale comparative tests between different methods given these definitions is extremely difficult, since it requires that someone go through the entire database in order to determine the "correct" answer to the queries in order to evaluate the performance of the access methods being tested, because of the great variation in the kinds of queries issued by users (and the great variation in the performance of many access methods from one query to another), and because of the very subjective nature of relevant documents (since even experts do not always agree on whether a given document is relevant to a given query, and the judgment of the experts may still not agree with the judgment of a typical user who is not a subject expert). At the same time, we should recognize that while this is a hard problem, experimental results for various retrieval approaches on a wide range of large databases would be of enormous interest and value.

34 Full text access is also helpful in dealing with the fact that controlled vocabularies grow and change over time as new areas of interest emerge within a discipline and new discoveries and developments occur, but these changes tend lag substantially behind the events that cause them; this is a well-recognized problem with the Library of Congress Subject Heading controlled vocabulary list, for example. In most cases, the cost of updating subject terminology used in existing database records to reflect changes in the terminology is prohibitive; only a few very high quality databases such as the National Library of Medicine's MEDLINE do this. Often, one can find terms used in the abstract or full text of an article long before they become established in the indexing vocabulary of a discipline.

the apparatus of bibliographic organization to occur simultaneously with the act of (electronic) publication. It is also interesting to note in this connection that part of the problem is the size of the literature base that most comprehensive abstracting and indexing services attempt to cover (which goes hand in hand with their lack of evaluative information-they will help you find all the documents on a given subject, but not the three best surveys). If we see the development of large numbers of limited scope, highly selective and highly evaluative citation lists/bibliographies offered by subject experts as proposed elsewhere in this paper, we may find that these specialized lists are also much more timely than the traditional abstracting and indexing services.

Clearly some types of electronic material, such as newsfeeds, will require automated indexing; human indexing will introduce so much delay that much of the time value of the material would be lost. Multimedia information—images, video feeds and the like-present an additional set of issues. Today, we have very limited capabilities to perform useful computer based indexing of multimedia; general image classification is beyond current technical capabilities,[35] though automated transcription of speech (audio, or the audio track of a video segment) may become a production technology within the current decade, and this soundtrack could provide a very valuable access point for video information. Already, today, closed-captioning tracks in video material are being indexed and used to provide access points to broadcast information. And there is technology in experimental use that separates pictures from text in bitmapped images of printed pages [Lesk, 1991], or that attempts to detect scene transitions in video clips .

There are many different full text based retrieval methods [Tenopir & Ro, 1990]. The simplest provide searching for exact words that appear in the text, often with the option of including truncation (match only the beginning of a word), Boolean operators (i.e. AND and OR), and proximity operators (to require that two words appear close to each other) in queries. These full text access methods are easy to understand and predictable, although they often lead to rather poor retrieval results. Much more sophisticated methods have been developed in the information retrieval research community and are now starting to appear in large scale production systems. These range from statistically based methods pioneered by Gerald Salton and his colleagues over the last three decades[36] [Salton, 1988], which are based on frequency of word occurrence along with some very superficial language processing (word stemming) through much more complex techniques that combine statistical analysis with various syntactic and semantic analysis techniques from natural language processing (for example, analysis of parts of speech, identification of proper nouns or noun phrases, or

---

35 Certainly there have been great advances in image recognition in very specific problem domains ranging from quality control in manufacturing processes through target identification for smart weapons systems, but more general problems, such as identifying the objects in a picture, remain intractable to the best of my knowledge. Further, really useful classification of images for general purpose retrieval involves a great deal of cultural knowledge as well as simply the ability to identify things: identifying a photograph of the President of the United States shaking hands with the Mayor of New York is far more useful than simply recognizing that a photo depicts two men shaking hands.

36 In the past few years there have been a number of proposals for more sophisticated and computationally intensive statistically based indexing algorithms, such as the Latent Semantic Indexing techniques developed by Bell Labs [Deerwester, Dumais, Furnas, & Landauer, 1990; Foltz, 1990].

even attempts at actual language understanding). Recently, a major focus on the application of these sophisticated hybrid methods in large production textual databases by the DARPA TIPSTER [Harman, 1992] and TREC [Harman, 1993a; Harman, 1993b] projects has produced some impressive successes and may encourage their transition from research efforts to more broadly deployed systems. The difficulty with all of these sophisticated methods, however, is that their operation is incomprehensible to almost all users. It is very difficult to predict what they will retrieve and what they will ignore. Some critics of these approaches have termed them "information retrieval as magic". These technologies raise very real integrity and access issues in that they work reasonably well often enough to be useful but seldom work perfectly; worse, they fail drastically in a reasonable number of cases. And information seekers not only have no idea what these retrieval systems are doing, but very little sense of when they are or are not working right; and, as they move from one system to another (as will be increasingly common in a networked information environment) they also have no sense of the specific features and idiosyncrasies of a given retrieval system. And, unfortunately, little effort seems to have been invested in researching effective means for these systems to explain and document their processes to their users; such features would help a great deal.

To some extent, these sophisticated "voodoo" retrieval systems have been kept from the general public by groups like librarians who are sufficiently infomation-retrieval literate to recognize the problems and be alarmed by them. The general public won't care; as soon as these developing technologies become effective enough to provide a useful answer most of the time, the public will accept them (and swear at the "stupid computers" in cases where they don't work), unless we see an unprecedented rise in public literacy about information and information retrieval techniques. The unreliability of probabilistic and statistically based retrieval algorithms is today not a problem that the public understands; without such understanding they may well become victim to their limitations simply because they are easier to use than more traditional, deterministic approaches.

## 6. Access to and Integrity of the Historical and Scholarly Record

One can consider a printed work as knowledge bound at a given time. For example, an encyclopedia published on a certain date represents the common wisdom of society about a number of topics as of some point in time. Indeed, old encyclopedias, obsolete textbooks, out of date subject heading classification guides and other literature represent primary databases for cultural research[37] and for understanding our culture's view of the world at a given time. The scholarly record in any given area, viewed as a series of frozen artifacts narrowly spaced in time can be viewed as such a historical record.

The same issue applies to mass media. The daily, weekly and monthly publications of popular journals provide a nearly continuous chronology of the shifting perceptions of any number of cultural issues. The selection criteria for what is published are themselves a very important part of the cultural record, and represent very definite

371 am indebted to Professor Michael Buckland for illuminating this point.

biases (in some cases, one selects information sources precisely for the benefit of those editorial biases). Further, as information technology has made publishers more **agile,** as we have moved more to broadcast media (where only the present exists, in a real sense, and it is very difficult to go back and look at the media's content at earlier points in time) and as means of monitoring audience response have become more precise and more timely, content can be changed almost continuously in response to audience interests and preferences rather than reflecting a consistent editorial position Indeed, this content shift may take place hour by hour in the popular media: one can envision services such as the Cable News Network (CNN) shifting perspective from one broadcast of the news to the next (every half hour) based on viewer feedback and sensitivities .38

In a real sense, electronic information resources invite an Orwellian, a historical view of the world. Consider an electronic encyclopedia that is updated weekly or monthly; entries for countries and political movements are freely replaced. Rather than a series of views of events fixed at specific times, the entire view of the world is now subject to revision every week or two. There is no a priori reason why the implementation of such an electronic encyclopedia must ignore the past, but this is the simplest implementation; overlay the obsolete with the present.

Within the database management system community a concept sometimes termed "time-travel databases" has been developing; these are databases that can be viewed based on their contents as of a given moment [Stonebraker & Kemmitz, 1991]. As the database is updated, older versions of database records are retained, along with information as to when updates were applied and when information is replaced. Records are not actually ever deleted in such databases; rather, an indication is stored that notes that a given record has become invalid as of a given point in time. Such versioning or time travel databases are still at the research stage, however, and most commercial DBMS software does not support the necessary range of functions to allow production implementations of databases that incorporate a historical record of database evolution. Further, even if software becomes available, there are substantial costs in disk storage and retrieval efficiency that must be paid in order to provide historical views of database content. Libraries, facing continued financial pressures, will be hard put to justify investment in these technologies. Yet the ability to retrieve the state of knowledge or belief about a topic at a given point of time is an essential element of the historical scholarly record, and indeed a critical part of the data needed for a wide range of research endeavors.

The shift from sale and copyright law to contract law also raises issues where integrity and access combine in complex ways. Once a library purchased or otherwise obtained a physical artifact (for example, through a donation) that it made part of its collection, this artifact became part of the library's permanent collection. With the replacement of the transfer of artifacts by licensing of electronic information, it becomes much more difficult for a library to maintain early editions, erroneous distributions and other

[38] Continuous news broadcast services such as CNN currently modify about 6-8 minutes of their coverage from one cycle of the news to the next, dropping stories, adding stories, or making editing changes to stories that are repeated from one hour to the next (with these editing changes not necessarily being necessitated by new news developments).

materials that may be a part of the historical record which the publisher of a given information resource is not necessarily eager to have generally available. One need not assign any malice to the publisher wishing to withdraw out-of-date versions of their publications from circulation; the publisher may be doing this with the best motives, such as ensuring quality control. A pharmaceutical database publisher may want to ensure that incorrect or obsolete information (which may, indeed, be dangerous—for example inaccurate dosage data) is corrected promptly and comprehensively. More generally, a the publisher of an electronic newsletter may simply want to ensure that the corpus of published material is as accurate as possible; there is an inherent conflict between quality of a published corpus and the accuracy of the electronic publication as a historical record. The integrity of the historical record becomes far more subject to the desires of the publisher.

Earlier in this paper, copyright was identified as a potential barrier to access in the electronic environment. In the context of integrity, however, it can serve a very valuable purpose for authors by providing a basis for the author to ensure the integrity of his or her words over time, and preventing later "amendments" or "corrections" to published works. The right to make changes, like other rights (such as republication or translation rights) is subject to negotiation between author and publisher.

Of course, by the same token, one can imagine situations where a publisher (for example, a government or some other entity) uses its license control over material to effectively rewrite the historical record; certain material is simply declared "inoperative" and removed from circulation.[39] This is another illustration of the extraordinarily strong position of publishers, authors and other rights holders in the electronic information environment and the loss of public policy control of the balance between the rights of creators (or rights holders) and the public.[40] Ultimately, there may well have to be a rethinking of the definitions and meaning of publication; in the print world there is a strong sense that once something is published, some copies are distributed and available to the public permanently. Even in cases where a lawsuit is successfully brought against a publisher for one reason or another, while a result of the judgment may be that the publisher ceases to sell the work and destroys existing stock, there is really no practical way to recall copies already sold. Publication in the print world is generally viewed as an irreversible act;[41] at least under some definitions of

39 While slightly outside of the main focus of this paper, one area that I find particularly interesting and troublesome in this context is control of the news. The primary record of historical events is copyrighted material owned by newspapers and the broadcast media. As the various trends discussed in this paper lead to a situation where less and less of this material is held by libraries, it raises the specter of situations where for whatever reasons the primary historical record of events in a given area might well become inaccessible to researchers.

40 Copyright is not the only issue here, however. For example, an executive order was signed during the Reagan administration that permitted the government to reclassify previously declassified material in cases where they were able to regain control of all copies of the declassified document. In a print environment this is quite difficult; in an electronic environment it would be much easier.

41 Because of this irreversible nature of the act of publication, the scientific community has had to develop the practice of withdrawing a previously published paper that is later been found to be erroneous, for example; this is accomplished by printing a notice in a subsequent issue of the journal. This is not necessarily very effective, since a reader of the withdrawn paper may be unaware of its status. In an electronic environment, it is interesting to speculate how a withdrawn paper would be handled. Would it continue to be distributed, but bearing a prominent notice that the author has subsequently withdrawn it, or

"publication" this is not the case in the electronic environment. The new electronic environment is likely to create a great demand for what are perceived to be neutral parties to maintain the historical and cultural record (and I believe that most people view libraries as falling into this category), as well as various forms of audit trails so that revisions of the "record" can be tracked and evaluated.

The ability of a library to acquire access to data rather than copies of information resources also threatens the historical record of scholarship and culture. One vision of electronic information resources calls for publishers to mount databases of journal articles on the network, rather than supplying copies of the information to libraries that subscribe the these journals. In such a world traditional journal subscriptions are replaced by contractual arrangements that allow libraries to retrieve articles from the publisher provided servers under various terms (pay per view, or unlimited access to articles from the journal through a "subscription" price, for example). This may represent an economy for libraries, in that they do not have to receive journal issues, and they do not have to pay for local storage space to house these journals. But what happens when a publisher decides that a given journal (or specific back issues of that journal, or even specific journal articles that appeared in the journal) are no longer being used enough to make it profitable to provide access to the journal on the network? Or, perhaps, the publisher goes bankrupt, or is acquired by another publisher, or becomes entangled in litigation? In such cases it is quite possible that the publisher will cease to provide access to some or all of the contents of journals without notice, and without any recourse by the library community; the back issues simply become unavailable, for example. In the old print world, if a library somewhere held these back issues, they would still be available to patrons through interlibrary loan, but in the new electronic environment, unless some library had made local copies of the material (thus losing out on the economies that make electronic distribution of the matetial attractive) this material could be lost to the user community for all time.

Copyright law may again provide a useful tool for ensuring preservation of the scholarly record. Historically, deposit of a copy of a work with the Library of Congress has been a requirement for establishing copyright protection; while current copyright law has removed this requirement, to a great extent, a return to such deposit requirements could help to ensure the long-term accessibility of electronic material.

In the print world archival access to material was the responsibility of the library and archival communities. In a world of licensed access to electronic information, libraries cannot unilaterally continue to accept and discharge this responsibility. It may well be that the networked information environment will call for a new compact of responsible behavior between publishers and the library community, in which publishers make a

---

would distribution cease? In the OCLC/AAAS ***Current* Clinical Trials** electronic journal, the reader of any article that has had subsequent corrections or comments receives a very prominent warning that such supplementary information exists; however, it is in some sense **easy** for ***Current Clinical Trials*** to make such linkages visible to the user, since the journal is not simply distributed as content but includes a integral OCLC-supplied viewing **interface. Unless the historical record is actually altered in an electronic journal (at least to the extent of indicating as a note in an article that a correction or withdrawal notice was later issued, and when) electronic journals distributed as pure content (without a user interface to make such** links) **are** likely to offer only the same weak ability to notify users of subsequent corrections that characterize the print publishing world.

copy of their material available to some organization serving (and governed by) the library community so that the library community can assure itself of continued availability of material. Or a publisher might agree that if it removes material from availability on the network, it will offer this material to some access provider of last resort that is financed and governed by the library community (perhaps a network analog of the Center for Research Libraries for example). But the problem here is that while it is reasonably straightfoward to find solutions in an environment of cooperation between libraries and publishers in which all parties behave responsibly, there is the constant threat of irresponsible behavior on the part of publishers, or of external, uncontrollable events giving rise to the loss of key parts of the scholarly record.[42] National attention to the role of national libraries or other organizations in ensuring the preservation of, and access to, the scholarly record, is of vital importance in gaining the confidence of the user community in abandoning printed formats for electronic ones.

It should also be noted that there is another, more crass, issue that is raised by the transition to a networked information environment in which publishers are the primary providers of their inventory. Print is an inherently distributed medium, whereas in the electronic environment a technically inept publisher might stand to lose their intellectual property holdings through various types of catastrophe like fire, earthquake, or corruption of a network server by computer hackers. While the user community would not loose the rights to their material, practical access to this material might well become permanently lost.[43] From a business point of view it might mean that the publisher went bankrupt, but from the broader perspective of the scholarly community, it means that the material is lost and is no longer a part of the scholarly record. Given the numerous relatively small publishers, such as professional societies that issue one or two journals, loss of information due to failures of the publisher to adequately back up or protect their material should be viewed as a very real issue.

Natural disasters and business failures are not the only issues. As libraries move from providing access to their own local physical collections to a set of networked resources international issues must also be considered, for example. One can readily imagine situations where a national library in a foreign country provided access to the majority of the literature related to that nation, until suddenly some international political problem (an obscenity dispute, a war, a change of government, or whatever) caused that national library to cease providing the information in question. Even if there was no central point of control such as a national library access to information provided in one nation could be cut off for other nations by government action. Access may not just be interrupted; more subtle changes are possible. Imagine a fundamentalist religious government taking power in some nation; they might order the destruction of some

---

42 The issue of the scholarly record in electronic form should not be viewed in an entirely negative light. Today, in print, retractions and corrections probably rarely reach those who read the original article. In an electronic environment where we can track who has read (or downloaded) a given paper, the possibilities for disseminating retractions or corrections to the readers who most need to be aware of them is greatly improved.

43 Some publishers have argued that downloading in the Internet environment implies that there is always likely to be some copy of a publication stored on some individual's workstation, and that in this sense electronic publication on the Internet is also an irrevocable act. But, I would suggest that there is a great difference between continued access to material by some random member of the scholarly community and continued access by an institutional agency (such as a library).

materials (or at least the removal of these materials from an electronic information archive) and in other cases change indexing terms in such a way as to distort the functioning of traditional bibliographic access apparatus.

Access to the historical record is not merely an issue of ensuring access to the primary material. In a very real sense, as already discussed, the coverage of the abstracting and indexing services defines the literature of a discipline for many information seekers. But, in almost all cases, abstracting and indexing services began creating computer-processable records in the late 1960s or 1970s. Except for monographs covered in library online catalogs the literature prior to those dates is inaccessible through computer-based retrieval tools, and, for all intents and purposes, might not exist in the mind of many library users. In some fields, particularly the humanities and some social sciences, programs will have to be established to make the remainder of the scholarly record accessible in electronic form and thus place it on an equal basis with the recent publications that are abstracted and indexed by electronic databases. Considerable work is needed in establishing priorities for investment in both the abstracting and indexing of older print material and the conversion of the source material itself to electronic form; these priorities must consider both the mandates of preservation (creating electronic versions of material that is currently deteriorating because it was printed on acid paper, for example) and the programmatic demands of the scholarly communities that will use the materials.

## 7. The Impact of Micropublishing, Narrowcasting, and Information Feeds

Publishing is becoming increasingly fragmented under the dual pressures of advertisers and readers. For advertisers, a given publication venue becomes increasingly attractive to the extent that it can offer the advertiser a very specifically focused, homogeneous readership-families in the San Francisco bay area that make over $50,000 per year, that are employed by service industries, that collect stamps, and that are shopping for their first house. From the perspective of the reader, overwhelmed by the ever increasing flood of published information, publications that are highly specific to the subscribers interests are of much greater value than more general periodicals. Improvements in the technologies of composition and publication have facilitated this fragmentation to attract readers and advertisers. It is quite common today to see mass market periodicals published in a large number of regional and industry-specific editions; newspapers now offer a wide range of regional editions. Indeed, today's regional editions are composed out of common article databases, but the interest templates are established by region, industry or other criteria rather than individually. The inexorable march of technology seems to point towards ever greater specialization to the audience, ultimately all the way down to the individual reader obtaining a custom product; this trend is manifested in developments that range from the experiments conducted at the MIT Media Lab in the composition of "personal newspapers" based on filters to newsfeeds [Brand, 1987]. Apple's experimental prototype Rosebud system[44] which allows users to define and operate software agents under the metaphor of "reporters" that cover specific information sources [Kahle, Morris, Goldman, Erickson, &

[44] The information access model pioneered by the Rosebud project has recently resurfaced in the commercial AppleSearch product.

Curran, 1992b] or cable television narrowcasts (for example, the slotting of five minutes of local news into a CNN broadcast every hour). Other variations on this model are already coming into large scale use within corporations; for example, in financial and securities markets that are highly sensitive to news, real time news feeds are being licensed for distribution over corporate local area networks, and these news feeds are filtered to allow near-instantaneous delivery of relevant stories to the appropriate individuals very rapidly [Belkin & Croft, 1992; Goldberg, Nichols, Oki, & Terry, 1992; Marshak, 1990].

These trends create enormous problems for libraries, which will only become worse as more information is cast into electronic forms suitable for personalized retrieval. In the case of print, it is becoming very difficult to track what has appeared in the published literature, and who might have been aware of its appearance. Different libraries hold different editions of newspapers and magazines. Abstracting and indexing services do not typically cover all of the various regional editions of a publication; rather, they select one (sometimes without much consideration, and without being very clear about which edition they have selected) for indexing. As we move beyond print editions to the databases from which the specialized print editions (and future personalized extractions) are generated, the situation changes again. It now becomes possible for a library to obtain or purchase access to what is in essence the composite intellectual contents of all of the various editions, but the researcher coming to these databases years later may well loose any understanding of how items from these databases were selected (or, sometimes more to the point, not selected) and presented to any given audience or what the impact might have been on readers. The selection of available materials into a given edition represent a value judgment about the importance of specific subjects in time and space, and the record of this judgment is of critical importance to researchers. The key questions, as we attempt to mine these databases for research, will be what material a given reader of a given print edition learned from that print edition, and what material would likely have been selected from a given database at a given time by someone filtering the database with a given interest profile. To make matters worse, the filtering typically occurs (at least today) close to the end user rather than the publisher, and there is a great diversity of filtering tools ranging from rather simplistic keyword matching all the way through sophisticated research tools that perform semantic and linguistic analysis to identify material that may be relevant.

There is also a serious problem with the accuracy of citations while we remain in a transitional stage between print and electronic databases. n the fully electronic world, one might simply reference article ID X (added on such and such a date) in the Wall *Street Journal* article database. But today, this would most likely be referenced by its title, and the date and issue number of the issue that the article appeared in, without providing the critical bit of additional information—was it the East Coast, West Coast, European, or Far Eastern edition?

The recombinant nature of materials that are maintained in electronic databases but presented to the reading public through specialized printed vehicles is not limited to newspapers and magazines. McGraw-Hill has embarked on a program called PRIMUS in which they supply databases of articles that can be combined into course readers, with the inclusion of optional commentary or additional material by the compiler of the

course reader. Here again it is becoming increasingly difficult to determine the provenance of material, its timeliness, or the conclusions that the reader might have drawn from it based upon context.

The ready availability of desktop publishing (now being combined with transmission via fax or electronic mail across telecommunications facilities) has led to an incredible proliferation of very narrow audience newsletters; many of these have very high time value, very high subscription costs, and sometimes rather high impact on the communities they serve. Yet to those outside of these select subscriber communities, these publications are almost invisible. They are not available in libraries, except occasionally for specialized corporate libraries (which often will not circulate the material through interlibrary loan), and material in them is not indexed in commercial abstracting and indexing services that researchers would use to try to obtain information about a subject, or to reconstruct events after the fact.

A surprising parallel can be drawn to the infamous "gray literature" of technical reports in some areas of science and engineering; these are poorly indexed, difficult to obtain documents that play a key role in circulating information within an insider community, but are not readily accessible either through libraries or through the bibliographic control and access apparatus of abstracting and indexing databases. Yet, for an active researcher, they are invaluable, and the information propagated through such technical reports can have an extraordinary impact on a scholarly community. The contents of the technical reports will appear in the traditional, "archival" published literature only years later in many cases, and often in an edited (reduced) form due to page limitations in the archival journals. In some disciplines, important work that is described in technical reports does not always make its way to the mainstream literature, with the authors simply viewing it as too much trouble to manage a manuscript through formal submission, peer review and publication when the material is already somewhat stale and dated to the authors. In some academic disciplines, it is probably not too strong a statement to argue that were it not for the need to publish in traditional archival print venues for tenure and promotion purposes, most of the effort would go into producing material that would be part of the gray literature. Increasingly, in part as a recognition of the importance of such material in scholarly communication and the diffusion of research, such reports are being made accessible through the Internet from file transfer (FTP) or GOPHER servers, most often through initiatives at the departmental level. Of course, this leads to an essential literature that is very hard to obtain access to, particularly for those who are not comfortable with the relevant technologies, and the long term accessibility of this literature is questionable, since there is today little institutional commitment to ensure continued availability of it; partially in a response to these trends ARPA is funding a major project to improve access to technical reports in computer science through the network. 45

---

[45] If anything, the scope of the gray literature is becoming larger and more confusing as we move into an electronic environment when anyone can make material directly available for anonymous FTP from a personal workstation or departmental storage server. Not only are members of the academic community mounting technical reports but in some cases also the text of material that they have published, often in violation of the transfer of copyright agreements that they have executed with the print publishers of the work as a condition of publication. Thus far, this has occurred only at the level of individual authors and not on an institutional basis and to the best of my knowledge the print publishers have not attempted to enforce their control of the rights with these individuals, perhaps feeling that it is not worth the expense or fearing that such an action would mobilize the community of academic authors to pay more serious

Over the past few years, the research library community has begun to devote resources to improving access to the technical report literature as part of a recognition of its importance. Several major research libraries have developed extensive technical report collections, and Stanford University has developed an extensive bibliographic database of technical reports which is available on the Internet. Yet the newsletter literature remains almost inaccessible to the academic community and the general public.

While somewhat outside of the primary focus of this paper, mention should also be made of real-time information feeds that go beyond human-generated intellectual property such as newswires. Other types of real time information feeds are beginning to appear on the network; these information feeds, just like newswires, can be analyzed by personal workstations to sift out events of interest. If ever knowledge was power, these applications illustrate the axiom; in some cases, the advantage of timely knowledge goes beyond financial or professional gain to matters that are literally life and death. The ownership, access to, and integrity of these resources is of potentially critical importance, and may in future become a central public policy issue. Consider the following examples of information feeds that might pass over the network:

•Newswires. These are clearly intellectual property (representing reporting of news events) yet timely access to news may permit an individual to obtain substantial financial or professional gain. Related to newswire feeds will be a wide range of real-time audio and video telemetry of current events of interest, similar to what is sometimes found on the C-SPAN cable television network or carried on current experimental services such as Internet Talk Radio. The intellectual property rights involved in these audio and video feeds is far less clear; in some cases the events are public (for example, Congressional hearings) and in many cases they are simply captured through fixed video cameras and microphones, thus involving little creative work.

---

attention to copyright transfer agreement terms. Often, it is extremely hard to tell precisely how the electronic version available from the author is related to the printed publication; in some cases, the electronic version does not even provide a citation to where the work appeared in print. And, of course, the electronic version is often available months or even years prior to the availability of the published printed paper. Another interesting area of expansion of the gray literature is doctoral dissertations. Most universities use University Microfilms Inc. (UMI) as a means of ensuring general availability of theses, and request (sometimes essentially require) doctoral candidates to file their dissertations with UMI as part of the process of filing the dissertation. The UMI agreement gives the author non-exclusive print rights, including the right to publish articles or books based on the thesis, but reserves *exclusive* electronic distribution rights to UMI. Yet theses, or versions of theses that are issued as technical reports by academic departments are starting to appear for public access as well. Again, it is difficult to tell if these are precisely the same as the formal filed theses; the print versions that one obtains from UMI normally include the signed cover sheet and are identical to the versions filed with the university. H should also be noted that while the incidents discussed here of making dissertations (or variants of dissertations) available electronically are isolated individual cases, there are also institutional level discussions going on between a number of universities and UMI under the auspices of the Coalition for Networked Information about how to make dissertations available electronically through the network and what (if any) UMI'S role should be in such an enterprise. These discussions might well result in a situation where UMI serves as an electronic publisher for dissertations on behalf of the university community (in print, UMI is a very cost effective means of providing access to dissertations, with a typical dissertation only costing about $35); alternatively, it might result in universities negotiating for changes to the UMI copyright transfer agreement and mounting dissertations themselves.

•Stock or commodities prices, or other financial data (for example, purchases of rare coins). It is much less clear that this is intellectual property rather than brute facts, but again timely access to and analysis of such information can offer an individual substantial advantage. Currently, one can obtain subscriptions to the stock market trading "ticker" (at considerable expense).

•Time synchronization information. The US Government, as a public service, offers very precise time information suitable for setting clocks in a distributed computing environment so that multiple computers can operate on a common time scale. (These efforts are supported by NIST, based on programs that date back to when NIST was the National Bureau of Standards). A related service is the Global Positioning System (GPS) satellite network, which allows a user to locate the position of a receiver with a tremendous degree of accuracy. GPS was deployed to support military applications, but has extensive civilian uses that include search and rescue, navigation (for planes, boats, and even potentially automobiles or individuals on foot). Currently, such information is publicly available, and is of substantial value; will private commercials services have a role here in future?

•Weather information. Much of this is collected from sensors (obseving stations, satellites, etc.) financed through public funds. Yet subscriptions to weather information are in many cases through private information services (though the University of Michigan, for example, makes information derived from such private services available to the general public across the Internet through the Weather Underground service). In some cases this information is merely interesting (e.g. whether it will rain today); in other cases it is of financial value (implications for commodities prices). But in a few cases—tornado watches, flood warnings, and the like—it is a matter of great importance to recipients, if they can obtain the information in a timely fashion and act upon it.

•Surveillance data. We are awash in digital imagery ranging from data generated from satellites (either government or commercial, such as the French SPOT system) through output of surveillance cameras in our workplaces and homes. Again, awareness of such information can have value ranging from insight into commodities or securities investment through convenience-traffic congestion information, for example—to personal safety.

•Earthquake warnings. Earthquake shock waves propagate from an epicenter rather slowly compared to the near speed-of-light propagation of information across copper or fiber optic trunks. Some states, such as California, have sensor networks deployed which could, at least in theory, propagate information about the occurrence and epicenter location of earthquakes across the Internet in such a way that locations that will be hit by a major earthquake might obtain as much as 60 to 90 seconds advance notice of the event. This is information which, if identified and acted upon in a timely fashion, could save not only a great deal of money (for example, by "safeing" processes ranging from parking heads on disk drives or stopping elevators in a graceful fashion at floors through shutting down industrial operations such as chemical refineries) but could also save lives. It is unclear who would own such information, or could provide it to the network.

Ownership is not the only issue with such information or other sensor feeds; integrity is equally important. It is vital that such information be not only correct but also accurate and authenticatable. The notion of someone "simulating" a major earthquake through the network, for example, is clearly unacceptable. In this connection, it is interesting to note that most of the existing cryptography based authentication technology could be quite problematic in these applications, unless considerable care is applied to address scaling issues; for example, imagine every workstation in Northern California simultaneously trying to obtain the public key of the earthquake information server to validate an earthquake warning, throwing the entire Internet into overload at precisely the time that smooth operation is most needed.[46]

## 8. Privacy issues in access in the networked information environment

### *Confidentiality and Anonymity of Access*

Within the library community, confidentiality of collection use information is a well established principle. Library practice, as defined by the American Library Association, defines circulation records as highly private, to be revealed only under a court order—if then. Indeed, practice goes further-typically a library will only store the information that a given patron has borrowed a given book during the period while the book is on loan; once returned, only statistical information based on the patron's statistical categories is retained for subsequent management analysis. Most libraries, even under court order, can provide little or no history about the books borrowed by a given patron. Through statistical analysis, they may be able (if their circulation system is well designed and well implemented) to provide lists of the hundred most popular (in the sense of most frequently borrowed) books, or the ten books borrowed most often by high school students in the past year. It is also usually possible to find out how often a given book has circulated, or how many items a given patron has borrowed in the past year. In fact, such information is very important for the tuning of collection development strategies, for deacquistions decisions, and for overall management and budgetary planning.

Similar principles about privacy in the networked environment are far less clear; there is no general consensus about norms of behavior. Most users have a tendency to assume that their privacy is protected-either by legislation or generally accepted practice-to a greater extent than it probably really is, perhaps making this assumption by analogy to library practices and other situations, such as video rental records.

---

46 Another example of the use of the network to provide information to control machines is provided by the "smart power" technologies that are under discussion in projects such as the Blacksburg Electronic Village effort in Virginia [Bull, Hill, Guyre, & Sigmon, 1991]. The basic idea here is that under heavy load the power company must purchase additional power from other power companies on the national electrical grid at very high prices and it is very advantageous to them to be able to reduce loading during those times; additionally, their pricing, particularly to residential customers, does not let them recover these premium costs directly during periods of very heavy load. Instead, residential costs are to some extent averaged over long periods of time in setting rates. The proposal is that consumers would install smart appliances and controls (thermostats, refrigerators, air conditioners, etc. ) that would be connected to the network. During periods of heavy power demand, the power companies would broadcast alerts through the network and these devices would reduce power consumption temporarily. Apparently, preliminary studies on the Blacksburg project have suggested that if the power company actually paid for smart thermostats (assuming that the network infrastructure was in place) they would recover their costs within two years.

Service providers, including libraries operating online catalogs and institutions supporting anonymous FTP archives, have little legal or policy guidance and view the situation with a considerable degree of unease.[47]

If one considers libraries, which at least have a historical policy context that might help them to develop policies for privacy in access to new online information services, one finds a variety of practices and motivations behind them. Many online catalogs provide anonymous access simply because it is easier than having to maintain user files for a large, and, in a university, rapidly changing user community, and not because of any policy commitment to the right of anonymous access to the online catalog (as distinct from a possible policy position on the right to privacy of searches; in other words, the library is saying that it will protect privacy, perhaps, but not providing the absolute guarantee of privacy that anonymous access gives to the user). Some institutions controlled access simply as a means of regulating resource utilization; these controls sometimes required users to identify themselves through user IDs and in other cases preserved anonymity by using controls such as originating network address to determine whether a user had access. As online catalogs have expanded to include abstracting and indexing databases and other electronic resources licensed to specific user communities, it has become necessary for many systems to implement some type of user identification mechanism in order to control access to these licensed resources in accordance with the license agreements. A few institutions, such as the University of California, have developed approaches to user identification that provide for patron anonymity, but many have simply gone to a user ID mechanism, often based upon the library card number. To some extent the questions about accommodating anonymous access tie back to the library's overall priorities; in a period of intense pressure on budgets and library resources, many libraries are articulating strategies that place priority on serving their primary clientele (for example, members of a given university community) and provide access to other users on a lower-priority basis.[48] Members of the primary clientele will typically be registered with the library and this registration process provides them with user IDs that can be used for authentication (thus shifting the issue from guaranteed confidentiality through anonymity to policy confidentiality provided by the library). As online catalogs grow into ever richer and more complex mixtures of public and restricted access licensed information resources, it is much simpler to abandon the attempt to provide anonymous access when feasible and move towards a uniform authenticated access model, which is less problematic for the

[47] T. a great extent library patrons are protected more by practice than by law; libraries do not collect information such as what books a patron has borrowed beyond the point that he or she returns them. There may be some statistical information used for collection development that links demographic characteristics of patrons to borrowing records, but the old sheet of paper or card in the back of the book in which the names of those people who have borrowed the book over the years has largely been eliminated by libraries, at least in part in response to growing concerns about patron privacy. In the electronic environment, we may see a clash of cultures; telephone companies, for example, typically gather and retain very detailed records of who each customer has talked to and when; these are easily accessible with a court order.

[48] This is also being done for networked information resources; for example, some FTP sites limit access during the day by users outside of a specific set of network addresses that are viewed as defining the primary user community, or the limit the amount of anonymous traffic to ensure that resources are available to serve the primary clientele.

primary clientele than to other outside users who wish to make occasional, casual use of the library's information system through the network.

There are other reasons why online catalog designers are moving towards (at least optionally) authenticated access as online catalogs become more sophisticated [Lynch, 1992]. This is needed for current awareness services that electronically mail notification of the arrival of interesting new materials to users, for intelligent interfaces that track a user's history with the system and his or her preferences, or that tailor the dialog to the user's familiarity with the system based on how long and how often he or she uses it. Again, it is certainly possible to support both authenticated and anonymous access modes, and even to permit users to store preference files external to the library information system, importing them when they start an anonymous session and exporting them again at the close of the session, but all of these options add considerable complexity to the system design, the cost of which is certainly subject to question, particularly in the absence of any policy or community consensus that underscores the importance of offering an anonymous access mode.

Matters are complicated by several conflicting sets of demands on service providers. Indeed, this conflict goes beyond operational needs to a basic conflict of values between the library community's tradition of free access to information and the computer community's emphasis on tracking, auditibility and accountability. Computer and network security practices stress the importance of audit trails and other monitoring tools to protect systems from intruders, and system security and integrity are major issues for any service provider on the Internet. In fact, there seems to be consensus among many of the regional network service providers that anonymous access to the Internet is unacceptable (for example, providing access to terminal servers that can TELNET to any Internet host without first identifying the user at the terminal server so that attempts to break into system can be tracked back to an individual at that institution—but note here that there is no requirement that the information be propagated outwards from the source terminal server, only that it be maintained so that by cooperation among organizations a trail can be defined back to an account at the first institution). Certainly, there are many systems that permit anonymous incoming access, but in order to satisfy these restrictions they limit access going back out to the network to specific, limited sets of hosts that have agreed to permit anonymous incoming access. For applications where there is recharge for information access, careful tracking of users is needed to allow discrimination among user groups. This question of anonymous access to the network has sparked bitter arguments between the library community and the networking community, as many libraries view themselves as potential access points to the Internet, and at least some libraries have taken the position that they should not have to require users to identify themselves in order to access resources on the net that are willing to accept these incoming connections. It seems likely that as "public-access" resources on the network multiply, and particularly as federal, state[49] and local government information becomes more

[49] In California, Assembly Bill 1624 is currently under consideration, which, if adopted, would require that various legislative information be made available at little or no cost to the general public through the Internet. One serious proposal by some members of the legislative staff is that the identity of those members of the public requesting this information be tracked for various reasons.

commonplace that the conflict between security and the right to anonymous access will continue to be troublesome.

Many of the information services being offered on the Internet are viewed as somewhat experimental; indeed, we are all still learning how to build user friendly and effective information retrieval and navigation tools, and analysis of user sessions is a key tool in improving the quality of such systems, as well as more routine tuning and capacity planning efforts that are part of the operation of any large scale service. Finally, it is important to recognize that not all information providers on the Internet are institutional; for example, it is quite common to find academic departments, research groups or even individual faculty members setting up anonymous FTP directories to permit people to obtain copies of their papers and research reports. They view this as not much different than responding to postcards asking for offprints or orders for technical reports, and retain a natural curiosity about who is reading their work (which was evident in the days when they responded to requests for printed copies).

Ironically, part of the problem is the development of the distributed computing infrastructure. Ten years ago, when online catalogs were initially being deployed by most libraries, access was primarily from within the library, or perhaps from a few large timesharing hosts on a university campus; if the library was recording searches, it would typically only know that a given search came from terminal 43 within the library or from machine X (which might have 500 registered users). The identity of individual users accessing resources on the network was effectively hidden behind these large multi-user timeshared hosts, and, while a given network resource might require a user to identify him or herself in order to use that resource, the user was aware when such a request was issued by the remote system-one was asked to log in, or provide a password. Very little information about the identity of individuals accessing a remote service could be determined autonomously by the remote service; if the service offered anonymous access (that is, it did not ask for information from the user accessing it) then the user could have a reasonable degree of confidence that access really was anonymous (barring collusion between the user's local host and the remote host; statistical analysis of who was logged onto the user's local timeshared host in comparison to when a remote service was accessed from that timeshared host could, over time, probably allow a sufficiently interested analyst to trace accesses back to individuals, but such activities are rare, and most users view them as too much trouble to represent a serious threat to anonymous access). As we have migrated to a world of personal workstations, the origin address for a search (or a request to fetch a copy of an electronic document) is linked to a specific host address, and increasingly this host, which is now a workstation, is now in the service of a single master. In the new networked environment, the source of a query or a request suddenly provides a great deal of information about the identity of the individual issuing that query or request. This should not be narrowly viewed as a matter of personal privacy; in fact, in the network environment, it is often hard to identify an individual but easy to identify the individual's organizational affiliation by the network number in the incoming Internet address. While people outside of organization X may find it hard to determine that a given address is person Y's workstation, everybody can tell that the access has come from organization X. This may be a matter of competitive intelligence rather than personal privacy.

Certainly there are technological solutions to the problem of one's address revealing one's identity. The simplest is to carry forward the time honored method of mail drops (post office boxes, or the mail forwarding services that various newspapers have long offered in conjunction with personal advertising). Electronic mail based dating services offering such anonymity through the agency of a mutually trusted third party are already operating on the network; a similar service could easily be set up for TELNET. But, as the number of protocols multiply and distributed system architectures become more complex, the development of general purpose anonymity services will become quite problematic. Further, one must wonder whether the vast majority of users will recognize when their use might be appropriate; the example of dating services is a good one since it is simply a recreation of existing practice in the electronic environment in a fairly direct way, and consequently its use in the electronic environment is appealing to the same people who would likely have used it in a non-electronic world. Whether users will recognize the new risks introduced by the development of new electronic information services, or the redesign of old services for the electronic environment remains an open question.

We are only beginning to explore the challenges that distributed computing raises for individual privacy in the context of "anonymous" remote terminal access becoming increasingly easy to trace back to an individual as more users use their own personal workstations rather than large timeshared hosts. At least in the remote terminal emulation environment—be it TELNET or more modern X Window system based applications-the user employs widely available, well documented, industry standard utility software that is written according to publicly available specifications and which can be used with a very wide range of remote services. Often, software to implement protocols like TELNET and the X Window system is available from multiple sources for a given platform (both commercial software suppliers and public domain or "shareware" sources). While there are some true distributed client-server protocols that are well documented national or international standards, such as the Z39.50 information retrieval protocol, and these protocols are implemented in multiple client software applications that can again be used with a wide variety of remote servers, in the developing client-server oriented distributed environment we will see providers of information services implementing custom software clients. These clients will be distributed to users in executable form only; they will employ proprietary protocols, and will be needed to obtain access to specific information servers. In essence, the user of such a service is expected to execute a program of largely unknown function *which typically has full access to the files on his or her personal workstation, given the current state of the* art *in the operating system* software *that runs on most of these workstations,* and which opens and uses a communications channel to a remote service as part of its normal, expected behavior. This is already the case in some commercial services, such as Prodigy [Burke, 1991].

The opportunities for collection of information are endless; for example, such client software might upload a list of what software the user has installed on his or her hard disk,[50] or the list of USENET newsgroups to which the user is subscribed.[51] Unlike

[50] Lists of software installed on machines is useful not only for marketplace research or marketing demographics (for example, to identify people who might be interested in add-on software to an existing product or in competing products) but for other purposes like identifying illegal copies of software: a

general purpose utility software (for example a TELNET-based terminal emulator), the covert information collection activities of specialized client software may be very difficult to identify and monitor,[52] and while very sophisticated users or institutions may be able to address this problem legally with the supplier of the service (and the client software), most users will likely remain unaware that the problem even exists. We may see organizations giving away client software and access to certain remote services through that client software just to be able to get users to run the client and unwittingly export information that the service provider can use directly or resell to others. We may find a direct contradiction between realization of the distributed computing potentials of the Internet and individual user privacy.

There is another interesting relationship among pricing, privacy and the capabilities of systems supported by information providers in the distributed computing environment. Currently, information providers frequently charge based on the amount of information that is exported from their systems; they offer filtering tools of varying degrees of sophistication. On a purely technical basis, some users of some system choose to do fairly unselective extractions from the information providers and then do ranking and filtering on their local machines; this has the effect of preserving some privacy (since the fine details of what the user is interested in are not conveyed to the information provider) but also tends to run up a large bill since the information provider assumes that everything that is exported is of value to the user and will probably actually be examined by the user, rather than filtered by a computer program running on the user's machine. As information provider capabilities improve, the decision as to how much information to give the information provider in order to permit the provider to perform filtering will likely be based in part on how specifically the user is willing to reveal his or her interests to the information provider; privacy (gained by the method of asking vague questions) will have a price. Balancing this, however, we should note that the trends in technology are towards user clients that act as integrators for *multiple* information providers, not just one providers, and such an integration function obviously cannot be pushed outwards to the providers, since no individual provider has the full range of information necessary to do the ranking and filtering of information from multiple sources.

---

company making multiple products could use one to scan for the presence of copies of others, and then check its registered user files.

51 The suggestion that a local client could exploit information about a user's subscriptions to USENET newsgroups is due to Simon Spero, although he proposed it in the context of client software using this as hints in developing a user profile which could be used to help tune information retrieval applications, and not as a mechanism for invasion of privacy.

52 Many personal workstation operating systems can now be equipped with virus protection software which can detect and warn the user of unexpected *modifications* to critical files on the user's machine, but I have never seen one which monitors access. The user does have some countermeasures, such as keeping critical files on a separate disk and never mounting that disk while running software that he or she does *not* trust, or encrypting critical information when it is not being used, but the cumbersome nature of these measures makes them impractical outside of very high security environments with very security-conscious users.

### *Who Owns Access Histories?: Privacy and Market Research*

The analysis of consumer behavior has become a major focus of attention in the business world. Supermarkets have on the one hand implemented laser scanners that track the products being purchased by shoppers (and linked them to systems that automatically issue a set of custom tailored discount coupons at the checkout register) and on the other hand now encourage payment with credit cards, allowing the development of databases that track consumer purchases in tremendous detail [Mayer, 1990]. Companies like American Express that have access to extensive histories of customer spending practices and preferences are now marketing finely tuned customer lists to their business partners-for example, I might receive mailings from American Express that offer me airline upgrades on airlines that I don't fly regularly, based on statistical analysis of my purchasing profile which indicates that I spend over $25,000 per year on airline tickets and that none of these charges go to certain airlines. Similarly, in many industries there is now an intense focus on what goods are selling, and in what marketplaces, and this information is employed in very complex pricing decisions (consider again airline seats as an example.) The practice of "data mining" from customer histories has begun to be viewed **as** simply effective exploitation of previously untapped corporate assets [Piatetsky-Shapiro & Frawley, 1991]. In addition, we are now seeing considerable use of multi-source data fusion: the matching and aggregation of credit, consumer, employment, medical and other data about individuals. I expect that we will recapitulate the development of these secondary markets in customer behavior histories for information seeking in the 1990s; we will also see information-seeking consumer histories integrated with a wide range of other sources of data on individual behavior.

The ability to accurately, cheaply and easily count the amount of use that an electronic information resource receives (file accesses, database queries, viewings of a document, etc.) coupled with the ability to frequently alter prices in a computer-based marketplace (particularly in acquire on demand systems that operate on small units of information such as journal articles or database records, but even, to a lesser extent, by renegotiating license agreements annually) may give rise to a number of radical changes. These potentials are threatening for all involved. Consider just a few examples:

•For the first time, libraries should be able to **easily** collect reliable data on how often specific journals are read, or even the pattern of access to specific articles within these journals. This information can be used to decide not to subscribe to journals, which worries the publishers.

•Publishers can employ this usage information to set prices on journals or even specific journal articles based on popularity. This leads to price instability, which worries the libraries.

•While citation data as a measure of the impact of a publication has been controversial (though it is already considered in tenure and promotion decisions at some institutions) usage data is less ambiguous; if nobody reads a publication, it is unlikely to have had much impact. This is of great concern to authors.

•Usage data makes it much easier for authors, publishers and libraries to rapidly reflect the short-term interests of the user community by keeping track of what is popular and trying to produce or obtain more of it.[53] To some extent, this is at odds with the development of the scholarly record and the integrity of scholarship. Archival publications are not necessarily read a great deal, but some would argue that it is of vital importance that they continue to exist.

•There is a tendency in systems that stress popularity to ultimately reduce diversity; if everybody else is reading something, then one concludes that one needs to read it also. The temptation to select as one's reading the ten most popular articles of the week is very dangerous to the development of a diverse body of ideas. It is also worth noting that producers of abstracting and indexing databases are increasingly considering the subscription patterns of libraries in deciding what journals to cover; this seems to make the databases more marketable. If these producers were to emphasize heavily read journals, these abstracting and indexing databases will tend to become less comprehensive guides to the literature (and, indeed, pathways to material in less well known journals).

•There is a danger that the system of statistics collection can be manipulated by those that understand it. This can range from authors repeatedly accessing their own works to get their statistics up through more sophisticated approaches (for example, including many popular keywords in an abstract even if they have little to do with the actual subject of the work so that many people will retrieve and view the work).

These examples have emphasized applications of data about the use of information resources such as viewing or downloading journal articles. However, the availability of searches is also of great value: it tells information product designers about the kinds of information that people are looking for, and also the means that they are using to locate it. This is invaluable market research data for designing new information products, and for marketing and improving existing ones.

The ability to collect not only information on what is being sought out or used but also who is doing the seeking or using is potentially very valuable information that could readily be resold, since it can be used both for market analysis (who is buying what) and also for directed marketing (people who fit a certain interest profile, as defined by their information access decisions, would likely also be interested in new product X or

---

53 It is interesting to note how each technological development that undermines privacy seems to be complemented by a technological countermeasure that supports privacy. Consider the case of pay telephones. At one time, these were a wonderful way to obtain anonymity; one simply deposited cash. and the source of the call was untraceable. Now, of course, most **pay** phone users are using credit cards because they are so much more convenient, not realizing that if they use these cards all their calls can be tracked in great detail. (In fact, many public phones will not even take cash anymore, due in part to the expense of collecting the cash and the fact that the cash is an invitation to vandalism. ) In France, vendors now offer a phone card which has a specific "cash" value; one pays cash for it, and it is debited as one makes phone calls using it. This is a form of "electronic cash" which facilitates anonymity. (It also has some other important advantages; for example, while one **can** loose the card, one cannot incur the virtually unlimited bills against one's account that can be caused by a stolen phone credit card number. ) Another example of this technological balance is the development of Caller ID facilities by the phone companies; these were quickly complemented by facilities that allowed a caller to block the display of the Caller ID to preserve anonymity.

special offer Y). While such usage (without the informed consent of the recipient of the advertising) may well offend strong advocates of privacy, in many cases the consumers are actually quite grateful to hear of new products that closely match their interests. And libraries and similar institutions, strapped for revenue, may have to recognize that usage data can be a valuable potential revenue source, no matter how unattractive they find collecting, repackaging and reselling this information.

Competitive intelligence is a burgeoning field promoted by any number of consultants. One aspect of competitive intelligence is knowing in what areas competing corporations (or, in academic world, research rivals) are seeking information. For example, it is valuable to know, if one is a coporation in the pharmaceutical industry, that a competing corporation is seeking articles about the effects of a given drug. Of course, once one recognizes that one may be a target of competitive intelligence, it is possible to deliberately offer disinformation that will lead the competition to an incorrect assessment of one's interest, and even deliberately send a competitor down false trails. Clearly, the type of information that can be collected about information seeking and use in the networked environment is invaluable for competitive intelligence. And it is worth noting that even fairly highly aggregated information can be of value in a competitive intelligence activity: for example, from the aggregated article access information for a given university (without any indication of who within that university accessed the material) it is quite reasonable to draw conclusions about the research directions of specific research groups that are very likely to be correct.

Some of these examples seem farfetched. But consider a number of trends. As electronic information providers license information rather than simply selling it, they can *require* usage reporting as a condition of license.[54] This is done in other areas. Libraries and universities are both aggressively seeking new ways to generate revenue; the resale of statistics about electronic collection use and/or searches, particularly if they can satisfy themselves that some level of privacy is being maintained by not including the identity of the user (if they know it) could be a very attractive revenue source. It is unclear whether these institutions have any legal obligation to ensure confidentiality of this information.[55] If one signs a contract with a commercial information service such as Dialog, issues of confidentiality can be negotiated in advance as part of the contract; but when one is accessing (anonymously or otherwise) a public-access information service, it is unclear what to expect, and in fact at present there is no way to even learn what the policy of the information service provider is. As the secondary markets develop it is even conceivable that when accessing a for-fee resource one might pay more for privacy, and that when accessing a public-access resource the user's client and the server for the public-access system might well negotiate various levels of confidentiality (no logging, statistical compilation only, actual

54 Such conditions could be imposed either directly on a library licensing the information for local mounting, or, less visibly, thorough contractual constraints on a third party such as Dialog or OCLC that makes the information available to the library community.

55 Resale of information about who is searching what type of information is not the only issue. A financial information service might be a good investment for a brokerage house or a merchant bank for example; they might internally exploit knowledge that could be gained from records of what customers were searching information about what companies or products.

text logged for searches but without ID, no resale or reuse outside of the internal operations of the information provider, etc.)

A final aspect that should be mentioned is that in the print world the library served as a shield for its user community in the sense that it purchased materials such as journal subscriptions. The act of purchasing and the cost of purchase might well be an act of public record discoverable under a Freedom of Information Act. But only the library knew who was using the material, and that information (in the form of circulation records) was protected. Further, because information was acquired in highly aggregated forms such as an annual subscription to a journal, the act of purchase revealed very little about the interests of the library patrons—indeed, there is no a priori reason to assume that purchasing decisions are always directly driven by the short term needs of specific patrons in the case of a research library. Now, consider the electronic world, where a library frequently acquires rather specific information (such as a single article from a journal) in response to the specific request of a user. This purchase, as an external business transaction, may be a matter of public record. Further, if the end user rather than the library as intermediary acquires the article, it may be possible to rather directly link information use to individuals or departments. The electronic information environment may well call for considerable reassessment of the definition of public records, particularly in the context of state universities, as these records are defined by federal and state Freedom of Information Acts.

The uses described for information about searches and access patterns are simply extensions of well established practices such as market demographic analysis and competitive intelligence. New uses for this information, unique to the networked information environment, are also being researched. For example, Professor Mike Schwartz at the University of Colorado has been exploring the concept of resource discovery-automated methods of discovering or locating potentially interesting network resources [Schwartz, 1989]. One of the techniques that he has studied is the examination of access pattern of other members of a user% affinity groups; for example, a botany professor might be interested in resources that other members of the botany faculty are utilizing regularly but that he or she is unfamiliar with. Such research may ultimately lead to new tools for locating information resources which will call into question the appropriate balance between privacy, competition, and cooperation in various communities.

### *Privacy, Intellectual Property and Electronic Mail Enabled Communication*

Electronic mail based discussion groups—sometimes called electronic journals in academic circles if their editorial policies parallel those of traditional printed journals —have become extraordinarily popular on the Internet. These fall into two major categories— LISTSERVs and mail reflectors. Mail reflectors are simply special user IDs; when one sends electronic mail to such a user ID it is redistributed to the subscribers of the mail reflector automatically. Maintenance of the subscriber list is typically done manually or semi-automatically, with the convention being that if the mail reflector's address is of the form user@ hostname then there is an additional mailbox in the form user-request@ hostname to which requests to join or leave the mail reflector are directed. LISTSERVs are based on a program that was originally developed for the IBM

Conversational Monitor System (CMS) environment.[56] There are thousands of such mailing lists on the Internet; in addition, many such lists are reciprocally gatewayed to Usenet newsgroups, which are essentially very large collections of publicly readable electronic mail messages that are propagated through the Internet (and beyond), organized by topic (newsgroup name). Like electronic mail lists, some Usenet newsgroups are moderated; others are completely open. Unlike directly electronic mail enabled services (LISTSERVs, LISTSERV imitators, and mail reflectors), Usenet newsgroups do not appear to the reader as electronic mail, but rather as continually updated databases that are viewed through a program such as a newsgroup reader; electronic mail only comes into play when a reader wants to enter a Usenet Newsgroup discussion by posting a message. The privacy and intellectual property aspects of these mailing lists and newsgroups are very interesting, and probably largely ignored by most participants in them.

Some lists deal with topics that are controversial—for example, the Usenet news group ALT.SEX.BEASTIALITY or the recently established LISTSERV on gay and lesbian issues in librarianship, to mention only two examples.[57] Currently, USENET newsgroups offer a moderate degree of privacy; a newsreader running on a client machine uses a protocol called NNTP (Network News Transfer Protocol) to pull list postings down from a local NNTP server to the user's client. The list of newsgroups that the user is interested in is stored on the client, and one can find out what a given user is interested in only by looking at his or her preference files on that client machine or by monitoring data transfer from the NNTP server, both of which are relatively difficult.[58] LISTSERVs, on the other hand, require interested parties to actively subscribe in order to receive electronic mail that is posted and include options which allow anyone to look at who has subscribed to a given list (except for those users who have explicitly chosen to conceal their identities; these individuals are invisible except to the system administrators or list administrators). Of course, the vast majority of LISTSERV subscribers are blissfully unaware of the fact that their identities can be easily discovered, or of their option to conceal their identity. It is only a matter of time, in my

---

56 More recently, software has been developed for the UNIX environment which emulates most of the functions provided by the LISTSERV program in the IBM CMS environment.

57 There has been enormous controversy about the appropriateness of various universities making such newsgroups available to their communities; these have been well documented in the Computers and Freedom mailing list postings by Carl Kadie. To my mind, these controversies help to illustrate the gulf between the library tradition of not only intellectual freedom but of free access to information and the values of the computing establishment. It seems likely that if these were print works that were owned by the libraries of the Universities in question there would have been little debate about the right of these libraries to own them as part of their collection and to make them available to the university user community; this would have been a clear case of intellectual freedom on the part of libraries. But when such resources are made available thorough institutional computer systems (where there is little philosophical basis established for determining appropriate content) major controversies quickly erupt.

58 Interestingly, traffic analysis between client and NNTP server is probably easier than breaking into the file system on the client, and this can be viewed as another illustration of the way in which the deployment of the distributed computing environment has exposed individual's activities to much more scrutiny. If the client is on a large time shared host then it is not clear why specific newsgroups are being transferred to that timeshared host; if the client is on a personal workstation, however, it is relatively easy to *assign* responsibility for the transfer of material from a specific newsgroup. There is also a program called "arbitron" written by Brian Reid which publishes regular statistics about the usage levels of various newsgroups; this is again, I believe, based on traffic analysis techniques.

option, before some organization begins to make use of LISTSERV subscriber lists as a means of identifying groups of individuals that the organization wishes to communicate with. In a very real sense, one can view the LISTSERV system as a very public definition of the interests of many of the individuals on the network today. Put simply, one monitors a Usenet newsgroup, and the fact of that monitoring is between the user and the local Usenet distribution host; one subscribes to a LISTSERV and the fact of that subscription is generally known throughout the Internet, unless the subscriber takes a positive action to conceal it.

A second issue has to do with the ownership of material that participants post to these discussion lists or newsgroups. Currently this is a highly contentious issue, and positions range from organizations that sponsor discussion lists (such as the Well service in the San Francisco Bay area, which simply states that posters own their own words) through individuals who argue that they automatically own their own words and affix copyright notices to postings stating this option. When one considers the text of a LISTSERV discussion in the aggregate, it is even less clear who owns rights to complications copyrights. While a rather complex consensual etiquette is developing which suggests that one should not repost from one list to another or reuse a list posting without the author's permission, the legal (as opposed to consensual and moral) basis of these conventions remains extremely unclear. Many LISTSERV are beginning to view this in some sense as a contractual matter; upon subscription they present subscribers with the assumptions about reuse of postings on the list.

## 9. The Internet Tradition of "Free" Information: Quality, Integrity and Liability Issues

Fee based services are a relatively recent development on the Internet. Prior to the last few year, for both policy and historical reasons, such services did not exist on the net; certainly, there were machines, services and resources that were restricted to specific user communities (for example, super computer centers where time was allocated through a grant-like mechanism, or machines that belonged to specific universities and were used by communities at that university), but this was considered to be a very different situation from a vendor that provided service to anyone on the net who was willing to pay for such service. The recent presence of commercial information providers such as Dialog and BRS indicates that these policies are a thing of the past, and that current policy at the very least welcomes vendors supplying services to the research and education community. However, there is a strong philosophical bias towards the use of "free" information on the network among most of the network user community. This is a particularly comfortable fit with the values of the libraries that have been appearing on the network both as information providers and information organizers: the Internet tradition of free information is consistent with the library ethos of both intellectual freedom and free access to information. And there is a great deal of free information available; in fact, much of the development in software tools (which were themselves typically free, public domain software, at least in their initial stages) to facilitate the mounting of networked information resources (for example, WAIS and Gopher) has been to facilitate the ability of organizations on the network to offer access to an ever growing array of publicly-accessible networked information resources. This bias towards free information is evidenced by the rather minimal billing and access

control facilities in these software systems,[59] and indeed throughout the Internet generally.

While there are any number of organizations which have the dissemination of information to the public as part of their mission (including most types of government), it is important to recognize that the strong bias in the Internet user community to prefer free information sources provided by these organizations is not without problems. These problems include a tendency by network users to use relatively low quality information (a "you get what you pay for" argument), a lack of accountability for the quality and integrity of information offered without charge to the Internet community, and the potential for various forms of bias to find their way into the most commonly used information resources on the network. The ultimate result a few years hence—and *it may not be* a bad *or inappropriate response, given the reality of the situation-may* be a perception of the Internet and much of the information accessible through it as the "net of a million lies", following science fiction author Vernor Vinge's vision[60] of an interstellar information network characterized by the continual release of information (which may or may not be true, and where the reader often has no means of telling whether the information is accurate) by a variety of organizations for obscure and sometimes evil reasons.

The first issue with "free" information is that it is, of course, not really free, but rather subsidized. Free information might be subsidized by a government agency as part of that agency's mission. It might be subsidized by a not-for-profit organization as part of that organization's mission to communicate its viewpoint to the public. A university might make information available as part of its missions to support research, education and public service. Some public information resources might be subsidized by a for-profit corporation as part of a public relations campaign.[61] It might, as discussed elsewhere in this paper, be provided as a means of acquiring market research data or mailing lists of people with specific interests. Following traditions in both the print and broadcast media, it might be subsidized by advertisers as a means of delivering advertising.[62] As an extreme case, one can envision the Internet analog of television

---

59 There is some evidence of a new focus on fee-for-service information resources on the network. The University of Minnesota, which funds Gopher development, has recently begun the implementation of licensing agreements for the Gopher software that assess substantial charges for organizations that wish to provide information—particularly for-fee information—using the Gopher software, in conjunction with an upgrade of the software to Gopher+, which includes facilities to address billing and user authentication. While Thinking Machines Corporation placed the initial version of WAIS in the public domain, Brewster Kahle, one of the original developers of the system, has recently formed a company called WAIS Incorporated which is seeking to commercialize the system (or at least the server software) and to work with information providers who wish to offer their information through WAIS servers-often for a fee.

60 This is described in his 1992 novel *A Fire Upon the Deep* [Vinge, 1992]

61 Print publications have tried to establish conventions that clearly identify advertising material as advertising; for example, when a corporation purchases space on a newspaper's editorial page for a corporate statement on a public-interest issue, the print publication typically goes to some lengths to try to indicate that the material is not part of the publications editorial content but rather paid "advertising" (communications). There will clearly be a need to develop similar conventions for Internet information resources.

62 Advertiser-supported material might be viewed with particular caution. A number of authors have explored the effect of advertising subsidies on the popular media (both print and broadcast) and have suggested that advertisers have a significant effect on content and editorial positions taken by these

"infomercials" where one obtains some information (probably of questionable accuracy and/or value) along with a very long sales pitch on some given topic, such as getting rich through selling electronic real estate to house out of copyright books. In a sense, one can regard much of the current crop of "shareware" and demo versions of commercial software as forms of advertising promotions.

It is also difficult to entirely separate "free" content from the mechanisms that provide access to the content. One of the properties of networked information distribution is the ability to suddenly and simultaneously make new information available to an enormous user community; a community that is perhaps far larger than the ability of the computing system supporting the information resource to service at initial peak load. New documents, new virus definitions for a virus protection program, new software releases or bug fixes may be provided free by the information provider, but the public access resources supporting access to this material may saturate under the demand levels of initial public release.[63] In these situations, users who have a real need for timely access to the newly available information may pay a premium to some access provider (perhaps a service like CompuServe or Applelink) rather than retrying and continually being refused access to some public FTP archive. Or they may be willing to accept some advertising, or the collection of their address for future marketing purposes as a condition of obtaining timely access to the information.

Another very real issue is lack of responsibility and accountability in making information available on the networks. Tools like WAIS and Gopher have made it very easy for anyone to offer an information resource to the network community at large; one simply implements a WAIS server or a Gopher server on one's personal workstation, for example, using publicly available software. Whether this information is accurate, and whether the institution or individual that initially made it available feels any responsibility to ensure that it is accurate or current is unclear. A recent problem that caused a considerable amount of discussion the LISTSERV PACS-L is indicative of the problem. Someone on the network went looking for a copy of the periodic table of the elements. Much to their delight, they located one that someone had made available through one of the networked information access tools. Unfortunately, upon closer inspection, this periodic table was missing entries for a number of the elements. Unfortunately, it was not clear that anyone felt much responsibility to remove the incorrect information from the network, or to update it to be accurate. While the readership of the list PACS-L now

---

media. See for example *The Media Monopoly* [Bagdikian, 1992]. One of the most pernicious aspects of this advertiser influence is that it is hard for most viewers to identify and subtle in nature. In a networked information environment where advertising may be more easily ignored by viewers of information resources, it may be even less clear who the advertisers are.

63 The accepted community practices for access to this type of information distribution are complex and quite interesting. For example, some public FTP archives limit the number of concurrent accesses from a given institutional network, with the idea being that if information on the archive is very heavily used by a given user community, that community should import the information and then make it available internally to reduce load on the public FTP archive. Unfortunately, there is little automated assistance to facilitate such actions; in an ideal world, an institutional network might recognize that a file was being frequently requested from a (globally accessible) public FTP archive and automatically import (cache) the relevant file and then redirect requests for copies of the file to the institutional file server; later, when demand died down, the local copy would be discarded, and requests would go to the globally available archive. But this type of automated implementation of institutional responsibility for sharing in the resource commitment to distribute such files does not exist today.

probably realizes that they should not trust the information in this public-access periodic table, it seems probable that any number of new citizens of the network will rediscover this inaccurate resource in the coming months and years; hopefully, they will quickly realize that it is unreliable, but there are no guarantees. This problem is closely related to a number of currently unclear issues having to do with liability for incorrect software or databases.[64] While liability for failures to provide correctly functioning software and accurate electronic databases are likely to provide fertile ground for litigation in future, many of the issues in this area are focused around some sort of implied quality or appropriateness of a product in the context of a sale or licensing agreement: it is particularly unclear how these issues apply to information resources that are provided on the "use at your own risk" basis that characterizes many (particularly not institutionally supported) "free" Internet information resources today. [65]

## 10. Access to Electronic Artifacts: The Problems of Versions, Presentation and Standards

### *The Nature of Electronic Works*

When we consider the contents of digital libraries, we encounter a bewildering menagerie of different artifacts: interactive services, computer programs, text files, images, digital movies, networked hypertext information bases, real time sensor feeds, virtual realities. The taxonomy for these artifacts has yet to be firmly established; indeed, one of the greatest sources of difficulties in developing a consensus vision of the electronic libraries we hope to make available through the NREN. There is some basic conception of a "work", however—not necessarily in the legal sense of intellectual property law, but rather an electronic "object" or "document" (in the broad sense), as distinct from a service. Such works can be viewed, copied or referenced; they may be electronic texts, or multimedia compound documents. The are the analogs and generalizations of books, journal articles, sound recordings, photographs, movies or paintings.

The printed page or image can be apprehended with no more technology than the human eye; other works are recorded on storage artifacts such as videotape or phonograph records which require technology to convert them to a form that can be apprehended by the human senses. Yet these storage formats and technologies are fairly well standardized (at least in comparison to the digital world), and, at least to some level of approximation we believe that the content of such a work can be moved from one storage medium to another (say, from 35 mm film to videotape or from LP recording to CD audio disk), perhaps with some degradation of quality but without

---

[64] There is reasonably well established liability for provision of incorrect information in commercial settings, such as credit bureaus; the situation for individuals or even institutions that make information available for free (with no promises of accuracy or maintenance) is much less clear. See, for example, [**Samuelson** 1993] for a further discussion of liability issues.

[65] Resources that are institutionally supported are likely to be of better quality than those that are support by the actions of an individual. But it is also important to recognize that there are a number of resources that are offered by third parties, particularly information access providers such as Dialog, that also have major problems with completeness and accuracy.

loosing the "essence" of this content. This relatively high level of standardization is facilitated by the fact that methods of presentation are well established and fairly simple-one listens to sound and views movies. The experience is not interactive. But, when we consider the new electronic "works," it is clear that they can only be apprehended by the human senses and the human brain interacting with a computer system that includes software and various input and output devices. The experience of these works is complex and interactive; a work can be viewed or experienced in many different ways. Further, other intuitive measures of a work are lost; for example, browsing a printed work gives the browser a sense of the amount of information that the work contains. It is unclear how to measure the amount of information that is contained in a multimedia database.

There is enormous variation among the capabilities of these computer mediators, and too few (or too many) standards. Material may be converted and transferred or displayed in many forms; these transformations can cause major changes in both the presentation and the intellectual content and thus threaten the integrity of the work. Worse, in many cases the content of the work is inextricably bound with the access methods used to view (and navigate within) the work. It is impossible to separate content from presentation of viewing. This inability to isolate intellectual content calls into question the long term and broad based accessibility of works.

### *Presentation and Content Standards*

Standards are needed that permit intellectual content to be encoded into a work in a way that is independent of any specific access software; this is needed to ensure that material can be viewed from any appropriately configured platform, and ensures that the work can outlive specific hardware or software technologies. This is both a preservation issue-ensuring that the work will be available for the foreseeable future-and a portability issue, ensuring that the work will be available from multiple hardware platforms today or in the near future. Today, few useful standards exist, and, for various reasons, those that do are not widely adopted as vehicles for distributing electronic information. Worse: it is becoming increasingly clear that there is a serious lack of consensus as to where intellectual content stops and presentation begins,[66] and the extent to which information providers such as publishers are prepared to distribute content as opposed to specific presentations of content. This controversy has had a considerable impact on the development and acceptance of relevant standards, and most standards currently seeing broad use tend to encode presentation along with content rather than separating the two.

[66] This ambiguity between form and content has very significant implications fOr copyright. A number of information providers seem to be taking the position that by simply reformatting or marking up out of copyright works they can return the specific representation of the work to copyright control. This is not necessarily a bad thing, as the availability of out of copyright works in marked up forms such as the encoding specified by the Text Encoding Initiative program adds substantial value, and the copyright protection protects the investment of the companies doing the markup, but it is a source of considerable uncertainty and confusion.

A great deal of electronic information today comes packaged with access mechanisms. In CD-ROM publishing, access software is almost always part of the disk.[67] In other environments one finds information packaged in formats such as Hypercard stacks for the Apple Macintosh.[68] Increasingly, we are seeing the use of programs that provide access to books and journals—Superbook from Bellcore, the RightPages system from ATT Bell Labs and similar systems. Such integration of retrieval apparatus with content raises serious concerns about the ability of libraries to provide access to this information in the long term and across multiple platforms. At the same time, other information providers, such as the Government Publishing Office (GPO) and the Bureau of the Census are creating other problems by issuing data, such as the Tiger files, without any access software, which raises serious questions about the ability of libraries to meet the objectives that are mandated by program such as the Federal Depository Library legislation.

Purchasers of electronic information such as universities and libraries are particularly eager to obtain information in presentation-independent formats such as SGML that allow the information to be reformatted for a wide range of display devices, and even easily processed by computer programs that perform various manipulations or indexing on the information. By contrast, many publishers have expressed a great deal of concern about the loss of control over the presentation of their works; they are worried about both quality and integrity questions. Thus, publishers are in many cases much more comfortable distributing information in formats like bitmapped images or PostScript files that allow reproduction of the information only in a specific presentation (that is, the information can only be replicated either in print or in a screen display with exactly the layout, fonts, and similar properties that were defined by the publisher when the information was created). Another factor in representation choice is that formats such as bitmaps, while they preserve the integrity of the print page, require a much more sophisticated support infrastructure of storage servers, networks and display devices than simple, unformatted text, and thus bar a large number of current network users from access to the material. SGML, while less demanding of storage capacity and network bandwidth, requires sophisticated viewing systems which today serve as a barrier to very broad access.

Even the transfer of files in presentation-specific formats is problematic. Standards for bit-mapped images are still immature; while the Internet Engineering Task Force recently came up with a basic standard for monochrome bitmapped images of pages, [Katz & Cohen, 1992] there is still no agreement on the broader structure that should

---

67 This explains, for example, the great difficulties that libraries and other organizations have encountered in networking CD-ROM databases. Most CD-ROM software is designed to run in the PC environment which does not currently integrate easily or well with commonly used wide area networking technology (i.e. TCP/IP and related protocols). Yet the content of the CD-ROM is so closely integrated with the access software that it is not feasible for purchasers to write their own access software that is more compatible with the network environments in use within the purchasing institutions, and, in fact, the information publishers regard data about how their information is formatted on the disk that would permit the writing of alternative access and presentation software to be a trade secret in many cases.

68 I am not aware of any software in common use that permits browsing of Hypercard stacks on other hardware platforms such as the PC, although there are of course programs that provide the ability to construct and browse similar databases on the PC.

be used for sets of pages[69]. Reproducing such bitmapped image files exactly is still tricky due to variations in the resolution of display and printing devices, which may require interpolation or decimation of the image. The difficulties of printing PostScript files created on one system and then transferred to another system are well known, and include not only header file incompatibilities and PostScript interpreter problems but also problems with fonts.[70]

SGML is frequently suggested as a good prospect for a content standard that can also incorporate some presentation information. However, SGML is really a standard which defines a language for defining document schemas (called Document Type Definitions, or DTDs) and for marking up documents according to a given DTD. While there are numerous attempts to use SGML as a basis for developing industry or application specific document markup standards (for example, the publishing industry has developed ANSI/NISO Z39.59, sometimes called the "electronic manuscript format" which is aimed at the transfer of manuscripts from author to publisher and for use by publishers during the editorial process; the Air Transport Association has developed DTDs for applications such as aircraft maintenance manuals; and the Text Encoding Initiative is developing standards for deep markup to support textual analysis processes of various scholarly disciplines in the humanities and social sciences), it is unclear to what extent these will be accepted.

Currently, most authoring systems do not support SGML (though there is some modest evidence that this is slowly improving). Most documents are either authored using word processors—Microsoft Word, WordPerfect and many others-or using various markup systems such as Troff and Tex. There are converters that move from one of these formats to another, but usually with some considerable loss of information for complex documents. Further, languages like Tex and Troff have many variants and enhancements.

Effectively, if one looks at how documents are being distributed on the network today, there is typically a canonical version of the document that provides as much content markup as possible-typically right out of the authoring system that created it. Users of the same authoring system can use this version. Then there are typically two derivative versions-one in pure ASCII text suitable for viewing on lowest-common-denominator terminals—and one in a PostScript or other print-ready format that (with a good deal of

---

69 The IETF standard allows multiple page images to be transferred as a group, but access to a specific page within the group is awkward. Some publishers are distributing collections of pages as sets of files, with one page per file, using the IETF format. Other groups are looking at higher-level structures that can be used to transfer groups of pages.

70 Fonts represent a particularly subtle problem. While a given font is not protected, certain representations of font families can be copyrighted, as I understand it. Today, in order to facilitate reproducibility, many PostScript files are shipped with their font definitions (which make the files huge), violating the copyrights of the fonts' rightsholders. There are a series of technologies being deployed at present, such as Adobe's SuperATM and Acrobat systems which permit documents to be shipped without fonts. The receiving workstation can use some parameter information to "regenerate" or interpolate a font that is supposed to be similar to the font actually used in the document, thus avoiding not only the copyright problem but also the very real practical problem of receiving a document that employs fonts one does not have on one's local machine. However, this very convenient substitution is performed at the expense of degrading the integrity of the document, and, again, most users may not really even be aware of what is happening, other than that the document looks a little strange.

luck) can be used to produce a printed version of the document. Most of the presentation, and often a great deal of the content, is lost in the ASCII document; the print-image document may or may not be usable, but if it is it preserves presentation integrity at the cost of making most of the intellectual content inaccessible.

While image information does not face the same dichotomy between content and representation a number of the same themes reappear. There is an enormous proliferation of formats for various types of digital images—TIFF (various dialects), GIF, PICT, etc.—which can be intermixed with various compression schemes (CCITT Group 3 or 4, JPEG, etc.). The formats themselves are in some sense more of a nuisance than anything else, and there is software available that converts from one format to another without much, if any loss of information. The compression standards are another matter entirely; JPEG, for example, includes "lossy" compression modes in which the size of the compressed image can be traded off against accuracy of reproduction, with more compact images offering a less detailed reproduction of the original image. Accuracy of reproduction of images—particularly color images-depends of course on having a sufficiently high quality display and/or printing device. But there are also more subtle issues: different monitors, for example, display the color palette differently, and a painting will look quite different when displayed (at identical resolution) from one monitor to another. This is a substantial concern for many color imaging applications, particularly in the fine arts.

There are on the order of twenty different standards for the storage of digital sound at varying degrees of quality and compression; size of the stored sound object depends on sampling rate, dynamic range and the compression scheme used (and whether that scheme is lossy or lossless), among other factors. Many of these schemes are platform specific.[71] More general multimedia standards—for example, for video or other compound objects—are still in their infancy, with a number of platform-specific standards such as Quicktime for the Apple line and Video for Windows coming into use.[72]

As one views these standards in the context of digital library collections one can see a conflict between integrity and access emerging. While it seems likely that over the next few years the industry will continue to adopt more platform-independent standards and that at least some publishers will move towards standards for electronic text that provide more content information and are less presentation specific, it is improbable that at any time in the near future we will find that the user community has installed a base of access devices with homogeneous, high-quality capabilities for reproducing

71 Part of the problem here is that hardware for sound reproduction is not universally available on various platforms (it is now standard within some vendors' product lines, but not across vendors) and the hardware used to play back sound is also not well standardized yet.

72 Quicktime is now available on Windows platforms, at least for viewing movies, and it seems likely that over time most of what come to be the better established digital movie standards will be ported across multiple platforms. This will become easier as the processing to support material that is formatted according to these standards is more readily done entirely in software with acceptable performance and reproduction quality. Right now, these formats push the capabilities of the basic hardware on most common platforms, and often benefit greatly from specialized hardware assistance. It also remains to be seen whether translators from one format to another become readily available for movies as they have for the popular image formats.

sound, displaying color images or viewing digital movies. It is also important to recognize that the capabilities of a user workstation are not the only issue in full quality reproduction of electronic information; network bandwidth is also a factor. The larger the object the longer it takes to transfer it. A high quality digital image or sound recording will take longer to move to the user's workstation for viewing than a lower quality one. In some cases this may translate into a cost issue; if the user is paying connect time for some service, or paying for network bandwidth on a usage-sensitive basis, he or she may be unwilling to pay for a full quality image, or may wish to browse lower-quality **simulacra** before selecting objects to move in full resolution. For some services, bandwidth limitations may be absolute in the sense that using a full-quality service will make it too slow to be usable. Consider users connecting from home; in most cases they are limited to rather slow connections (9600 bits/second in most cases today, even with relatively expensive and modern modems; perhaps in the not too distant future ISDN at 64 or 128 KBits/second will become a generally available and affordable reality); at these speeds, many users find it more satisfactory to communicate with systems by emulating older character-based terminals rather than running available graphical interfaces based on the X Window system technology. The issue is not that they cannot support the X Window system on their workstations; rather, it is that they don't have the bandwidth on their network connections to run it effectively. It seems likely that libraries, as institutions, will obtain high-bandwidth links to the Internet much more quickly than most end users. and that the issue of available bandwidth to the end user will be a critical factor in determining when information can be delivered directly to the end user in electronic form; it is likely that there will be a long transitional period during which libraries will have to act as staging sites for information on its way to the end user, or to some printing facility in cases where the end user does not want to move the information all the way to his or her workstation prior to printing it.

Thus, information servers such as libraries will face the need to decide, as they design their systems, how much downward conversion they are willing to do in order to permit viewing of material in a degraded format, and how they will make the user of that material aware that he or she is indeed receiving a degraded presentation. Some degraded versions of content will obviate the need for specialized software viewers, thus adding value. Also, while the libraries or other information providers may make the decisions about the spectrum of capabilities available in access systems, rights holders—publishers and authors—will clearly, through license agreements, have a voice in the extent to which libraries will be allowed to apply these capabilities to deliver degraded-quality representations of works to users who cannot view the "canonical" full quality version of the work.

### *The Problems of Structured Information*

There is a natural tendency to view the contents of digital libraries as primarily information intended for direct human apprehension—text, sound, movies and the like. In fact, it is clear that a major class of electronic information resources will consist of structured databases—not databases of abstracting and indexing records or of fulltext, but of genetic sequences, weather observations, commodities prices, chemical compounds and their properties, menus, plane schedules, biographical summaries, and thousands of other groups of information. This information may be viewed by humans

(with the aid of a viewing program) but also be used extensively by other programs that manipulate information, summarize it, reason with it, and correlate it with information from other sources on behalf of human users. In order to be able to access and operate on such information reliably it needs to be rigorously structured; the human eye and the human mind are far more capable of compensating for variations, applying common sense, and inferring than most software. There are a growing array of tools available for defining data structures that can be moved across the network, such as Abstract Syntax Notation One (ASN.1); however, there will be a growing need to standardize semantics for data elements that may appear in records retrieved from information servers on the network. Indeed, the success of such standardization efforts will be central to the success of large scale, distributed scientific programs in areas as diverse as systemic biology and botany, molecular biology, global climate change, and the initiative to map the brain. While some communities, such as the library community and more recently the various groups interested in electronic data interchange (EDI) have substantial experience in this area and have made some progress within limited domains, this is a new concept for many scientific communities and one that these communities are having considerable difficulty making rapid progress in realizing. Further, while many of these scientific and research communities clearly recognize the need to develop such standards for data elements and a clear focus exists within the research community to address the issue, it is far less clear who will do the work to develop and maintain similar standards for the more mundane or commercially-oriented standards .73

Sadly, until such time as these structured information interchange standards become defined and widely implemented, we will be unable to realize one of the major potential payoffs of creating libraries of electronic information—the ability to view these as knowledge bases that can provide a foundation for the development of more capable software agents that can act on our behalves, but that can function relatively independently, without the need for a human being's eyes, mind and judgment to preprocess large amounts of information.

## 11. Digital Images and the Integrity of the Visual Record

In a very important recent book titled The *Reconfigured Eye: Visual Truth in the Post-Photographic Era* Mitchell, 1992], William Mitchell traces the historical role of the visual image as a record of events. Prior to the development of photographic technology, painting served the cultural role of recording things, events and people. In the 19th century photographic methods were developed which were far more accurate than painting in reproducing and recording; indeed, the author argues that the development of these technologies freed painting to explore new and less literal reproductions of visual reality. While "trick photography" and various photographic effects were developed very early in the evolution of photography, there seemed to be a fairly well understood implicit consensus that these uses of photography to capture unreal images

73 Indeed, for the commercial standards, in the worst case we are likely to see a number of different commercial information suppliers attempting to establish incompatible de facto standards in order to capture markets and ward off competition. The notion that open standards actually establish and foster the growth of markets in information is not yet generally accepted by most of the commercial information suppliers.

would be applied primarily in artistic frameworks where their use was clearly identified to the viewer of the photograph. In the 20th century, as photographs (and later moving images) became an increasingly central part of advertising, alterations to photographs without warning to the viewer became more commonplace.[74] In the past few years, however, photography has begun to migrate to digital technology; this shift, in conjunction with the ease with which digital images can be altered, combined, or even generated from scratch by computer graphics programs has created a situation where the implicit evidentiary value of images (without regard to their provenance) has been lost.

It is not entirely clear that this loss of evidentiary value has fully penetrated the public perception, despite recent developments such as advertising which interposes modern-day celebrities into historical films in commercials, very sophisticated special effects in movies, and even virtual reality technologies to which the public is now regularly exposed. But the fact remains that it is almost impossible to tell whether an image represents a record of a real event given current technology; moreover, while twenty years ago it required a great deal of sophistication to alter images or films, easy to use software that can be employed to perform such image manipulation (and, increasingly, image creation) is now widely available on inexpensive personal computers.

This change in the meaning of images has a number of implications for digital libraries. When viewing an image, one must always harbor a certain degree of skepticism about whether the image actually represents a real event or thing and to what extent this representation may have been altered. In essence, when using images, one must constantly be concerned with the source of the image, whether the image has been altered subsequent to its capture, and the purposes for which the creator of the image is making it available. Without facilities to track and verify the source of images, they have become meaningless as a part of the historical record. 75

## *12.* Authenticating versions and sources

In the print environment a great deal is taken for granted about the integrity of documents. If an article appears in a journal it is extremely rare that the authorship that the journal lists for the article is called into question; when this happens it is typically framed either in the context of scientific and scholarly misconduct such as plagiarism and/or results in a lawsuit. Outside of scholarly circles the issue is probably more likely to revolve around the publisher's right to publish (that is, whether the rights holder indeed gave permission, or whether the claims to rights on the material are valid—for example, in the case of unpublished archival material) than whether the author

[74] While advertising was perhaps the greatest culprit in the undermining Of the Integrity of visual images as a record of events, there were many other contexts in which altered photographs were used: politically motivated changes and sensationalistic news reporting are two other common areas.

[75] Of course, just because an image cannot be verified as a representation of an actual thing or event does not mean that the image is not of interest. Computer generated or computer altered images are of vital importance as works of art, as hypothetical models of things (for example, a reconstruction of a dinosaur) and as records of culture. The problem is that the viewer of the image needs to know the context within which to understand the image.

attribution is false. With the exception perhaps of certain tabloids one would not normally assume that there was much reason to question authorship, Similarly, print publication naturally tends to produce specific editions of a work; if two people both have the same book or issue of a journal there would be little reason to question whether the two copies of the publication were indeed identical in content.[76] It is not that a publisher couldn't deliberately publish variant copies of what is labeled as the same edition, or deliberately misattribute authorship of material, but rather that it does not happen often and when it does the publisher is typically readily identifiable and can be sued by the aggrieved parties. Further, there is little motivation (other than general malice) to motivate most publishers to do this; a publisher would have to go to considerable trouble, expense and risk in order to do it.

Perceptions and concerns in the world of networked information are quite different. It is very easy for someone to distribute information over someone else's name, and hard to trace the person who does it in most cases. It is very easy to replace an electronic dataset with an updated copy, and, since there is no automatic system which distributes multiple copies of the original version to different sites (such as the distribution of an issue of a printed journal or a book) the replacement can have wide-reaching effects. The processes of authorship, which often involve a series of drafts that are circulated to various people, produce different versions which in an electronic environment can easily go into broad circulation; if each draft is not carefully labeled and dated it is difficult to tell which draft one is looking at, or whether one has the "final" version of a work. Because of the ease with which material can be taken from one document and inserted into another which can then be circulated to a large number of people quickly, there are concerns about quotation from obsolete or draft ("unpublished") versions of a work. Visionary authors such as the late Ithiel De Sola pool[77] have written that the world of networked information would lead to the demise of the "canonical" form of documents that print publication created, and that documents would become living texts that were continually adapted and annotated. Events thus far have suggested that De Sola Pool may have overstated the case. While it is common within small groups to have people annotating drafts of a document, they are typically ultimately brought to a final, "canonical" form. Further (and it is unclear whether this is due to current limitations in information technology such as groupware and collaborative authoring systems or whether it is a more basic problem having to do with the limits to the size of a group that can collaborate effectively) such continual annotation typically occurs among a relatively small community of collaborators or reviewers, and not among the full community of interested individuals on the net. In cases where a large scale, long term collaborative effort is taking place to develop and manage a "living document" such as On Mendelian Inhetitance *In Man* at Johns Hopkins, a fairly complex and formal editorial structure is set up to manage this collaboration, and considerable care is taken to validate and track updates and their

76 While there would be little question in most people's minds about what they could expect from printed literature, this is not necessarily the case with other media, such as videotapes, audio recordings, or computer programs, although even here it would probably not occur to most people to doubt that identically labeled copies of a work were in fact identical.

77 See his work on "The Culture of Electronic Print", reprinted/adapted in his book *Technologies* of freedom [de Sola Pool, 1983].

sources. In a real sense, these efforts are as much undertakings to create and manage databases as they are efforts to author documents.

Interestingly, aside from a few pranks and malicious acts (most commonly people sending electronic mail with false origin addresses) it is unclear whether the fears that people seem to harbor about the deceptive and mutable nature of the electronic environment are justified by real occurrences of problems. Also, those publishers who have risked the Internet environment have had less problems with piracy than one expect given the experience of software vendors, for example. The simple fact is that several people are successfully distributing publications on the Internet today for profit (though I don't know how much real profit they are making, they are still in business after several years in some cases). Nonetheless, it seems clear that if network based information distribution is to become a widely accepted context for the sorts of archival materials that libraries currently acquire and provide access to in print these concerns must be addressed. Certainly, the development and wide implementation of technologies and practices to address these concerns will, at the least, lead to a far more robust and reliable environment, although from a strict cost-benefit analysis it may be difficult to fully justify the costs of addressing some of the fears one hears voiced.

To clarify the focus here, it is important to recognize that while the network will undoubtedly be used extensively for transacting commerce (including, as just one example, commerce in the form of acquisition of electronic information by individuals and organizations, which may involve activities such as the identification of the parties involved, the exchange of credit information for billing purposes, the assessment of charges against some type of account, and even the acceptance of the terms of a license agreement for copyrighted material) and there is, I believe, strong justification for ensuring that these commercial transactions are conducted in a technical environment that protects the security and confidentiality of all parties, the issues involved in protecting transactions are somewhat different from those involved in ensuring that a user of the network who finds a document somewhere can verify that the authorship attribution is true and that the copy which the user is looking has the same content as the version that the author "published." The issues of protecting network commerce generally are outside the scope of this paper, other than simply to observe that for a market in network-based digital information to develop it will be necessary to develop and implement adequate measures to protect commerce on the network and also to conduct some form of electronic rights clearance. The remainder of this section will address issues of verifying authorship and integrity of contents, and the state of the art in technologies to accomplish these objectives.

Public key crptography and various higher level protocols layered above the basic cryptographic algorithms offer methods that can be used to effectively address both of these needs. A public key cryptosystem can be used to attach a signature to a digital object in such a way that the contents can be associated with a given individual or organization at a given time. There are well-established algorithms for computing "digests" of digital objects in such a way that it is extremely unlikely that any change to the object can be made without changing the value of the digest computed from it. Thus, by checking whether the digest for an object in a user's possession is the same as the digest value that the author has signed and makes available as the signature of the current version of the work, it is straightforward to check whether one has the same object as the author or publisher distributed. These systems offer the additional feature

of non-repudiation; it is possible to include capabilities so that one can prove that a given author actually distributed a document at a given time even if that author later denies it. Such capabilities can be seen today in the Internet in the privacy-enhanced mail system [Balenson, 1993; Kaliski, 1993; Kent, 1993; Linn, 1993].

While the basic technology exists to solve the problems in question (at least as long as one is satisfied with literal bit-by-bit equivalence of two digital objects as a definition of having the "same" document, which is often really overly restrictive, since it prevents any reformatting, character code conversion or other activities that might be needed to successfully move the document from one machine to another, even if these do not change the "content" of the document in any way) the operational problems of implementing these technologies on a large scale in environments such as the Internet are far from solved. There are least four barriers:

•Standards are needed. While the algorithms are well understood at a general level, in order to ensure interoperability among implementations agreements need to be reached at a much more specific level of detail and documented in standards. Parameters such as the precise types of signatures need to be defined, along with lengths of public/private key pairs, the exact computational algorithms to be used, and the supporting protocols and data interchange formats. The IETF specifications for message digest algorithms [Kaliski, 1992; Rivest, 1992a; Rivest, 1992b] are an important step in this direction, but more work is needed. It is also important to recognize that there is more to an effective system for authentication and verification than simply algorithms; there are application protocols which use these algorithms to be defined, along with an accompanying infrastructure of service providers (see below). An additional problem that must be resolved in the standards area is the seemingly continual conflict between the standards proposed or established by the Federal Government through the National Institute for Standards and Technology (NIST) and the standards that are favored by the commercial and research and education communities .78

•Patent issues need to be addressed. What is currently widely accepted as the best public key cryptosystem is called RSA (named after its inventors, Rivest, Shamir and Adelman); this was patented and commercial rights to this patent are licensed, as I understand it, to a company called RSA Data Systems incorporated. Similarly, Public Key Partners holds patents to a variety of public key and message digest algorithms; to make matters even more confusing the National Institute of Standards and Technology (NIST) has filed patent applications for some of the algorithms that it has developed and adopted as federal standards and proposed licensing these (on an exclusive basis) to Public Key Partners; again, there seems to be a provision for free use of the patents for at least some types of personal or non commercial use. The net effect of these patents on basic cryptographic technology is to promote considerable uncertainty about the status of the algorithms and to inhibit their incorporation in software of all types (but particularly public domain software; while large corporations can negotiate and pay for

78 Recent examples of this problem include NIST's controversial adoption of the Digital Signature Standard Algorithm and the Secure Hash Standard (FIPS 180), as well as the widely-denounced proposal for the Clipper chip and its accompanying key escrow system. The continued unwillingness of the government to recognize the RSA public key algorithm as a standard despite its widespread use is another example.

license agreements with the rightsholders, individual software developers or university based software development groups that wish to distribute their work without charge typically are unable or unwilling to do so), despite the relatively liberal positions that the commercial rightsholders seem to be taking on personal use and use by the research and education communities. Further, the patent filings by NIST are regarded with considerable suspicion in some quarters; there is concern that in future these patents might be used as a means of controlling the use or implementation of the technology.

•Cryptographic technology is export restricted. This has caused two problems. The commercial information technology vendors have been somewhat reluctant to develop products which can only be marketed in the United States without major export complexities, particularly given that authentication and digital signature technology tend to be very basic building blocks for distributed systems. In addition, because the world of the Internet and of networked information is very clearly viewed as a global rather than a national enterprise, system developers and standards makers have been reluctant to use technologies that cannot be freely used internationally. The issue of the justification and implications of applying export controls to cryptographic technology is a very complex one that is well outside the scope of this paper; however, the impact of the current restrictions must be recognized. In addition, it should be observed that while the position of the United States on the export of cryptographic technology is crucial because of the nation's leadership in so many areas of information technology, other nations may also have laws related to the import, export and use of cryptographic technology that also create barriers to the free use of authentication and digital signatures on a global, Internet-wide basis.[79]

•Critical mass and infrastructure. Like so many things in the networked environment, these technologies will not come into wide use unless they are available on a large part of the installed base. Authors want to communicate; publishers and libraries want to make information available. If this information is not readily used without specialized cryptographic software that is difficult and/or costly to obtain, or that cannot be used outside the United States, it is unlikely that authors, publishers or libraries will use them. While in the specific applications under discussion here of verifying authorship and integrity of objects it should not be necessary to have specialized cryptographic software support simply to view the material but rather only to conduct verification,[80] it is really more of a question of whether the investment is worthwhile because enough

[79] There is a major public policy debate currently taking place about the appropriate balance between the rights of individuals and the private sector generally to privacy on the one hand and the desires of law enforcement and intelligence agencies to be able to monitor communications on the other. While the details of this debate are outside the scope of this paper, the interested reader might wish to review the history of export restrictions on the RSA algorithms, the recent proposal by the Clinton administration for the Clipper chip, and the deployment of the PGP ("pretty good privacy") computer software both inside and outside the US.

[80] In cases where cryptographic technology is being used in conjunction with rights clearing some proposals do call for the distribution of encrypted documents that cannot be read without both software to decrypt and the appropriate key. Somewhat similar approaches are being used today where vendors will distribute a variety of locked digital information on a CD-ROM (such as programs or font libraries) and then issue decryption keys on a file-by-file basis as the customer purchases these keys; one advantage to this approach is that one can phone order the information by providing a credit card and getting a key, without waiting for physical delivery of media containing the information being purchased.

people will make use of the services. In addition, it should be recognized that there is a substantial infrastructure needed to make public key cryptosystems and the applications constructed using them work on a large scale, including key providers, certification authorities, directories of public keys, and "notary public" servers (third parties that can witness signatures and contracts, or that can record the fact that a given entity had a given document at a given time and date as a registry function). The precise details of this infrastructure will vary depending on the standards, protocols and procedures that develop in support of an implementation of the technology; how these details change from one proposal to another are not important here, but recognizing that an investment in support infrastructure must be made is vital. Further, as indicated earlier under the discussion of standards, the conflicts between federally endorsed standards and standards favored by much of the user community are having the effect of fragmenting and confusing the user community, and greatly delaying the achievement of the necessary critical mass.

It is interesting to place the issues of authenticating authorship and document integrity in the broader context of the way in which migration to a networked information environment is beginning to suggest an "unbundling" of the services that publishers have traditionally provided in the print world. Print publishers serve as selectors of material; they prepare the material for distribution, distribute it, manage rights and sometimes royalty payments, and authenticate authorship and integrity, among other functions. In the network environment, it is clear that distribution (at least through a mechanism as crude as making a document available for anonymous FTP) can be done by anyone. It is clear that services that help people to identify material of interest such as abstracting and indexing services, reviewers, bibliographers, and ratings servces are likely to play an enlarged role in the electronic environment, and that these services can be quite separate from the persons or institutions that make material available. It may well be that authenticating and verifying the integrity of a document is at least optionally a separate, and separately priced (and perhaps costly!) service from simply obtaining a copy of the document.[81] If so, it will be interesting to see how much use is made of such services (outside of some specific environments, such as litigation, where cost is typically not much of a factor and such issues simply must be unambiguously established) and particularly the extent to which people are willing to pay to allay their fears about the electronic information environment.[82]

---

81 The integrity of published works is not an entirely new problem. A number of publishers currently provide loose-leaf services for areas such as tax law; determining whether a user has the most current version of such a service is today an important question with a potentially high payoff.

82 It is also necessary to consider other implications of establishing a chain of provenance for an electronic document. As discussed previously, technology for tracing the provenance of a document is well established, and depends on sophisticated cryptographic technology. It seems likely that US government, and perhaps other governments have agencies that are monitoring most international traffic, and that any encrypted traffic will attract their attention., since at least at present it is relatively rare In some nations use of cryptographic technology may be illegal, either across international boundaries or even within the national boundaries. Even if it's not illegal, it may attract attention. Are scholars prepared to attract the attention of such communications security agencies as a consequence of maintaining verifiable ties to the scholarly record?

## *13.* Citing, Identifying and Describing Networked Information Resources

As networked information resources are integrated into the body of information, both scholarly and popular, it will be necessary to extent traditional print-based methods of citation to accommodate reference to these new network-based resources. Here, the objective is to continue the functions served by citation in print literature: to permit a work to make reference to the contents of another work with sufficient specificity to permit the reader to obtain a copy of the cited work and locate the part of that work being referenced; to give the reader of the citation enough information to make some judgments about whether he or she is already familiar with the cited work, and to provide some information about the cited work such as date of publication, title and author which might help the reader to determine if it is worth obtaining a copy of this cited work. It is important to note that traditional print citations today serve both of these purposes; for example, citations consisting simply of document numbers assigned by some document registry are not typically used because while they would allow the reader of the citation to obtain a copy of the cited document, they don't tell the reader anything about the cited work to help in making a decision whether to obtain a copy of it. 83

At the same time that the need to cite electronic information resources is being recognized, several other closely related requirements are emerging. These include the desire of libraries, bibliographers and other organizations and individuals that organize information to catalog the increasingly valuable and common electronic information resources; essentially, to extend the existing mechanisms of bibliographic description and control to facilitate access to these resources. The needs here are closely related to those of citations, but more extensive in that there is usually a requirement to include more information about how to obtain access to a given resource once identified, and also requirements to include subject access or other classification information.

It is interesting to note that both for citation and cataloging purposes a number of people have expressed a desire to have the citation or cataloging record include some information (such as document digests or signatures, as discussed earlier in this paper) that would permit the user to check that he or she had retrieved the same version of the electronic object that the creator of the citation or descriptive record had originally described (at least as an option: when one is talking about electronic documents this makes sense, but when one is making reference to a database that is continuously updated at the level of an information resource, rather than referring to the contents of a specific record in that database at a specific point in time, such version information does not make sense). Logically, this requirement makes little sense. Reasoning by analogy with the print world, if a citation specifies the second edition of a specific work, it is possible that the publisher might change the contents of the work and reprint it without updating the bibliographic information or date of publication, in effect creating two editions of the work that have different content but are not identified as distinct

---

83 It is worth noting that in some areas of scholarship historically citation systems have been used that only address the identification of a work, or passages from it, without referring to specific editions. Examples include biblical scholarship and some types of literary criticism. Usually in these situations there is an implied canonical text, so it is not necessary to specify the specific edition of the intellectual content.

editions.[84] However, this does not happen often (at least for materials that are extensively cited and where very precise citation is important) in the print world and people don't generally worry about it much.[85] The emergence of this requirement for version verification in the electronic information world simply underscores the general perception that electronic information is more volatile and more easily changed, and that the contents of electronic objects cannot be trusted to retain their integrity over time without introducing special verification processes into the system of access and management of these resources. It is also worth recognizing that on a technical level this problem of version verification is largely unsolved as yet; while the digital signature and digest algorithms discussed earlier can readily ensure that a document is bit-for-bit the same as the one cited, citation typically is more concerned with intellectual content. As we move to an environment where software and protocols for retrieval of electronic documents (in the broad sense of multimedia objects) becomes more adaptive and mature, transfer of documents from one host to another may commonly invoke format translations and reformatting of various types automatically,[86] while such translations would presevee the intellectual content of the document (perhaps at varying levels of precision, depending on whether the transformations were lossless and invertable), the transformation would of course change the actual bits comprising the document and thus cause it to fail a version comparison test based on such bit-level algorithms.

---

84 T. be clear: current library cataloging rules direct catalogers to explicitly differentiate works that are different even if the publisher has not done so.

85 Indeed if anything, the problems today with citation to printed material, as discussed earlier in this report include the difficulty that the creator of the citation often does not realize that the publisher is producing multiple editions targeted for different geographical regions or for different subsets of the readership (for example, trade magazines that include special advertising sections targeted at readers who work in specific industries) and hence doesn't create a sufficiently specific citation. From the publisher's point of view, there is often great economic incentive to keep repackaging and reissuing content with minimal changes as new editions or even new works; the notion of going to the trouble of producing an unadvertised and unlabeled new edition and quietly introducing it into the marketplace is relatively rare, at least for print; this practice does occur sometimes with electronic publications such as software, where minor corrections or improvements are sometimes shipped automatically without much publicity, although even there the publisher usually changes the version number. There are a few examples of audio materials where different versions have been shipped with the same cover and same publisher catalog number.

Also, in the print world, in cases where a citation is to a work where there is some question about the precise final form of the work, conventions have been developed such as indicating "unpublished draft" or "in press" to alert the user of the citation that there may be some problems. Of course, such citations are the exception rather than the rule.

86 T. provide only a few examples of such translation, a document might be changed from one character set to another (ASCII to EBCDIC or UNICODE); fonts might be substituted, since fonts are copyrighted and the workstation receiving a document might not have the fonts used by the author, so it might be necessary to substitute similar fonts that are either in the public domain or that are licensed to the receiving workstation; an image or digital sound clip might be converted from one format to another, and the resolution or sampling rate might be altered; or more extensive format changes might occur, such as the rendering of a postscript document into a bitmapped image prior to transfer. The extent to which these transformations preserve the intellectual content of the work are highly dependent on the nature of the transformation and also the use to which the document will be put when it is transferred; for example, if it is only to be viewed, then a transformation from SGML markup to a bitmapped image makes no difference to the content in some sense, but if that same document is to be edited or analyzed by a postprocessing program, then there is a very large loss of information in the conversion from SGML to bitmapped representation.

A third set of requirements are more technical in nature but address some of the needs for both cataloging and citing networked information resources; while they solve neither problem they provide tools for developing solutions. In addition, a solution to these technical requirements is needed to enable the widespread development and deployment of a number of important networked information applications. These technical requirements are based on the need for standards so that one object on the network can contain a computer-interpretable "pointer" or link to another object on the network. This is needed for network-based hypertext systems such as the World Wide Web. It is needed so that document browsers can automatically follow references in a document when these references are to other network resources. It is needed so that bibliographic or abstracting and indexing records that describe electronic information resources can include information about where to find and how to access these resources. This last case is particularly important for a number of projects that are now underway where large bitmapped image databases of material are being created; because of the size of these databases it is desirable to store them at only at most a few sites on the network and to retrieve page images from them on demand; yet multiple databases of descriptive records, developed by multiple organizations, need to include links to these image databases. Further, in some cases, the descriptive records are being distributed under different license terms than the actual content; for example, some major publishers are exploring scenarios where they give away brief records analogous to tables of contents in printed journals, and then charge transactionally for retrieval of the actual articles.

The idea is that these pointers to networked information resources should be representable as an ASCII text string, permitting their inclusion in both electronic documents and in printed documents, as well as their easy transfer from machine to machine and from one application to another within a machine (for example, via cut-and-paste facilities now available in most graphical user interfaces, with the idea being that a user might view a document or an electronic mail message or a screen display from an online bibliographic database in one window, find a reference to a document that he or she desires to fetch, and simply highlight and drag the citation to another application, which would then fetch the object or open a connection to the service, using whatever access protocol is required).

These technical requirements are being addressed by a working group of the Internet Engineering Task Force. While the technical details of the IETF standards proposals are beyond the scope of this paper (and, indeed, some specifics of the standards are still under active debate within the IETF Working Group **as** of this writing) there seems to be some substantial consensus on the overall approach to be taken. It should also be recognized that there are some very substantial research problems in dealing with these technical requirements in full generality, and thus the IETF work should be regarded as a beginning and a framework that will undoubtedly undergo a great deal of extension and refinement in the coming years based on operational experience with the first generation standards, improved understanding of the theoretical issues and abstract modeling questions underlying the standard, and the continued development of protocols and applications for accessing networked information resources of various types.

Roughly, the IETF proposals call for the definition of a syntax for what they call a locator, which is an ASCII string that identifies an object or service that is hosted on a specific machine (typically specified by its domain name) on the network, the service (such as FTP, electronic mail, Z39.50 database query) that is used to obtain the object, and the parameters that are to be passed to that service to identify the specific object to be obtained (for example, in the case of FTP the fully-qualified filename). There are several problems with locators as a basis for citation, however. Machines on the network come and go over time, and files are migrated from one machine to another. Some commonly used files are duplicated on multiple machines; from the point of view of citation, one wants to refer to content and not instances of content, and thus should no more list machines containing copies of a file than one would list libraries holding copies of a book in a citation. An object may be accessible through multiple access methods (for example, FTP and database retrieval); indeed, the method of access may change over time and in response to improved technology, but the content being accessed remains unchanged. Further, one cannot tell whether two different locators actually refer to the identical content.

Thus, the IETF working group has proposed the definition of identifiers, which are strings assigned by identifying *authorities* to refer to content. An identifier for an object, then, is just a two-component object consisting of a specifier for the identifying authority (these would be assigned centrally, as a service to the Internet community, much like top-level domains or network numbers) and the identifier that the authority provided. These identifying authorities (and other organizations) may offer services that provide a mapping from an identifier to a series of locators, which could then be used to actually obtain access to a copy of the object. Some mapping services, particularly those operated by specific identifying authorities, might only resolve identifiers assigned by the operating identifying authority; others, perhaps operated by organizations such as libraries, might attempt to resolve identifiers issued by multiple identifying authorities into sets of locators. Locators would be viewed as relatively transient; at any time one could obtain a fresh set of locators corresponding to an identifier. Identifiers would be used in citations and other applications. It is important to note that the IETF model explicitly recognizes that deciding whether two instances of an object are "identical" is a subjective issue which is highly dependent on the objectives of a given identifying authority, and that there will be a multiplicity of such identifying authorities, which might include publishers, service organizations, libraries, or industry-wide standards implementations (such as the International Standard Book Number in the print world). The same content might be assigned identifiers by multiple identifying authorities; in some cases two objects might be viewed as identical by one identifying authority (meaning that the authority would return locators for both objects in response to its identifier) and yet viewed as distinct by another identifying authority. 87

Rules for citations are typically set by editors of journals, or sometimes by professional societies (for groups of journals) or by style manuals (such as the *Chicago Manual of Style).* While a number of journals (both print journals and electronic journals) have

87 As a specific case in point, one identifying authority might view a bitmapped image and a Postscript file of the same document as identical; another might view these as different objects. The issue of format variations and the extent to which these variations, as well as multiple versions of documents, should be recognized by and integrated into the locator and identifier scheme is still an active area of discussion.

already defined practices for citing electronic information resources it seems very likely that these practices will be altered over time to include identifiers which followed the IETF standard in order to facilitate both the identification and the retrieval of electronic objects. As these identifiers come into wide use, some of the other material that is currently specified in citations to electronic resources (such as the name of a machine holding a file available for anonymous FTP and the file name) might well be dropped. Some of the traditional citation data elements that help the reader to identify and evaluate the intellectual content of the cited work, such as author, title, and publication date, will almost certainly be retained. A few data elements used in some citation formats, such as the number of pages in a work, are problematic in an electronic environment; while it is clearly useful for the reader to have some sense of the size of a cited work, it is unclear how to most usefully measure this in an electronic environment that may contain multimedia works. The transformation of citation rules is likely to be a gradual process; it is important to note that, at least in practice, citation formats are really not national and international standards, but rather working rules that serve various communities, and there are a fairly large number of citation formats in common use.

Cataloging practices for networked information resources is an area that is currently under very active discussion. Several groups within the American Library Association (in particular, MARBI and CC: DA) are studying this issue and working on guidelines in association with groups that include the Library of Congress, OCLC, the Coalition for Networked Information, and the IETF. Some of the issues involved here are very complex, and not yet well understood; indeed, some of the questions involve very basic considerations about the purposes and objectives of cataloging. Taxonomies for classifying networked information resources are also needed, and still poorly understood. The current drafts [Library of Congress, 1991b; Library of Congress, 1993] from the American Library Association's MARBI committee again recognizes the use of the IETF locator and identifier structure as an appropriate means of encoding some needed information, and foresees a conversion to these standards as they are established, while also supplying some provisional field definitions that can be used by catalogers who wish to experiment with cataloging network resources in the interim.

It is also important to recognize that cataloging is only a part of the broader question of how to provide information to help users to identify and select networked information resources. Cataloging is concerned primarily with description and organization of materials (for example, through assignment of subject headings within some classification structure and vocabulary, or through the development of name authority files that bring together works published by the same author under different names, or different variations of a single name); equally important information which would allow someone to obtain evaluative information about a resource or to compare one resource to another is outside the scope of cataloging. Such information is provided by book reviews, consumer information services, ratings services, critical bibliographies, awards given by various groups, sales figures and other tools. All of these services-and new ones, such a certification that software works properly in a given environment or is free from viruses, for example—will need to be evolved into the networked information environment but with some new and challenging additions. One key objective will be to preserve, and if possible to expand the diversity of evaluative sources that information seekers can consult if they wish; just as one promise of the networked information

environment is an increased pluralism in available information resources, a parallel diversity of facilities for selecting from these information riches is essential.

The tools and methods of selection and evaluation must become more diverse and flexible. Today, virtually all evaluative information is intended for direct human consumption; a person reads a review or rating service and then perhaps makes a decision to acquire a product or use a service. It seems clear that in order to manage the overwhelming and dynamic flood of information that will occur in the networked environment we will need to develop software tools to help us in selecting information resources and navigating among them. Encoding and knowledge representation for evaluative information, and in fact even the definition of appropriate data on which to base selection decisions are areas in which research and innovation are desperately needed, along with all of the accompanying issues of algorithm design for software to assist in such decision making ; indeed, the lack of progress in this area may prove to be a significant limiting factor achieving the promise of a large scale networked information environment.

## 14. Directories and Catalogs of Networked Information Resources

As networked information resources multiply, one of the central issues will be locating appropriate resources to meet various needs for information [Lynch & Preston, 1992]. There are many tools that have evolved for identifying various types of information resources for various purposes, and many organizations that produce these tools for many reasons.

Libraries have played a role in this area through their collections (and the choices they have made in selecting and acquiring these collections), their catalogs, and the bibliographies and directories that they make available to their patrons. However, in the electronic environment, the role and content of these tools for locating and identifying information are changing. One important and problematic issue is the relationship between library catalogs and networked information resources. In the print world, one can distinguish the cafalog, which describes and provides access to material held by a given library from the *bibliography,* which defines and provides access to the literature on a given subject without regard to where that literature is held (and typically does not provide the user of the bibliography with any information that would help this user in physically obtaining access to material listed in the bibliography) [Buckland, 1988] .88

88 Basically for economic reasons, the coverage of library catalogs is typically limited. Since the early part of the century, libraries have typically been unable to afford to catalog the individual articles in journals that they receive, so they only catalog at the journal level. Bibliographies (or abstracting and indexing databases, which are simply the electronic successors to printed bibliographies) are used to obtain access to journals at the article level; library **catalogs are then used to determine if the library holds the journal containing the desired articles. So-called online library catalogs today typically at large research libraries offer access not only to the library's catalog, but also to some abstracting and indexing databases (bibliographies); a few systems offer the ability to view the bibliography as a form of catalog by permitting users to limit searches to articles in journals held by the library.** This is accomplished by having the library's online information system link the library's catalog database to the journal titles covered by the abstracting and indexing database. A few systems, such as the University of California's MELVYL system, or OCLC's EPIC/FirstSearch service have gone a step further and also linked the journal holdings of other universities to these bibliographies, thus in some sense transforming the bibliography into a union catalog of holdings in a specific discipline (though not a comprehensive one, since there are undoubtedly journals

Some leaders in the library community have discussed the transition to networked electronic information as a transition from the role of libraries in creating physical collections to a new role as providers of access to information that may be physically stored anywhere but is available through the network. In this new environment, the role of the library catalog in permitting users to identify relevant electronic information is problematic. One scenario calls for libraries to include in their local catalogs descriptions of networked information resources that the library chooses to logically "acquire" (either simply by selecting them and placing descriptive records for them in the local catalog, or in the case of fee-based services paying some type of license fee, or subsidizing transactional usage fees on behalf of the library's user community in addition to adding the descriptive record to the local catalog). An alternative scenario calls for libraries to simply provide their users with access to external catalogs, directories or bibliographies of networked information resources and to assist patrons in accessing these resources; in this scenario the "selection" or "acquisition" decisions of the library are accomplished at two levels: first, by the choice of external databases that they offer their patrons which describe available networked information resources, and secondly by the extent to which the library allocates both staff and financial resources to helping patrons to use different networked information resources, and to subsidize the costs incurred by use of these resources. Complicating the picture in either case is the inevitable development of various directories and bibliographies of networked information resources by other organizations that will be accessible to the library's patrons, in some cases for free and in other cases for fee.

It is also important to recognize that there will be a lengthy transitional period where libraries may provide access to directories of information resources and abstracting and indexing databases in electronic form, but during which most of the primary material, such as journal articles, will continue to exist in printed form. Linkages from electronic directories, bibliographies, abstracting and indexing databases, and online catalogs to the print holdings of libraries will be of central importance for at least the next decade. Experience has shown that these linkages are difficult to establish without human editorial intervention by simply matching on unique numbers such as the International Standard Serials Number (ISSN); yet the establishment of such linkages reliably will be of central importance in providing access to current library resources. Additionally, such links are essential in making effective, economic interlibrary loan and document supply services feasible.

Realistically, it seems likely that libraries will seek a compromise solution with regard to the representation of networked information resources in their local catalogs, probably including descriptive records for resources that they believe are important enough to spend money acquiring access to on behalf of their user community and for some carefully selected free public-access resources deemed to be of significance to their patrons. For access to other resources, patrons will be guided to external databases on the network, and libraries will develop policies about the extent to which they will subsidize and assist use of these external directories and the resources listed in them by various segments of the library's user community (in much the same sense that university research libraries today will go to considerable lengths to obtain access to

---

relevant to the discipline that are not covered by the producers of the abstracting and indexing databases). So, there is already growing ambiguity as to the boundaries between bibliographies and catalogs.

arbitrary material through interlibrary loan or purchase for faculty,[89] for example, but might charge students for a similar service if they offer it at all).

Not all identification or use of networked resources will take place through libraries, of course. Just as today people also identify and/or acquire material by reading advertising, browsing in bookstores, scanning book reviews, joining book clubs or by word of mouth, similar routes will be taken to electronic information resources. The only cause for concerns here are those of balance. While university research libraries are actively addressing access to networked information resources, the vast majority of public and school libraries lag far behind and lack the resources or expertise to address these new information sources; indeed many such libraries are today struggling just to survive and to continue to provide their traditional services. For many people without access to major research libraries, the primary routes to identifying networked information of interest may not be through libraries at all, but rather through information services on the network. But the level of these network information services has been disappointing, up till now; perhaps in future competing commercial services will improve the level of service, but at the cost of reducing equality of access.

But consider: while libraries, depending on their mission, budget, and patron community will vary in scope and depth of collections, one of the primary tenets of library collection development is to provide a broad, diverse, and representative selection of sources on areas that are within the scope of the library's mission. It is unclear to what extent other groups providing directories of networked information resources will reflect these goals of libraries; some directories will undoubtedly be forms of advertising, where a resource provider pays to be listed and is listed only upon payment of such a fee. Some databases of resources may be essentially the electronic analog of bookstore inventories, with all of the criteria for inclusion that such a role implies. Other directories may be built as "public services" by organizations with specific agendas and specific points of view to communicate. Services will develop that provide very biased and specific selection criteria for the material that they list in their directories; this will be a very real added value for their users, who in some cases will pay substantial sums for the filtering provided by these review and evaluation services There is nothing wrong with such directories; indeed they provide real value, offer essential services, and also ensure the basic rights of individuals and organizations to make their points of view

89 Specific mention should be made of the changing nature of the use Of interlibrary loan to Permit a library to obtain material on behalf of its users. Consider first the major research library; historically, interlibrary loan was used primarily as a means of providing fairly esoteric research materials to faculty when they were not held by the local library. With the growing inability of even research libraries to acquire a the bulk of the scholarly publications in a given area, we are seeing the use of ILL even to support requests from undergraduates. ILL is no longer used simply for esoteric research materials. Another important issue is the independent scholar—this might be an individual conducting independent research, a staff member at a small start up company that does not have a library, an inventor, or even a bright high school student: in all of these cases, the information seeker will most likely use a local public library and the ILL system to obtain access to the research literature. Such requests are relatively rare, and a decade or two ago were accommodated fairly routinely through the ILL system when they occurred; today, with the increased emphasis on cost recovery as a reaction to the overloading of the ILL system, the barriers to access by such disenfranchised patron communities are multiplying rapidly. There is a real danger than within the next few years the research literature will be essentially inaccessible to those library patrons who are not part of the primary user community of a major research library. This is a major threat to equitable access to knowledge, and one that may have some serious long-term societal implications, ranging from frustrating bright young students through handicapping the independent inventor or scholar.

known. But there is, I believe, cause for concern if the at least relatively "neutral" service offered by libraries is not among the options for seekers of information in the networked environment.

## 15. Conclusions

### *Integrity and Access Issues in the Broader Networked Information Context*

Before attempting to summarize or draw conclusions from the material covered in this paper, it is vital to put the issues reviewed here in perspective. This report has concentrated on problems and open issues. In some cases it sketches a rather bleak picture, particularly in regard to the role of libraries as publishers move towards electronic information products. It has outlined a growing array of threats to information consumer privacy in the networked environment. Indeed, the purpose of the report is to highlight these issues and problems.

It is important to recognize and address these issues precisely because the potential of networked information is so significant. Realizing this promise is of central importance. Information technology and network-based access to a rich array of information resources can change our educational institutions (in the broadest sense not only of elementary and higher education, but of lifelong learning), our political system, our economic frameworks, and our culture. Visions of futures in which our children, anywhere in America, can browse storehouses of knowledge and cultural history available from electronic library collections, define goals which we collectively believe worthy; the question before us is how to achieve these goals. If the potentials were not so great, the issues defined here could be left to the evolving marketplace in electronic information and the continual redefinition of institutional roles that this marketplace is driving. But I believe that the promise of networked information demands conscious, deliberate choices, and, where necessary, investments to support these choices.

The other point that should be stressed is that we are in a very complex transitional period which is likely to continue to at least the end of the century. This is not only a transition from the traditional print publishing system (including the role of libraries in that system) to a system of electronic information distribution, but also to some extent a transition away from the existing system to new models for creating and controlling access to content. For example, government (at the federal, state or local level) may well commission the creation of content for use by the public, or license access to content on behalf of the public because access to this content is an essential element in the educational system (again, in the broadest sense of elementary, higher, and adult education). Authors may choose to make their creations widely available at little or no cost simply because they believe that access to these creations is of great importance to society, or because they are writing to communicate ideas rather than to make money. A new information distribution system, enabled by the ability of the network to make every participant a publisher and to disseminate materials in electronic formats widely and at very low cost, is starting to grow up alongside the traditional publication system even as this system of publication is itself transfigured. Depending on an author's goals in creating a given work, he or she may choose the traditional, copyright-controlled system based on publishers or the one of the new network-based

publication models as a distribution channel. Within this new, parallel, information distribution system using the network libraries will take on new roles and missions. This is a time of great creativity and experimentation, of exploring new roles and new models.

We are seeing signs that economics alone will not define the shape of the future. For example, in a networked environment there is a very strong tendency to centralize resources; the extreme case of this is the vision of a centralized electronic library in a given discipline that provides service worlwide.[90] While there are strong economic justifications for this sort of centralization in a networked environment since the presence of the network eliminates geographic-based use community affiliation and permits economies of scale that are amortized across national or international user communities, the predicted centralization is not clearly taking place. Rather, the networked environment is giving rise to a very pluralistic model of information storage and access; at one level, this is inefficient, as a good deal of information is stored redundantly, but at another level this is a comforting development since it re-enforces the value that we as a society place on distributed, democratic access mechanisms that lack central points of control. We have yet to fully comprehend the resolution of the conflicts between economics and cultural/institutional values.

Similarly, the destruction of the existing interlibrary loan system is not an entirely forgone conclusion; as authors, particularly authors of scholarly works, become more aware of the consequences of their actions, they are beginning to protest the confines of the existing scholarly publication system and in at least a few cases to explore alternatives, such as various forms of network-based electronic distribution of their works. There is a growing recognition that the publication system that has developed to support scholarship, teaching and research over the past centuries exists to seine these communities rather than to define their function. There is a perception within the research and higher education communities that they can define the future that they wish to live in, and that the members of these communities are responsible for defining that future. For example, I expect that there will be serious and occasionally bitter debates among the boards of scholarly societies in the next few years as the communities to which these societies are ultimately accountable wrestle with questions about whether these societies will have roles similar to for-profit publishers (perhaps subsidizing other activities of the society with profits from publication programs) or whether they will return to their original functions of facilitating communication and diffusion of new knowledge within scholarly communities, even if this means distributing their publications at little or no cost on the network and loosing the revenue that these publications generate (and presumably finding new financial models for supporting the society's activities and publications). This reevaluation of the roles of the existing system of publication in meeting the needs of the scholarly community is likely to be painful and acrimonious, since whatever their origins both commercial scholarly publishers and many professional societies which function as publishers are now very large and profitable businesses that will resist changes diminishing their size, income and influence.

90 Technically, such a facility is likely to be mounted on multiple hosts, probably at multiple sites, 'n order to provide some redundancy in case of disaster and to permit scaling to very large user communities. But, organizationally, the model is one of a single monolithic institution providing access to information.

As we look beyond the research and higher education communities, the picture becomes less clear, as the motivations of key stakeholders become more clearly profit-oriented and the sense of accountability to a community becomes weaker. When one considers the role of advertising, and the corporations that advertising serves in the development of the electronic mass media to date, one cannot be sanguine about predicting a future in which these media are held directly accountable for furthering the public good. Perhaps we can see the start of a divergence here between the research and education community and the general information consuming public (recognizing of course that many individuals participate in both communities to a greater or lesser extent at various points in their lives). The research and education community, which ultimately creates and can control most of the information it uses, is beginning to take responsibility for its own transformation into the networked information environment. On the other hand, the populace as a whole (including the public library system that serves this general populace) does not in any real sense create the information that it consumes, or control this information except in the most indirect ways (the power of the consumer's dollars in the marketplace and the power of the consumer's vote in developing public policy); content and the means of access to information are controlled by relatively unaccountable organizations like commercial corporations. In the general case, we are a society of information consumers who view ourselves at the mercy of information providers. The electronic information world of the general public may well be defined primarily by entertainment video libraries, interactive games, shop-at-home services that substitute for the printed catalogs that clog our mailboxes today, and "infotainment" segments advertising the latest in personal growth, weight loss, business success, and the like, with market researchers lurking in the wings to accumulate (electronic) mailing lists of qualified prospects. Here it is important that libraries, government information, and information from the scholarly community, as well as many diverse viewpoints from the general public on issues of importance maintain a presence among the information sources offered to the general public through the network, even if, following the patterns of today's broadcast mass media and print publications, such materials are only modestly used by the general public. Ensuring this continued presence is an important public policy objective. There is considerable precedent for this; for example, in the broadcast media the offerings of the Public Broadcasting System are not typically the highest rated programming, but they are offerings that make important contributions to our society in many different ways.

There is no question in my mind but that we will solve the problems and address the issues raised in this paper. The progress of information technology is inexorable; the promises and advantages compelling and the payoff enormous. It is clear that the private sector has now recognized the potential marketplace that networked information of various types represents, and has begun to commit massive financial resources to develop this marketplace. If not already the case, the scope of this private sector commitment will soon overwhelm the resources that the research and education community and the government have already contributed to seed and nurture development of the networked information environment. This will create additional pressures to address and resolve the issues quickly. It may also introduce a new pragmatism and expediency into the development of these solutions; while academics and policy makers sometimes debate issues at great length, the need to ship products,

launch services and recover investments is a great motivation to come up with some sort of practical solution and get it implemented in a timely fashion. The growing private sector pressures will also create considerable tensions and controversies, since solutions acceptable in the commercial marketplaces (and desired by the private sector) may not be entirely acceptable to the research and education community or to makers of public policy.

The challenge before us, then, is to ensure that we address the issues and solve the problems in the most timely way possible while, to the maximum extent possible, incorporating and balancing the interests and concerns of public policy, of the research and education community, and the private sector in these solutions. Speed is important; without timely progress we face the risk of being overrun by marketplace developments, which are not likely to reflect the balance of interests that I believe is essential for a future that will offer not only the commercial payoff but also the improvements in research, education, and the extent to which the public is informed. And balance is also vital: the interests of the various sectors involved are in many cases conflicting, and a deliberately and thoughtfully crafted balance among them will be needed to achieve the future that we desire. The importance of developing this balance is too great to be left entirely to the chance and marketplace forces.

### *Ensuring Access to Information in the Networked Information Environment*

Publication, whether in print or in electronic form, is the act of making a unit of information available to the public, perhaps at some price. These individual units represent intellectual property for which the authors and/or publishers are frequently compensated. This is as it should be. At the same time, when all of these publications are aggregated, they form a major part of our societal, cultural and scholarly record and serve as a repository for our collective knowledge. Ensuring that our children, scholars, researchers, indeed all of our citizens, have some reasonable level of access to this collective body of information both when it first appears and even many decades later is a vitally important public policy objective. Today, this public policy goal is implemented by the provisions of the copyright law and by institutions such as libraries. The copyright law and the doctrine of first sale help to ensure that libraries exist and can effectively function; however, with some relatively modest exceptions, while the operation of libraries seems to be generally accepted as a public policy goal, the libraries of America are enabled more than they are mandated by specific federal legislation.

As this paper has shown, the mechanics of "publication", its legal framework and perhaps even its definition are changing in important ways in the electronic environment. Further, as has been discussed, new forms of information rather different than the traditional published works collected by libraries are taking on increased importance: these include the contents of the electronic mass media and also the so-called secondary information sources (such as abstracting and indexing databases) which, when joined with the searching capabilities of computers, provide new and powerful tools for managing and navigating the growing primary literature. The public policy goals of creating and maintaining a reasonable level of citizen access to the published literature remain, but we may need to find new ways to achieve these goals. There are questions both of access to relatively current material and continued access

to the societal and scholarly record in the long term. This changing legal framework is making it very difficult for libraries to continue to fulfill the functions that they have traditionally performed in support of these public policy goals. Either changes must be made to permit libraries to continue to perform these functions, or some new or redefined set of institutions must be established and empowered to do so. There are many possibilities, some of which have been at least superficially mentioned in this paper, including changes to the copyright law (such as mandatory licensing), the creation of increased amounts of information or licensing of information at a national level, or changes to the depository provisions of the copyright law to ensure that copies of electronic works are registered with some institution responsible for their long term preservation. One can imagine a number of other legislative or regulatory approaches to addressing these issues.

One of the problems today is the general uncertainty surrounding intellectual property law as it relates to electronic information. This sense of uncertainty is both inhibiting progress and driving some developments that may well be undesirable from a public policy point of view (such as the increased use of contract law and licensing to control electronic information). Resolving these intellectual property questions in the courts will be a very slow and costly process and one that only increases the sense of uncertainty and risk surrounding electronic information. One alternative would be legislative action to clarify the issues and in some cases perhaps implement specific changes in support of public policy objectives. But, in an area as complex as intellectual property law changes will have to be made with great care and great wisdom; further, because intellectual property laws potentially impact so many areas of the economy and society (and also have important international implications) it may be difficult to develop a successful consensus on changes driven by the needs and public policy objectives related to networked information within this much broader community. There are other possible ways to make progress and reduce uncertainty, such as guidelines developed among the stakeholders which do not have the force of law, but which provide generally agreed upon rules of acceptable behavior; the model of the National Committee on New Technological Uses of Copyrighted Works (CONTU) with regard to the development of guidelines for interpreting the copyright law in the context of new technologies may be relevant here. CONTU both helped to clarify and obtain some consensus on issues, and also paved the way for subsequent legislative changes.

The purpose of this paper, however, is to make the reader aware of the growing problems in achieving the public policy goals related to access in an environment that is increasingly moving towards electronic information, and to provide background for an informed discussion of solutions, rather than to explore the ramifications of the various proposed solutions in detail. These problems are real and growing. But, I believe that our strength here is that as a society we have a reasonable consensus on the public policy goals, though there will always be debates about how much access is enough and how such access should be financed, as well as the nature of the implementation mechanisms and the continual tuning of the balance between rightsholders and the public.

Finally, I would note that federal government information has a very special role in the developing networked information environment. If it is made publicly available at little or no cost it will be a very widely used and important information source in the networked environment. Indeed, the creation and distribution of inexpensive high quality

information resources can be an effective instrument of public policy; one need only consider the enormous impacts that databases like MEDLINE from the National Library of Medicine and ERIC from the Department of Education have had in vastly improving access to and use of published information in the biomedical and health sciences and education respectively [U.S. Congress, 1990]. Federal leadership in information policy related to the electronic distribution of public information would also be helpful to state and local government in developing policies and recognizing the advantages that networked information access and distribution offer. Finally, large amounts of federal information can be used as a testbed for developing and proving standards, technologies and systems without the complexities, costs or limited and closed user communities that would typically be required if licensed commercial information was used in such experiments.

### *Privacy, Confidentiality and Anonymity in Access to Electronic Information*

If there is relatively good consensus on the importance of access to information to our society, I believe that there is much less consensus about issues related to privacy, confidentiality, and rights to anonymous access. This lack of consensus goes far beyond simply access to the published works and to the societal and scholarly record, and is clearly seen in the many public policy debates related to privacy and confidentiality generally (for example, credit reporting, medical records, public records, computer matching of various types of files, debates about cryptography) as well as the conflicts between the cultures and perhaps values of libraries, the computing and computer networking communities, and the commercial world that have been illustrated here. The ability of information technology to provide easy access to and permit the analysis of vast amounts of information has implications that we are just beginning to understand. Further, as this paper has illustrated, there are many subtle questions related to the use, compilation, and analysis of histories of access to and use of information even in cases where users may be anonymous.

Hopefully, the paper will give the reader a sense of the scope, complexity and subtlety of the issues in this area. While perhaps there are a few areas, such as confidentiality of some types of records, on which there is general consensus and which might be addressed quickly, my sense is that it will be necessary to conduct an extensive policy debate with the objective of defining public policy goals before a great deal of progress can be made. In the meantime, to some extent, the best that can be hoped for is that users of electronic information become more aware of the privacy and confidentiality issues involved in their use of electronic information resources so that they can make more informed choices.

Many of the privacy and confidentiality issues discussed in this paper are peculiar in that they can be addressed on two levels: the legislative/policy level and the technological level. The technological solutions are often in turn driven by marketplace perceptions about the value of privacy and confidentiality; if consumers recognize that a serious problem exists and are sufficiently concerned to pay for the implementation of a solution, that solution will often become available. Legislation can, of course, also mandate the implementation of technological solutions, but this is rare. In my view, the technological solutions are often more robust than the legal ones, because the legal

restrictions are very difficult to enforce. Consider as an example the controversy about scanners for cellular telephones, and let us ignore the issues about consumer use of encryption and exportability of products incorporating cryptographic technology touched on elsewhere in this paper. A cellular telephone user concerned about privacy could purchase an encryption device which would provide a high assurance of privacy, at the cost of some inconvenience and subject to the limitation that secure communication would be possible only with other owners of a compatible encryption device. A few cellular phone users did so. Legislation was passed making scanners to eavesdrop on cellular phones illegal; however, such scanners were widely available and it seems likely that anyone who really wants one could still purchase or build one. So, the effect of the legislation has been to provide most cellular phone users (who have not purchased encryption devices) a false sense of security; while cellular phone eavesdropping as a consumer "sport" has no doubt been curtailed, I would suggest that the real problem hasn't been solved. A more effective solution would have been to either establish standards for cellular phone encryption and encourage the marketplace to implement them (and mount a campaign to make sure that users were aware of the risks of purchasing a phone that did not implement encryption) or perhaps even to mandate the inclusion of such encryption devices in new cellular phones.

Very similar problems apply in the case of services that require users to import software onto their personal machines for access, and where that software may collect and export information back to the service in question. While it might be possible to craft legislation to prohibit such practices, this would have to be done very carefully so as not to prevent legitimate and valuable applications. Further, as discussed, the consumer might well be willing to permit export of certain information in return for other considerations such as free or discounted access to services. A better choice here, in my view, would be to combine efforts to inform consumers (including perhaps some sort of labeling disclosure requirement on commercial software that exports information) with investment to develop good technology to permit the consumer to monitor and control the export of information from his or her personal machines. The difference between the cellular telephone example and many of the problems discussed in this paper, however, is that we do not currently have good technological solutions ready to deploy, and thus research investments are likely to be required.

### *Infrastructure and Standards*

There is a tremendous amount that needs to be done to establish a viable infrastructure for electronic information and to ensure that it can become an effective, manageable part of our scholarly and societal record. A good deal of this work is neutral with regard to the public policy questions raised in this paper (although accomplishing these tasks will require that other public policy questions be addressed, such as those related to cryptography). Much of what is needed is simply funding (for research, experimentation and analysis and evaluation of experiments), standards development (discussed in more detail below), authoring and distribution of public domain computer software to help establish a critical mass of implementations in support of selected standards,[91] and to seed the construction of at least some parts of

91 There are numerous success stories in this area that deserve consideration. Software authored by universities and publicly distributed over the network without cost has led to the deployment of a number of important new network-based information services, such as Gopher. The availability of such software has

the infrastructure that will support networked information; the research and education community and the private sector are already working actively in these areas and are making considerable progress, as Section 2 suggests, but funding sources are few and funding is often a problem. I believe there is reason to be optimistic that some of the legislation currently under consideration will help to address these areas. Leadership in forging partnerships among the research and higher education communities, industry and government is also an important part of the effort required.

A few specific points should be emphasized with regard to the needs for funding. First, funding the infrastructure of the computer-communications network is certainly a prerequisite for the development of networked information, but there is additional infrastructure investment needed over and above that for the web of transmission facilities, switches and other technology necessary to create the communications network itself. This paper has discussed some of the areas in which investment will be needed, such as: systems to support integrity and authentication; systems to permit the location and identification of networked information resources; directories and catalogs to permit network users to find relevant information resources; systems to create, disseminate and view multimedia electronic works. Thus far, the vast majority of the funding invested in encouraging the development of networks at the federal level has gone towards building the communications infrastructure, and, while this investment has been quite successful to date (to the extent that there is serious discussion about when what parts of the communications infrastructure should transition entirely to the private sector) the facilities to support networked information are not nearly as extensive or advanced. The need for federal investment in the networked information infrastructure has not passed, and this should not be overlooked in discussions focused on the need for future federal support for the communications infrastructure.

Also, as a community I do not believe we yet understand how to solve a number of the technical and management problems related to networked information. There is a very real need for funding to support research and experimentation, including the implementation, testing and evaluation of a number of fairly large scale prototypes. The ability to test and learn from multiple approaches will be very important in guiding the development of technology in this area. In addition, we must be sure that there is funding not only for implementation but also for the follow-up evaluations and studies that permit us to really gain the full benefit from pilot projects. Further, there is relatively little basic theory to guide engineering projects in networked information, and much of the research in the field has a very pragmatic, near term focus on developing operational prototypes. A case can be made that this needs to be balanced by funding for more "basic" longer-term research.

---

**served as** a stimulus for additional software development by other institutions as well as widespread implementation of the services themselves. Industry has also made good use of this approach; one notable contribution here is the Wide Area Information Server (WAIS) system developed by Thinking Machines, Apple Computer, Dow Jones and KPMG. Finally, it is important to recognize that many people in the computer networking community believe that the funding that the Defense Advanced Projects Research Agency (DARPA) provided for the incorporation of the TCP/IP protocols into the UNIX operating system at the University of California at Berkeley during the 1980s was a critical factor in the success and explosive growth of both the Internet and the UNIX system.

Finally, this paper has not really discussed where the people will come from who will build and manage the networked information environment; while this is somewhat out of scope for a study of the integrity and access issues in electronic information (other than to point out the obvious, that there will be a need for trained and skilled individuals to manage the information and insure its integrity, and to help information seekers to gain access to it). From the point of view of developing the necessary technology and standards base and actually building the infrastructure, however, there is a developing shortage of people with the necessary combination of expertise. It is necessary for the higher education community to begin now to design and implement appropriate academic programs to develop a large pool of people who can contribute to designing and building the networked information enterprise; some universities have already begun this process, typically building on programs in library and information studies as a starting point. My view is that this is really in some sense a new field, though one that builds extensively on computer science, traditional library and information studies, communications technology, public policy and other disciplines. Funding to support academic research and the development of academic programs to support networked information can thus be viewed as part of the infrastructure investment that will be needed.

There are two particular problem areas impeding the development of the necessary infrastructure. The first, which has been discussed extensively in this paper, is the set of barriers surrounding the large-scale use of cryptographic techniques to implement the authentication and integrity functions that will be essential to the use of electronic information. The impact of these barriers is not limited to electronic information access and integrity; it also poses problems for a number of other network-based applications., including commercial transactions of various kinds. Resolving these problems, I fear, will require nothing less than the development of a rational, clearly articulated national policy on cryptography. There is, in my view, an urgent need for action in this area.

The second problem area is standards. As the paper has illustrated, standards are a key to developing the infrastructure, and also a central part of the strategy for ensuring that electronic information continues to be available in the fact of continual changes and improvements in the technology base used to house and deliver it. Yet the necessary standards are not in place yet in many cases and many of those that have been established are little used in the real world of large-scale, operationally deployed systems and products. Getting the appropriate standards defined, disseminated, and implemented in the marketplace is essential to progress in infrastructure.

There are five major groups of standards-developing organizations functioning today in areas relevant to information technology, electronic information and computer networking:

- International standards bodies, such as the International Organization for Standardization (ISO).

- National standards bodies in the US which link to the formal international organizations, such as the American National Standards Institute and its accredited standards writing bodies (for example, the National Information Standards Organization, NISO, which serves the library, publishing and information services communities).

•The National Institute of Standards and Technology (NIST; formerly the National Bureau of Standards, NBS) which develops standards for the federal government and also is charged to provide leadership in developing standards for the US generally in some situations where progress in standards is critical to US national interests and the private sector is not making sufficient progress.

•A growing array of ac-hoc industry standards development groups, consisting primarily but not exclusively of corporations; these are typically focused on a single problem. Examples include the UNICODE consortium, the Open Software Foundation (OSF), the Object Management Group, and many others.

•The Internet Engineering Task Force (IETF), an informal standards-writing group that manages standards for the Internet and is increasingly also concerned with developing standards to enable and facilitate the use of electronic information resources in the Internet environment.

There are major problems in the standards development system today. A full exploration of these is far outside the scope of this paper; the recent Office of Technology Assessment study Global *Standards: Building Blocks for the Future [U.S.* Congress, 1992], touches on a number of these problems but emphasizes the international perspective and standards in all areas, not just information technology and electronic information. Basically, from the perspective of building the networked information infrastructure, the speed with which formal standards (that is, standards within the ANSI/ISO structure and process) can be developed is too slow, leading to increased reliance on mechanisms like ad-hoc industry groups and the Internet Engineering Task Force. The costs for developing all types of standards have become very high; these high costs are largely precluding the effective participation of many of the communities involved in networked information in the standards development process. The refusal of the formal standards bodies to make their products available at reasonable cost and in electronic form has increasingly limited the usefulness of these products., particularly in disciplines like computer networking; by contrast, the IETF, which makes all of its work publicly available on the Internet, is gaining increased acceptance as a standards developer in many quarters, even though it is outside of the formal standards establishment. Finally, there is a growing perception among many of the people actually involved in building networks and the networked information infrastructure that the formal standards establishment has lost touch with engineering reality; the standards being developed by these groups are not being implemented in the marketplace and existing marketplace standards are not being reflected in the work of the formal standards bodies.[92] To some extent, at least, this problem is being created by conflicts between international demands, politics and commitments, and the policies of other nations regarding technology, standards, and the development of

92 This problem is perhaps most evident in the controversies surrounding the two competing networking standards suites: TCP/IP, which is the protocol that forms the basis of the Internet and is managed by the IETF, and is *not* a formal international standard, and the Open System Interconnection (0S1) protocol suite, which is a large and complex (and not yet complete) set of formal international networking standards that have been under development for about 15 years, but still have not gained large scale marketplace acceptance, despite attempts by various governments (including the US Government) to mandate their use. The history of this controversy is extremely complex and involves a number of political and economic as well as technical factors.

network infrastructure (and pressures of global marketplaces) and the requirements of the networked information community in the United States (which is certainly the dominant force in this area today) to build an effective infrastructure quickly and at reasonable cost. One consequence of this is that there is a considerable amount of at least partially duplicative work taking place, and the competing standards are causing confusion among the user community. There is also at least some anecdotal evidence that the private sector is increasingly turning away from the formal standards-setting process in frustration.

Some of the problems related to standards are funding problems; for example, it seems likely that relatively small investments, properly structured and applied, could help a great deal with the problem of broad community participation in networked information standards development and the speed with which such standards are developed. Encouraging marketplace implementation of standards is much more difficult; there is a fine art in developing and selecting standards that are viable and appropriate. To some extent this is the responsibility of purchasers, but today purchasers are often defeated in their attempts to acquire standards-conformant products by the lack of reasonable standards available to specify. Addressing the remain issues presents very complex policy and management problems for standards organizations and the communities that they serve, and, ultimately, for the United States government itself. A review of policies in the standards area as they relate to networking and networked information, with some specific emphasis on the relationship between policy choices and timely and effective progress on the construction of network and networked information infrastructure in the United States, is needed.

## 16. Recommendations for Possible Action

### *Legislative, Government and Public Policy Actions*

•It is clear that cryptographic technology will be required to ensure the integrity of electronic information. Incorporation of this technology on a broad basis in the networked environment (including internationally) has been effectively paralyzed by a series of issues concerning intellectual property, standards, and exprot controls. It is time to address these issues in a systematic way and develop policies that will guide and encourage the implementation of the appropriate and needed integrity and authentication functions.

•Intellectual property issues are central to ensuring access and integrity for electronic information, particularly **as** this information becomes an increasingly substantial and important part of our intellectual record as a society. Consideration needs to be given both to clarifying the existing intellectual property laws as they relate to various forms of electronic information, particularly in a networked environment, and also to a review of whether the current balance between rightsholders and the public as defined in the copyright laws will continue to accomplish public policy objectives in the networked information environment. Immediate legislative action may well not be the answer; rather, the formation of a group similar to the National Commission on New Technological Uses of Copyrighted Works (CONTU) in the late 1970s might be an effective way to make progress at this time.

•A policy for networked access to federal government information which encourages such access at very reasonable costs is important both in its own right and also to stimulate technology development, network use, and to serve as a model for information policy development by governments at the state and local levels.

•A public policy debate related to expectations of privacy, confidentiality and anonymity in access to electronic information is needed to establish consensus on the objectives to be achieved by policy development and legislation. In addition, actions are needed to help make the public aware of the current state of affairs so that informed choices can be made by individuals. It may be appropriate to cast this effort in a larger context of individual privacy.

•Funding is needed to support research, development, prototype experiments, standards work, and networked information infrastructure construction. This is not the same as funding the development of the networks themselves. A substantial part of this need may be addressed by legislation currently under consideration by Congress. I believe that it is important to recognize that a substantial research component is needed here, and that a number of diverse, moderate scale prototypes will serve us better at this time than simply subsidizing the construction of one or two very large operational prototypes of "digital libraries".

•A major policy review of the standards development process and the organizations involved in this process is needed, in the context of information technology, electronic information, and computer-communications networks with consideration of both international implications and national needs. This should be conducted jointly with the standards development community and also seek broad participation by users of standards.

•A number of steps could be taken to ensure the public's access to electronic information resources. This might include some specific funding to help the existing library system, license of certain collections of information for unlimited use nationally, support to help schools license access to information, or other measures.

•Further consideration should be given to ways in which government funded electronic information resources, particular in the networked information environment, can help to achieve public policy goals such as controlling the increase in health care costs and improving the effectiveness of the educational system and national competitiveness. Such resources have proven effective in the past.

•Efforts to ensure the preservation of the cultural, historical and scholarly record as this becomes increasingly composed of electronic information are needed. This involves not just copyrighted information from publishers and other information providers but also public information. There are a number of government and nonprofit organizations with interests in this area, including (to name only a few) the National Endowment for the Humanities, the National Archives, the Library of Congress (both in its role as a library and as the registry for copyright), the Commission on Preservation and Access, the National Science Foundation, the National Library of Medicine, the National Library of Agriculture and the Association of Research Libraries. Consideration needs to be given to how to most effectively coordinate efforts on these issues.

***Actions by the Stakeholders: Authors, Publishers, Libraries, Information Technology Providers and the Education Community***

•There is a need for continued discussion among the stake holders as to the meaning of publication in the networked information environment and community expectations about continuity of access to electronic works, integrity of such works, and related topics. Organizations like the Coalition for Networked Information can play a key role in facilitating such discussions.

•A greater understanding of the increasing diversity of publication or information distribution paths available to authors and the implications of choices among these paths for the library, research and education, and publisher communities is needed. This may include some reassessment of the valuation placed on electronic publication channels in areas such as university tenure and promotion decisions, for example.

•Increased investment in the development and implementation of appropriate standards to facilitate the authoring, distribution and preservation of networked information is required. Further, these standards should reflect the evolving consensus about integrity and access to networked information.

•Publishers and libraries need to attempt again to reach a compromise about the uses of new technologies and their relation to copyrighted works which addresses the concerns of both parties. This could take the form of agreements about some limited' use of licensed electronic information for interlibrary loan, for example.

•Libraries, publishers and the scholarly community need to begin discussions about their roles and responsibilities in preserving and ensuring long-term access to the scholarly and cultural record. Broadcast media providers also need to be brought into these discussions. This is an area where it seems the financial stakes may not be high (unlike, perhaps, the interlibrary sharing of licensed electronic information) but which has great societal importance, and where it may be possible to make significant progress quickly. Such discussions might also provide a basis for informing future legislative options if appropriate.

## Glossary

archie. Archie was developed at McGill University by Peter Deutsch and Alan Emtage, who have since established a private corporation (Bunyip) to further develop and commercialize the technology. In essence, archie automatically creates, manages, and offers access to a database of the contents of the major FTP archives worldwide and thus permits users to locate FTP archives holding files of interest to them. Note that archie is with a small "a".

*Ariel.* Ariel is a system developed by RLG (which see) that permits documents to be faxed using the Internet rather than the dial telephone network as a transmission medium. Essentially, documents are scanned into a file on the sending host, transmitted across the network using FTP (which see) and then printed (or viewed on screen) at the receiving ARIEL node. ARIEL is based on IBM personal computer hardware platforms.

*ARL.* The Association of Research Libraries is a not-for-profit organization representing 119 major research libraries in the United States and Canada. ARL supports a wide range of activities, including studies of the function and costs of the research library system, the development and articulation of policy positions on legislation and other government activities that are of interest to the research library community, and development of resource sharing agreements and policies among its members. ARL is also one of the three parent organizations that created the Coalition for Networked Information (CNI).

BRS. BRS, which is now a part of the Maxwell communications empire, is another commercial search service very similar to Dialog (which see), although somewhat less extensive.

*CARL.* The Colorado Alliance of Research Libraries (CARL) is an organization based in Denver Colorado that started out building an online catalog for a number of Colorado based libraries, but has recently expanded nationally, offering a linked collection of catalogs for major libraries from Hawaii to Maryland. In addition, CARL creates and offers access to a database of journal tables of contents, a companion document delivery service, and a number of commercial databases.

Center *For Research Libraries.* The Center for Research Libraries (CRL) is an organization set up by a group of research libraries that essentially serves as a central depository for very rarely used material that is considered to be important to retain within the research library community but which is not used enough to justify any single library in this community retaining as part of the local collection; in these cases the material is sent to CRL where it can be available to the entire research library community but can be stored in a relatively inexpensive facility.

*C/V/.* The Coalition for Networked Information (CNI) is a joint project of EDUCOM and CAUSE, two associations concerned with information technology in higher education, and the Association for Research Libraries, an association of the 119 largest research

libraries in North America. Its purpose is to advance scholarship and intellectual productivity through the use of networks and information technology, and it has played an active role in promoting the development of networked information resources and the underlying technologies necessary to implement and use them.

*CONTU.* CONTU was the National Commission on New Technological Uses of Copyrighted Works which was established by Congress to make recommendations on copyright legislation in light of developments in technology. CONTU issued its final report in 1978. The work carried out by CONTU helped to define community standards for the interactions between technologies such as photocopying (xerography) and the copyright law, and to help define appropriate practices and standards for the acceptable use of these technologies in contexts such as Interlibrary Loan. CONTU also studied computer software and database intellectual property issues. CONTU made a number of recommendations for legislative changes; some were implemented and others were not. While the CONTU guidelines do not have the force of law, the represent a very real community consensus on acceptable behavior. In some cases they have been used by courts to help to interpret the copyright law, although other courts have ignored the CONTU recommendations as having no legal status.

Dialog. Dialog is a commercial online service that dates back to the late 1960s; originally a subsidiary of Lockheed, it is now owned by Knight-Ridder. Dialog essentially acts as a service bureau to database producers, mounting their databases and providing interactive searching access to them. It also handles billing and training, and typically pays royalties back to database producers who make their databases available through Dialog. Dialog is well known both for its very high prices, which have kept its use within the academic community to a minimum, and for its very complex user interface, which, while offering very powerful search capabilities, is really intended for use by trained searchers and not by end users. Dialog offers access to hundreds of databases; some of these are exclusively accessible though Dialog; others are also available though other channels (for example, CD-ROM, tape licenses direct to institutions, or other competing online services).

*Fair Use.* Fair use is a provision of the US copyright law which permits copying of copyrighted material for specific purposes, such as personal research. More generally, the term "fair use" is used to refer to the entire set of specific copying exemptions in the law, which include some provision for making a copy of an out of print work that is deteriorating for preservation purposes (if no other reasonable alternative is available), satire and criticism, and other exemptions.

*FTP.* FTP is the file transfer protocol within the TCP/IP protocol suite and widely used on the Internet to copy files from one host to another. It is the access mechanism used with FTP archives, which are simply large collections of files that are stored on various hosts on the Internet and available for copying. FTP supports an anonymous access mode that allows users to list and copy public files from many hosts on the network without any need for pre-registry with the archive host manager.

Gopher. Gopher is a system that was developed at the University of Minnesota in the early 1990s. It is a distributed system consisting of client programs for a variety of hardware platforms (including PCs, UNIX, and Macintoshes) and servers (again for

various platforms) which allow a server manager to establish a database of menus. The menu entries can point to other menus, either on the local Gopher server or any other Gopher server on the Internet, to documents (stored anywhere on the net), to interactive services, or various other types of information objects or resources. Essentially, Gopher offers a fairly simple, low cost means for an institution or even an individual to provide menu-based access to a variety of networked information (including locally stored information).

IETF. The Internet Engineering Task Force is a self-selected volunteer group that meets quarterly to develop standards for the Internet. While it is not an officially sanctioned national or international standards development body, it operates primarily by consensus and has been tremendously effective in managing the standards needed for the Internet environment. Standards developed by the IETF are called Requests for Comments (RFCs) for historical reasons. The IETF has a number of active working groups dealing with various aspects of standards and architectural models for networked information resources.

*LISTSERV.* LISTSERV (an abbreviation for list server) is a program that was originally developed for the IBM CMS operating system; more recently imitation programs with very similar functionality have been developed for the UNIX operating system environment. Essentially, the LISTSERV program permits mailing lists to be established with a wide range of parameters. Typically, users can join or leave a mailing list by sending commands to the LISTSERV program by electronic mail. Once they have joined a given mailing list, they can send a mail message to the list which will be echoed to all other subscribers of the list. The LISTSERV program also supports a variety of maintenance functions, such as maintaining and offering access to archived messages, permitting people to get lists of the people subscribed to a given mailing list, and the like. LISTSERV lists can also be defined as moderated, which means that the list moderator must approve all postings to the list before they are sent out to other subscribers.

Mail Reflectors. A mail reflector is essentially a simple, manual form of a LISTSERV (which see). Mail reflectors simply re-transmit mail to a list of users who are tabled as part of that mail reflector; in this sense a mail reflector can be viewed as simply a shorthand for sending mail to a list of people. Maintenance of the list is typically manual (perhaps assisted by computer programs) as distinct from the completely automated list management of LISTSERVs. In addition, there is usually no easy way to find out who is signed up on a given mail reflector.

Multicast. Multicast technology is a network facility which permits information to be broadcast to a specific subscriber group of machines attached to the Internet. Rather than having the source send one copy of the information to each recipient in the multicast group separately, multicasting permits the source machine to simply transmit one copy of the data, addressed to the multicast group address, onto the network; the network routing services take care of duplicating it as required so that every machine in the multicast group receives its own copy. Network hosts can join and leave a specific multicast group at will. Currently, multicast technology is used primarily on an experimental basis in the Internet (although it is a basic network service within some types of local area networking technology, such as Ethernet) to carry audio and/or

video traffic to groups of interested recipients in real time. Only a portion of the Internet (called the MBONE, or multicast backbone) supports multicast service at this time.

NNTP. The Network News Transfer Protocol is the protocol that is used by news readers that wish to access Usenet news groups from a newsgroup server. It is defined by RFC 997 [Kantor 1986].

OCLC. The Online Computer Library Center (formerly the Ohio Computer Library Center) is a not for profit organization based in Dublin, Ohio which provides a wide variety of services to the library community. These include interlibrary loan requesting and tracking, shared cataloging using an enormous copy cataloging database, and recently access to a variety of online databases (in a sense competing with commercial firms like Dialog (which see)). Recently OCLC has also partnered with the American Association for the Advancement of Science (AAAS) to mount the electronic journal Current *Clinical Trials,* and is currently working on several other electronic journals in conjunction with various other professional societies. The model for *CCT* is that of a database, where OCLC makes available the necessary user interface software to read articles, view them, and, optionally, print them, rather than the more common model used by other electronic journals where the contents of the journal are often physically transmitted to each subscriber.

PostScript. PostScript is a language developed by Adobe Systems to communicate with printers, particularly high-quality laser printers. Typically, word processors "print" a document by converting it into PostScript form; this is then sent to a printer for interpretation which produces the actual printed page. Programs also exist to preview PostScript files on display devices. It is very difficult to go backwards from PostScript to a revisable form document, or even one that permits reasonable full text searching; in addition, all semantic level markup (and even most syntactic level markup) is lost when a document is converted to PostScript form. In this sense, a PostScript document representation can be viewed as quite similar to a bitmapped image, although it is more compact.

Project *Gutenbeg.* Project Gutenberg, managed by Michael Hart, creates (either by scanning or keyboarding) ASCII versions of out-of-copyright books and other works, and makes these materials freely available over the Internet via anonymous FTP.

RLG. RLG is the Research Libraries Group, a consortium of research libraries in north America. One of the major activities of RLG is the operation of RLIN (the Research Libraries Information Network), a service similar to OCLC (which see) that provides libraries with access to a large shared cataloging database, an interlibrary loan requesting and tracking system, and a number of scholarly databases targeted for end users in academic institutions.

SGML. Standard Generalized Markup Language is an international standard for defining markup tagging for text. This can be at a relatively superficial level, where only syntactic structures such as headings and paragraphs are marked, thus facilitating the reformatting of a document for presentation in multiple environments (for example, in print or on a display terminal), or the markup can define very deep semantic meaning,

as is being doing in the Text Encoding Initiative project, which addresses the needs of the humanities computing community to encode text for subsequent computer analysis.

TELNET. TELNET is a protocol within the TCP/IP protocol suite and widely used on the Internet to conduct a terminal session with an interactive service on a remote host. It can be viewed as the network equivalent of dialing up a remote host with a character based terminal.

TOPNODE. This is a project of the Coalition for Networked Information (CNI - which see) to experiment with the description and cataloging of networked information resources.

Usenet *Newsgroups.* These are collections of electronic mail messages that are distributed throughout the Internet and beyond using the Usenet news system. There are many hundreds of such newsgroups, with more being established all the time. Total Usenet traffic is now measured in tens of millions of characters daily. Usenet news groups include the infamous "ALT" newsgroups, which are unmoderated and which deal with rather controversial issues such as sex and drugs, and which have been subject to censorship from time to time by various institutions. One subscribes to and reads Usenet news groups using a news reader. For a more extensive description of Usenet, see [Quarterman].

*WAIS.* WAIS stands for Wide Area Information Server; this is a system that is based on the **Z39.50** information retrieval protocol (which see) that was originally developed in the early 1990s by Thinking Machines, Apple, Dow Jones and KPMG. The original code was publicly distributed. Since that time a startup company (WAIS Inc.) has been established by several of the original developers to commercialize the technology, while work on the upgrading and extension of the public domain version continues through organizations like the Clearinghouse for Networked Information Discovery and Retrieval in North Carolina. Essentially, WAIS permits full text searching (using sophisticated statistical ranking and matching algorithms) against databases distributed across the Internet using a common user interface.

WWW. WWW is the WorldWide Web system originally developed by Tim Berners-Lee at the CERN center in Geneva, Switzerland. It can be viewed as a network-based hypertext system that allows linkages between arbitrary information objects stored on hosts throughout the network.

**Z39.50.** This is a US national standard developed by NISO, the National Information Standards Organization, an ANSI accredited standards developing body serving the publishing, library, and information services communities. The standard addresses computer to computer information retrieval and provides a basis for the development of network-based applications that offer a common interface to multiple, autonomously managed information servers.

## Suggested Background Readings

For those unfamiliar with networking technology, networked information, and related issues the following brief list of suggested readings is provided. This list is not intended to be a comprehensive bibliography, even of survey works. Also included is a list of related Office of Technology Assessment reports that will provide helpful background on copyright issues, standards, cryptography, and other subjects related to the topic of this paper.

### *Networking Technology, Networks, and the Internet*


Cerf, V. (1991). Networks. *Scientific American,* ***265(3),*** September 1991,42-51.

LaQuey, Tracy, with Ryer, J. C. (1993). The Internet *companion : a beginner's guide to global networking.* Reading, MA: Addison-Wesley.

Malamud, C. (1992). Stacks: *Interoperability in Today's Computer Networks.* Englewood Cliffs, NJ: Prentice-Hall.

Quarterman, J. S. (1990). *The matrix: computer networks and conferencing systems worldwide.* Bedford, MA: Digital Press.


### *Networked Information Services and Networked Information*


American Society for Information Science (1992). *Networking, Telecommunications and the Networked Information Revolution, May 28-30, 1992, Proceedings of the ASIS 1992 Mid-Year Meeting. N.* Gusack (Ed.), Silver Springs, MD: American Society for Information Science.

Krol, E. (1993). *The whole Internet : user's guide & catalog* (Corrected Edition). Sebastopol, CA: O'Reilly & Associates.

Lynch, C. A., & Preston, C. M. (1990). Internet Access To Information Resources. in *Annual Review Of information Science And Technology, Volume* ***25 (1990),*** M. Williams (Ed.), pp 263-312.


### *Surveys of Topics Related to this Paper*


U.S. Congress Office of Technology Assessment. (1987). *Defending Secrets, Sharing Data: New Locks and Keys for Electronic Information:* Washington DC: U.S. Government Printing Office.

U.S. Congress Office of Technology Assessment, (1986). *Intellectual Property Rights in an Age of Electronics and Iformation.* Washington, DC: U.S. Government Printing Office.

U.S. Congress Office of Technology Assessment. (1988). *Informing the Nation: Fedeal Iformation Dissemination in an Electronic Age.* Washington, DC: U.S. Governement Printing Office.

U.S. Congress Office of Technology Assessment. (1990). *Ctitical Connections: Communication future.* Washington, DC: U.S. Government Printing Office.

U.S. Congress Office of Technology Assessment. (1990). *Helping America Compete: The Role of Federal Scientific and Technical Information.* Washington, DC: U.S. Government Printing Office.

U.S. Congress Office of Technology Assessment. (1992). *Finding a* Balance: *Computer Software, Intellectual Property, and the Challenge of Technological Change.* Washington, DC: U.S. Government Printing Office.

U.S. Congress Office of Technology Assessment. (1992). Global *Standards: Building Blocks for the Future.* Washington, DC: U.S. Government Printing Office.

## References


Alberti, B., Anklesaria, F., Linder, P., MaCahill, M., & Torrey, D. (1992). The Internet *Gopher protocol: a distributed document search and retrieval protocol.* Univeristy of Minnesota.

Arms, W. Y. (1992). The Design of the Mercury Electronic Library. *EDUCOM Review,* 27(6), 38-41.

Bagdikian, B. H. (1992). *The Media Monopoly* (4th cd.). Boston: Beacon Press.

Baker, S. K., & Jackson, M. E. (1992). *Maximizing Access, Minimizing Cost: A First Step Toward the Information Access Future.* Association of Research Libraries, Comittee on Access to Information Resources.

Balenson, D. (1993). *RFC 1423: Privacy Enhancement for Internet Electronic Mail: Part III: Algorithms, Modes and Identified*

Belkin, N. J., & Croft, W. B. (1992). Information Filtering and Information Retrieval: Two Sides of the Same Coin? *Communications of the ACM, 35(12), 29-38.*

Berners-Lee, T. J., Cailliau, R., Groff, J.-F., & Pollermann, B. (1992). World-Wide Web: The Information Universe. *Electronic Networking: Research, Application and Policy, 2(l), 75-7.*

Blair, D. C., & Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document-retrieval system. *Communications of the ACM,* vol.28,(no.3), 289-99.

Bowers, F. J., & Shapiro, N. R. ***(1992).*** CD-ROM Standards: Essential for Progress. *CD-ROM Librarian, 7(8), 33-6.*

Brand, S. (1987). *The Media Lab. New* York: Viking.

Buckland, M. (1988). Bibliography, Library Records and the Redefinition of the Library Catalog. *Library Resources and Technical Services, 33(4), 299-311.*

Bull, G., Hill, I., Guyre, K., & Sigmon, T. (1991). Building an Academic Village: Virginia's Public Education Network. *Educational technology, 31(4), 30 (7* pages).

Burke, S. *(1991).* Los Angeles DA investigating Prodigy service. *PC Week, 8(18), 119 (2* pages).

Council on Library Resources (1990). *Communications in Supped of Science and Engineering: A Report to the National Science Foundation from the Council on Library Resources.* Council on Library Resources.

Council on Library Resources (1992). *Final Report on the Conference for Explorition of a National Engineering Information Service, June 14-19, 1992, Palm Coast, Florida.* Cornell Information Technologies and Media Services Printing.

de Sola Pool, 1. (1983). *Technologies of Freedom.* Cambridge, MA: The Belknap Press of Harvard University Press.

Deerwester, S., Dumais, S. T., Furnas, G. W., & Landauer, T. K. (1990). Indexing by Latent semantic Analysis. *Journal of the American Society for Iformation Science, 41(6), 391-407.*

Dillon, M. (1993). *Assessing Information on the Internet: Towards Providing Library Services for Computer Mediated Communication.* OCLC,Inc.

Emtage, A., & Deutsch, P. (1991). archie-an electronic directory service for the Internet. *Proceedings of the Winter 1992 USENIX Conference, 93-110.*

Foltz, P. W. (1990). Using Latent Sematic Indexing for Information Filtering. S/GO/S *Bulletin, 11(2-3), 40-7.*

Garfield, E. (1979). *Citation Indexing - Its theory and application in Science, Technology and Humanities. New* York: Wiley.

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM, 35(12), 61-70.*

Grunder, T. (1992). Whose Internet is it Anyway? - A Challenge. *Online Magazine,* 16(4), 6 (3 pages).

Grycz, C. (cd) (1992). *Serials Review Special Issue on New Models in Serials Pricing; 18(1-2)*

Harman, D. (1992). The DARPA TIPSTER Project (Information Retrieval). *S/G/R Forum, 26(2), 26-8.*

Harman, D. (Ed.). (1993a). *The First Text Retrieval Conference (TREC-1)* (NIST Special Publication 500-207 cd.). Gaithersburg, MD: National Institute of Standards and Technology.

Harman, D. (1993 b). Overview of the First TREC Conference. In R. Korfhage, E. Rasmussen, & P. Willett (Ed.), *Sixteenth Annual International ACM S/G/R Conference on Research and Development in Information Retrieval,* (pp. 36- 48). Pittsburgh, PA, June 27- July 1 1993: ACM Press.

Jacso, P. (1992a). Author Agrees On Inspec Currency. *Database-The* Magazine *Of* Database *Reference And Review, 15(6), 55-55.*

Jacso, P. (1992 b). What Is In A(N) (Up)Date - Currency Test Searching Of Databases. *Database-The Magazine Of Database Reference And Review, 15(3), 28-33.*

Jacso, P. (1993a). A Proposal For Database Nutrition And Ingredient Labeling. *Database-The Magazine Of Database Reference And Review, 16(1), 7-9.*

Jacso, P. (1993 b). Searching For Skeletons In The Database Cupboard .1. Errors Of Omission. *Database-The Magazine Of Database Reference And Review, 16(1), 38-49.*

Jacso, P. (1993 c). Searching For Skeletons In The Database Cupboard .2. Errors Of Commission. *Database-The Magazine Of Database Reference And Review, 16(2), 30+.*

Kahle, B., Morris, H., Davis, F., & Tiene, K. (1992a). Wide Area Information Servers: An Executive Information System for Unstructured Files. *Electronic Networking: Research, Applications and Policy, 2(l), 59-68.*

Kahle, B., Morris, H., Goldman, J., Erickson, T., & Curran, J. (1992 b). Interfaces to Wide Area Information Servers. In N. Gusack (Ed.), *Networking, Telecommunications, and the Networked Information Revolution: ASIS 1992 Mid-year meeting,* Albuquerque, New Mexico: American Society for Information Science.

Kaliski, B. S. (1992). *RFC 1319: MD2 Message Digest Algorithm.*

Kaliski, B. S. (1993). RFC 1424: Privacy *Enhancement for Internet Electronic Mail:* Part *IV: Key Cetification and Related Services.*

Katz, A. R., & Cohen, D. (1992). *RFC* 1374: *File format for the Exchange of Images in the Internet.*

Kent, S. T. (1993). *RFC 1422: Privacy Enhancement for Internet Electronic Mail: Part II: Cetificate-Based Key Management.*

Keyhani, A. (1993). The Online Journal of Current Clinical Trials: An Innovation in Electronic Journal Publishing. *Database, 16(1), 14-15, 17-20, 22-3.*

Lederberg, J., & Uncapher, K. (1989). Towards a *National Colaboratoty: Report of an Invitational Workshop at the Rockefeller University, March 13-151989.* The Rockefeller University.

Lesk, M. (1991). The CORE electronic chemistry library. *SIGIR Forum,* 93-112.

Library of Congress (1991a). *Discussion Paper 49: Dictionary of Data Elements of Online Iformation Resources.* Washington, DC; Library of Congress.

Library of Congress (1991 b). *Discussion Paper 54: Providing Access to Online Information Resources.* Washington, DC; Library of Congress.

Library of Congress (1993). *Discussion Paper 69: Accommodating Online Systems and Services in USMARC.* Washington, DC; Library of Congress.

Library of Congress Network Advisory Committee (1992). Proceedings of the *Joint Meeting of the Library of Congress Network Advisory Committee and the Chief Officers of State Library Agencies Network* Planning Paper 23). Washington, DC; Library Of Congress.

Linn, J. (1993). RFC 1421: Privacy *Enhancement for Internet Electronic Mail: Part 1: Message Encryption and Authentication Procedures.*

Loken, S. C. (1990). *Report of the APS Task Force on Electronic Iformation Systems.* Lawrence Berkeley Laboratory.

Love, J. P. (1993). The Ownership and Control of the US Securities and Exchange Commission's EDGAR System. *Government Publications Review, 20(1), 61-71.*

Lucier, R. E. (1990). Knowledge management: refining roles in scientific communication. *EDUCOM Review, 25,(3), 21-7.*

Lucier, R. E. (1992). Towards a knowledge management environment: a strategic framework. *EDUCOM Review, 27,(6), 24-31.*

Lynch, C. A. (1989). Library Automation and the National Research Network. *EDUCOM Review, 24,(3), 21-6.*

Lynch, C. A. (1991a). Visions of Electronic Libraries. In The *Bowker Annual Library and Book Trade Almanac, 36th Edition 1991* (pp. 75-82). New Providence, NJ: R. R. Bowker.

Lynch, C. A. (1991 b). The Z39.50 Information Retrieval Protocol: an Overview and Status Report. *Computer Communication Review, 21,(1), 58-70.*

Lynch, C. A. (1992). The Next Generation Of Public Access Information Retrieval Systems For Research Libraries - Lessons From 10 Years Of The Melvyl System. *Information Technology And Libraries, 11(4), 405-415.*

Lynch, C. A., Hinnebusch, M., Peters, P. E., & McCallum, S. (1990). Information Retrieval as a Network Application. *Library Hi Tech, 8,(4), 57-72.*

Lynch, C. A., & Preston, C. M. (1990). Internet Access To Information Resources. *Annual Review Of Information Science And Technology, 25, 263-312.*

Lynch, C. A., & Preston, C. M. (1992). Describing and Classifying Networked Information Resources. *Electronic Networking: Research, Applications and Policy, 2,(l), 13-23.*

Markoff, J. (1993, v 142 (Thursday, March 4, 1993)). Turning the desktop PC into a talk radio medium. (Carl Malamud's Internet Talk Radio Program). New York *Times.* p. Al(N), Al(L), COI 1.

Marshak, D. (1990). Filters: separating the wheat from the chaff. Patty Seybold's Office *Computing Report,* 13(11), 1-16.

Mayer, M. (1990). Scanning the Future (marketing use of supermarket scanning information). Forbes, 146(8), 114 (4 pages).

Mitchell, M., & Saunders, L. M. (1991). The virtual library: an agenda for the 1990s. *Computers in Libraries, 11(4), 8-11.*

Mitchell, W. J. (1992). *The Reconfigured Eye: Visual Truth in the Post Photographic Em.* Cambridge, MA: MIT Press.

Nelson, T. H. (1988). Managing Immense Storage: Project Xanadu provides a model for the possible future of mass storage. Byte, 13(1), 225-238.

Olson, M. V. (1993). The Human Genome Project. *Proceedings of the National Academy of Sciences of the United States of America, 90(10), 4388-4344.*

Palca, J. (1991). New Journal will Publish without Paper ('The Online Journal of Current Clinical Trials' Starts as an On-line Peer-reviewed Journal). *Science,* 253(5027), 1480.

Pertiz, B. C. (1992). On the objectives of Citation Analysis - problems of Theory and Method. *Journal of the American Society for nfonmation Science, 46(6), 448-451.*

Peters, P. E. (1992a). The Coalition for Networked Information as a Model. In J. W. Y. Smith (Ed.), ˆ(Eds.), *Networking and the Future of Libraries: Proceedings of the UK Office for Networding Conference,* (pp. 165-170). Bath, UK: Meckler.

Peters, P. E. (1992 b). Making the Market for Networked Information: An Introduction to a Proposed Program for Licensing Electronic Uses. *Setials Review,* 78(1-2), 19-24.

Piatetsky-Shapiro, *G.,* & Frawley, W. J. (Ed.). (1991). *Knowledge Discovery in Databases.* Cambridge, MA: MIT Press.

Rivest, R. L. (1992a). *RFC 1227: MD5 Message Digest Algorithm.*

Rivest, R. L. (1992 b). *RFC 1320: MD4 Message Digest Algorithm.*

Roche, M. M. (1993). *ARL/RLG Interlibrary Loan Cost Study˙A Joint Effort by the Association of Research Libraries and the Research Libraries Group.* Association of Research Libraries.

Salton, G. (1988). *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Iformation by Computer.* Reading, MA: Addison-Wesley.

Saylor, J. M. (1992). NEEDS (The National Engineering Education Delivery System): If We Build It (According to Standards) They Will Come! In *Networks, Telecommunications and the Networked Information Resources Revloution. American*

Society for Information Science, *Mid-year Proceedings.,* (pp. 273+). Albuquerque, NM: American Society for Information Science.

Schwartz, M. F. *(1989).* The networked resource discovery project. *Iformation Processing 89. Proceedings of the IFIP IIth World Computer Congress, 827-32.*

Schwartz, M. F., Emtage, A., Kahle, B., & Neuman, B. C. (1992). A Comparison of Internet Resource Discovery Approaches. *Computing Systems, 5(4), 461-93.*

Schwartz, M. F., Hardy, D. R., Heinzman, W. K., & Hirschowitz, G. C. (1991). Supporting resource discovery among public Internet archives using a spectrum of information quality. *11th International Conference on Distributed Computing Systems* (Cat. *N0.91CH2996-7), 82-9.*

Simmonds, C. (1993). Searching Internet Archive Sites with archie: Why, What, Where and How. *Online, 17(2), 50 (5* pages).

Stonebraker, M. (1992). An Overview of the Sequoia 2000 Project. In *Digest of Papers. COMPCON Spring 1992. Thirty-Seventh IEEE Computer Society International Conference,* (pp. 383-8). San Francisco, CA: IEEE Computer Society Press.

Stonebraker, M., & Kemmitz, G. (1991). The POSTGRES Next - Generation Database Management System. *Communications of the ACM, 34(10), 78-92.*

Stout, C. (1992). *TENET: Texas Education Network* Texas Education Agency, Austin.

Strangelove, M. (1993). *Directory of Electronic Journals, Newsletters, and Scholarly Discussion Lists.* Association of Research Libraries.

Tenopir, C., & Ro, J. S. (1990). *Full Text Databases.* Westport, CT: Greenwood Press.

U S. Congress Office of Technology Assessment. (1987). *Defending Secrets, Sharing Data: New Locks and Keys for Electronic Information.*

U.S. Congress Office of Technology Assessment. (1986). *Intellectual Property Rights in an Age of Electronics and Information.* Washington, DC: U.S. Government Printing Office.

U.S. Congress Office of Technology Assessment. (1990). *Helping America Compete: The Role of Federal Scientific and Technical Information.* Washington, DC: U.S. Government Printing Office.

U.S. Congress Office of Technology Assessment. (1992). Global *Standards: Building Blocks for the Future.* Washington, DC: U.S. Government Printing Office.

Vinge, V. (1992). *A Fire upon the Deep. New* York: TOR.

Watkins, B. (1991). USC to Put Full Text of The Chronicle on network. *Chronicle of Higher Education, 37(27), A20.*

Watkins, B. T. (1992). Free-Net helps Case Western fulfill its community-service mission. Chronic/e *of Higher Education,* 38(34), A21 (2 pages).

White, H. S. (1989). The vaule-added process of librarianship. *Library Journal, 114(1), 62 (2* pages).

Wiggens,R. (1993). Gopher. *Public Access Systems Review. 4(l).*

# 1997 Traffic Crashes, Injuries and Fatalities - Preliminary Report

by

Ezio C. Cerrelli

**National Highway Traffic Safety Administration**

# Introduction

This report contains preliminary estimates of motor vehicle crashes and resulting injuries and fatalities.

The crash estimates for 1997 are based on all cases reported to the General Estimate System (GES) for the first nine months of 1997 and the cases reported for October, November, and December of 1996. The Primary Sampling Unit (PSU) weights were changed in 1997 to reflect the demographic changes that have occurred since the original survey design. The weights for the 1996 cases have been modified accordingly to insure a more realistic comparison between 1996 and 1997.

The GES obtains its data from a nationally representative probability sample selected from all police-reported crashes which occur annually. Although the GES file contains fatal, injury, and property damage cases, for the purpose of this report only injury and property damage cases are utilized.

The fatality estimates for 1997 are based on all cases reported to the Fatal Analysis Reporting System (FARS) as of February 1998, and on the preliminary reporting of the total number of traffic fatalities in 1997 by each state. The number of cases on file as of February 1998 are estimated to represent about 85 percent of the final total for the year. The combination of the two sources has been used for a number of years to produce fairly accurate estimates of detailed traffic fatality statistics months before the actual reporting of all cases is completed.

FARS and GES are sponsored and managed by the National Center for Statistics and Analysis (NCSA), an office of the National Highway Traffic Safety Administration (NHTSA).

The estimates of fatalities for 1997 represent an extrapolation of the data presently available in the FARS file. Extrapolation factors have been established for each of the twelve months using the expected final monthly fatality counts and the corresponding counts in the FARS file. Past experience has shown that most large estimates, counts of 10,000 or more, tend to be within one percent of the final figure. Smaller estimates are subject to relatively larger estimation errors.

For some data elements, e.g., Highway System, Land Use, Speed Limit and Vehicle Type, the reporting is not quite as complete at this early date and the file contains a large number of cases with unknown values. Additional adjustments were required for these data elements, in order to make proper comparisons with previous years' data.

The GES file is based on a sample survey; therefore, all estimates based on this file are subject to

sampling and nonsampling errors. Sampling errors represent the probable differences between the results obtained from the sample survey and those results that would be obtained from a census of the population. Sampling variability should be taken into account when interpreting year-to-year differences. The technical note (DOT HS 807 796), which is available from NHTSA, provides a complete description of the GES sample design, and estimation of sampling errors.

The GES does not provide estimates for individual states or NHTSA regions. Only FARS estimates of total fatalities in each state and region are presented in the report.

The report presents a series of estimates on crashes, injuries, and fatalities for 1997. The report contains three major sections. The tables in the first section reflect a comparison of selected 1997 estimates to the corresponding counts reported for 1996. The second section consists of a group of charts reflecting various trends in crashes, injuries, and fatalities. The final section is designed to provide a large and detailed set of 1997 estimates derived from the GES and FARS files.

The fatality trend, presented in the report, is based on monthly data from January 1975 through December 1997. These fatality counts have been combined with the travel estimates, over the same period of time, to compute the fatality rate trend. Vehicle Miles of Travel (VMT) data used to calculate the fatality rate were provided by the Federal Highway Administration (FHWA).

The year to year comparisons rely on the redistribution of unknown data for 1997. These results, therefore, may change once the completed version of the 1997 GES and FARS files become available.

# Principal Findings

## Crashes, Injuries, and Fatalities

The number of police-reported traffic crashes is estimated at 6,753,500 for 1997. This represents a small (1.3%) decrease over 1996. The number of injuries that occurred in these crashes is estimated at 3,450,000, which is 1.7 percent lower relative to 1996.

An estimated 42,000 people lost their lives in traffic crashes during 1997. This represents a 0.2 percent decrease from the 42,065 fatalities reported for 1996. The 1997 total is 1.4 percent lower than in 1983 and 17.8 percent lower than the count in 1980. The years 1980 and 1983 are used for comparisons as these years represent the previous high and low points in the fatality trend during the first fifteen years of the FARS file (1975-1990).

Based on the FHWA's estimated increase of 2.0 percent in vehicle miles of travel, the fatality rate for 1997 is estimated at 1.7 fatalities per 100 million vehicle miles of travel, the same as in 1996 but much lower than the 2.6 for 1983, and 3.3 for 1980.

The overall fatality trend (Page 25) displays a cyclical pattern. Since 1975 two cycles have been completed and 1997 may be the peak of the third cycle. The period of the cycle is approximately nine years with lows in 1975, 1983, and 1992, and highs in 1979, 1988 and, possibly, in 1997. Both highs and lows have been decreasing their values with each cycle. The fatality rate per 100 million vehicle miles of travel, based on the ratio of the monthly trend values of fatalities and travel, has continued to

decline at a steady rate, from the 3.5 level at the beginning of 1975 to about 1.7 by the end of 1993. The fatality rate has remained at the 1.7 level through the end of 1997.

Between 1996 and 1997 the number of pedestrian fatalities decreased by 5.4 percent while pedalcyclist fatalities increased by about 4.6 percent. Fatalities in single vehicle crashes and angle collisions remained at the same level as in 1996. Head on crash and sideswipe crash fatalities increased by 3 percent and 8 percent respectively, while rear end crash fatalities decreased by 1 percent.

In 1997 the number of fatalities associated with the presence of alcohol decreased with respect to 1996. More specifically, the number of fatalities in crashes where the BAC level was at .1 or above decreased by 3.5 percent, and a larger decrease of 5.9 percent occurred in crashes where the BAC level was between .01 and .09. Fatalities increased by 2.5 percent in crashes where no alcohol was present.

## Location

The number of crashes occurring on roads with a posted speed limit below 55 mph decreased by about 2.4 percent, roads posted at 55 mph and above show an increase of about 2.8 percent.

Crash injuries decreased by 2.6 percent on all roads combined with a posted speed below 55 mph, and increased by about 1.2 percent on roads posted at 55 mph and higher speed limits. All roads with a posted speed limit under 55 mph had about a 2.4 percent average decrease in fatalities from 1996 while the increase was 1.7 percent on roads posted at 55 mph and higher speeds.

Fatalities increased by 3.8 percent over 1996 in rural areas and decreased by 5.7 percent in urban areas. Fatalities increased by 5 percent on the Interstate system and by 1.6 percent on the U.S. routes that are not part of the Interstate system. The combination of state and county roads experienced an average decrease of less than 1 percent. Fatalities on local and other roads decreased by 1.8 percent.

## Drivers

An estimated 12,088,470 drivers were involved in police-reported crashes in 1997, a decrease of about 1.1 percent from 1996. About 2,182,660 drivers suffered some type of injury and an additional 24,860 were fatally injured, a 2.3 percent decrease over 1996 for the number of driver injuries and a 1.3 percent increase for the number of driver fatalities.

The number of drivers involved in fatal crashes in 1997 was 57,000, about the same number as in 1996. The proportions of male and female drivers in fatal crashes remained almost the same as in 1996. Between 1980 and 1997, the number of male drivers in fatal crashes has decreased by 20 percent while female drivers involved in fatal crashes experienced an increase of 28 percent.

Between 1996 and 1997, the number of drivers of motorcycles involved in fatal crashes decreased by 3 percent, while the number of drivers of passenger cars also decreased by 2.4 percent. During the same period the number of drivers involved in fatal crashes increased by almost 2.7 percent for light trucks and vans, while medium and heavy trucks combined experienced no change.

The number of driver fatalities increased by 1.3 percent overall, with males showing a 1 percent decrease and females a 2 percent increase. For drivers over 65 years of age, the number of fatalities

increased by 6 percent, while driver fatalities in the younger groups showed no change.

## Passengers

The number of passengers injured in traffic crashes is estimated at 1,125,890, a very small increase over 1996. The number of passengers that died in motor vehicle crashes is estimated at 11,040, 1 percent fewer than in 1996. The number of occupant fatalities remained the same as in 1996. The number of occupant fatalities decreased by 1 percent in passenger cars, increased by 4 percent in light trucks, and increased by 16 percent for medium and heavy trucks. Motorcycle driver and passenger fatalities, combined, also decreased by 3 percent.

## Nonoccupants

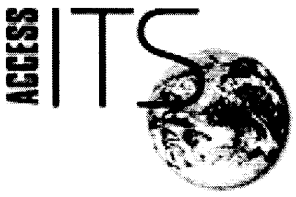
About 76,550 pedestrians were injured in crashes during 1997, a 6 percent decrease from the previous year. The number of pedestrians killed in 1997 was 5,300, about 5 percent lower than in 1996. Since 1980, pedestrian fatalities have decreased by 32 percent, the decrease being similar for both males and females. The number of pedalcyclists injured in 1997 is estimated at 64,900, a 7 percent decrease from 1996. The estimated 800 pedalcyclist fatalities for 1997 is 5 percent higher than in 1996. Since 1980, pedalcyclist fatalities have decreased by 17 percent.

For copies of the full report, please call (202) 366-4198 or toll free, 1-800-934-8517. Questions regarding the report may be directed to Ezio Cerrelli at (202) 366-5358.

Return to NCSA Reports and Research Notes Page

# Other Resources

## ITS Fair Information and Privacy Principles




This page was last updated on **Thursday, January 16, 1997**



December 13, 1994 Privacy Task Group Draft
Intelligent Transportation Systems Fair Information and Privacy Principles

These fair information and privacy principles were prepared in recognition of the importance of protecting individual privacy in implementing Intelligent Transportation Systems. They have been reviewed by the Legal Issues Committee of the Intelligent Transportation Society of America and are transmitted for adoption in draft final form by the Coordinating Council and Board of Directors.

The Privacy Task Group of the Legal Issues Committee will present these "draft final" principles for review and comment to organizations and groups interested in privacy and ITS outside of ITS AMERICA during 1995. They will then be submitted for final adoption to the ITS AMERICA Legal Issues Committee, Coordinating Council, and Board of Directors.

The principles represent values and are designed to be flexible and durable to accommodate a broad scope of technological, social, and cultural change. ITS AMERICA may, however, need to revisit them periodically to assure their applicability and effectiveness.

These principles are advisory, intended to educate and guide transportation professionals, policy makers, and the public as they develop fair information and privacy guidelines for specific intelligent transportation projects. Initiators of ITS projects are urged to publish the fair information privacy principles that they intend to follow. Parties to ITS projects are urged to include enforcible provisions for safeguarding privacy in their contracts and agreements.

1. INDIVIDUAL CENTERED. Intelligent Transportation Systems (ITS) must recognize and respect the individual s interests in privacy and information use.

ITS systems create value for both individuals and society as a whole. Central to the ITS vision is the creation of ITS systems that will fulfill our national goals. The primary focus of information use is to improve travelers' safety and security, reduce travel times, enhance individuals ability to deal

with highway disruptions and improve air quality. Traveler information is collected from many sources, some from the infrastructure and some from vehicles, while other information may come from the transactions like electronic toll collection that involve interaction between the infrastructure and vehicle. That information may have value in both ITS and non-ITS applications. The individual's expectation of privacy must be respected. This requires disclosure and the opportunity for individuals to express choice.

2. VISIBLE. Intelligent transportation information systems will be built in a manner visible to individuals.

ITS may create data on individuals. Individuals should have a means of discovering how the data flows operate. Visible means to disclose to the public the type of data collected, how it is collected, what its uses are, and how it will be distributed.

The concept of visibility is one of central concern to the public, and consequently this principle requires assigning responsibility for disclosure.

3. COMPLY. Intelligent Transportation Systems will comply with state and federal laws governing privacy and information use.

4. SECURE. Intelligent Transportation Systems will be secure.

ITS data bases may contain information on where travelers go, the routes they use, and when they travel, and therefore must be secure. All ITS information systems will make use of data security technology and audit procedures appropriate to the sensitivity of the information.

5. LAW ENFORCEMENT. Intelligent Transportation Systems will have an appropriate role in enhancing travelers' safety and security interests, but absent consent, government authority, or appropriate legal process, information identifying individuals will not be disclosed to law enforcement.

ITS has the potential to make it possible for traffic management agencies to know where individuals travel, what routes they take, and travel duration. Therefore, ITS can increase the efficiency of traffic law enforcement by providing aggregate information necessary to target resources. States may legislate conditions under which ITS information will be made available. Absent government authority, however, ITS systems should not be used as a surveillance means for enforcing traffic laws. Although individuals are concerned about public safety, persons who voluntarily participate in ITS programs or purchase ITS products have a reasonable expectation that they will not be "ambushed" by information they are providing.

6. RELEVANT. Intelligent Transportation Systems will only collect personal information that is relevant for ITS purposes.

ITS, respectful of the individual's interest in privacy, will only collect

information that contain individual identifiers which are needed for the ITS service functions.

Futhermore, ITS information systems will include protocols that call for the purging of individual identifier information that is no longer needed to meet ITS needs.

7. SECONDARY USE. Intelligent Transportation Systems information coupled with appropriate individual privacy protection may be used for non-ITS applications.

American consumers want information used to create economic choice and value, but also want their interest in privacy preserved. ITS information is predictive of the types of goods and services that interest consumers, for example the right location for stores, hospitals, and other facilities. However, that same information might also be used to disadvantage and harm a consumer. Therefore, the following practices should be followed.

o ITS information absent personal identifiers may be used for ITS and other purposes.

o Other unrelated uses of ITS information with personal identifiars may be permissible if individuals receive effective disclosure and have a user friendly means of opting out.

o Data collectors will only provide personal information to private organizations that agree to abide by these privacy principles.

8. FOIA. Federal and State Freedom of Information Act (FOIA) obligations require disclosure of information from government maintained databases. Database arrangements should balance the individual's interest in privacy and the public's right to know.

In determining whether to disclose ITS information, governments should, where possible, balance the individual's right to privacy against the preservation of the basic purpose of the Freedom of Information laws to open agency action to the light of public scrutiny. ITS travelers should be presumed to have reasonable expectations of privacy for personal identifying information. Pursuant to the individual's interest in privacy, the public/private frameworks of organizations collecting data should be structured to resolve problems of access created by FOIA.

Return to

Comments?
Questions?
Contact
editor
@itsa.org

©1999 ITS America.
Globe image ©1996 Photodisc Inc.
Legal notices

**ITS America**
400 Virginia Avenue, S.W., Suite 800
Washington, D.C. 20024–2730

202/484.4847 · Fax: 202/484.3483
**www.itsa.org**

**Drivers Privacy Protection Act**

**18 U.S.C. § 2721 et. seq.**

(Public Law 103-322)

**Section 2721. Prohibition on release and use of certain personal information from State motor vehicle records**

(a) In General -- Except as provided in subsection (b), a State department of motorvehicles, and any officer, employee, or contractor, thereof, shall not knowingly disclose orotherwise make available to any person or entity personal information about any individualobtained by the department in connection with a motor vehicle record.

(b) Permissible Uses -- Personal information referred to in subsection (a) shall bedisclosed for use in connection with matters of motor vehicle or driver safety and theft, motorvehicle emissions, motor vehicle product alterations, recalls, or advisories, performancemonitoring of motor vehicles and dealers by motor vehicle manufacturers, and removal of non-owner records from the original owner records of motor vehicle manufacturers to carry out thepurposes of the *Automobile Information Disclosure Act*, the *Motor Vehicle Information and CostSaving Act*, the *National Traffic and Motor Vehicle Safety Act of 1966*, the *Anti-Car Theft Act of1992*, and the *Clean Air Act*, and may be disclosed as follows:

(1) For use by any government agency, including any court or law enforcementagency, in carrying out its functions, or any private person or entity acting on behalf of aFederal, State, or local agency in carrying out its functions.

(2) For use in connection with matters of motor vehicle or driver safety and theft;motor vehicle emissions; motor vehicle product alterations, recalls, or advisories;performance monitoring of motor vehicles, motor vehicle parts and dealers; motorvehicle market research activities, including survey research; and removal of non-ownerrecords from the original owner records of motor vehicle manufacturers.

(3) For use in the normal course of business by a legitimate business or itsagents, employees, or contractors, but only --

(A) to verify the accuracy of personal information submitted by theindividual to the business or its agents, employees, or contractors; and

(B) if such information as so submitted is not correct or is no longercorrect, to obtain the correct information, but only for the purposes of preventingfraud by, pursuing legal remedies against, or recovering on a debt or securityinterest against, the individual.

(4) For use in connection with any civil, criminal, administrative, or arbitralproceeding in any

Federal, State, or local court or agency or before any self-regulatorybody, including the service of process, investigation in anticipation of litigation, and theexecution or enforcement of judgments and orders, or pursuant to an order of a Federal,State, or local court.

(5) For use in research activities, and for use in producing statistical reports, solong as the personal information is not published, redisclosed, or used to contactindividuals.

(6) For use by any insurer or insurance support organization, or by a self-insuredentity, or its agents, employees, or contractors, in connection with claims investigationactivities, anti-fraud activities, rating or underwriting.

(7) For use in providing notice to the owners of towed or impounded vehicles.

(8) For use by any licensed private investigative agency or licensed securityservice for any purpose permitted under this subsection.

(9) For use by an employer or its agents or insurer to obtain or verify informationrelating to a holder of a commercial driver's license that is required under the*Commercial Motor Vehicle Safety Act of 1986* (49 U.S.C. App. 2710 et seq.).

(10) For use in connection with the operation of private toll transportationfacilities.

(11) For any other use in response to requests for individual motor vehiclerecords if the motor vehicle department has provided in a clear and conspicuous manneron forms for issuance or renewal of operator's permits, titles, registrations, oridentification cards, notice that personal information collected by the department may bedisclosed to any business or person, and has provided in a clear and conspicuousmanner on such forms an opportunity to prohibit such disclosures.

(12) For bulk distribution for surveys, marketing or solicitations if the motorvehicle department has implemented methods and procedures to ensure that --

(A) individuals are provided an opportunity, in a clear and conspicuousmanner, to prohibit such uses; and

(B) the information will be used, rented, or sold solely for bulkdistribution for surveys, marketing, and solicitations, and that surveys,marketing, and solicitations will not be directed at those individuals who haverequested in a timely fashion that they not be directed at them.

(13) For use by any requester, if the requester demonstrates it has obtained thewritten consent of the individual to whom the information pertains.

(14) For any other use specifically authorized under the law of the State thatholds the record, if such use is related to the operation of a motor vehicle or publicsafety.

(c) Resale or Redisclosure -- An authorized recipient of personal information (except arecipient under subsection (b)(11) or (12)) may resell or redisclose the information only for a usepermitted under subsection (b) (but not for uses under subsection (b)(11) or (12)). An authorizedrecipient under subsection (b)(11) may resell or redisclose personal information for any purpose. An authorized recipient under subsection (b)(12) may resell or redisclose personal informationpursuant to

subsection (b)(12). Any authorized recipient (except a recipient under subsection(b)(11)) that resells or rediscloses personal information covered by this title must keep for aperiod of 5 years records identifying each person or entity that receives information and thepermitted purpose for which the information will be used and must make such records availableto the motor vehicle department upon request.

(d) Waiver Procedures -- A State motor vehicle department may establish and carry outprocedures under which the department or its agents, upon receiving a request for personalinformation that does not fall within one of the exceptions in subsection (b), may mail a copy ofthe request to the individual about whom the information was requested, informing suchindividual of the request, together with a statement to the effect that the information will not bereleased unless the individual waives such individual's right to privacy under this section.

**Section 2722. Additional unlawful acts**

(a) Procurement for Unlawful Purpose -- It shall be unlawful for any person knowingly toobtain or disclose personal information, from a motor vehicle record, for any use not permittedunder section 2721(b) of this title.

(b) False Representation -- It shall be unlawful for any person to make falserepresentation to obtain any personal information from an individual's motor vehicle record.

**Section 2723. Penalties**

(a) Criminal Fine -- A person who knowingly violates this chapter shall be fined underthis title.

(b) Violations by State Department of Motor Vehicles -- Any State department of motorvehicles that has a policy or practice of substantial noncompliance with this chapter shall besubject to a civil penalty imposed by the Attorney General of not more than $5,000 a day foreach day of substantial noncompliance.

**Section 2724. Civil action**

(a) Cause of Action -- A person who knowingly obtains, discloses or uses personalinformation, from a motor vehicle record, for a purpose not permitted under this chapter shall beliable to the individual to whom the information pertains, who may bring a civil action in a UnitedStates district court.

(b) Remedies -- The court may award --

(1) actual damages, but not less than liquidated damages in the amount of$2,500;

(2) punitive damages upon proof of willful or reckless disregard of the law;

(3) reasonable attorneys' fees and other litigation costs reasonably incurred; and

(4) such other preliminary and equitable relief as the court determines to beappropriate.

**Section 2725. Definitions**

In this chapter --

(1) "motor vehicle record" means any record that pertains to a motor vehicleoperator's permit, motor vehicle title, motor vehicle registration, or identification cardissued by a department of motor vehicles;

(2) "person" means an individual, organization or entity, but does not include aState or agency thereof; and

(3) "personal information" means information that identifies an individual,including an individual's photograph, social security number, driver identification number,name, address (but not the 5-digit zip code), telephone number, and medical or disabilityinformation, but does not include information on vehicular accidents, driving violations,and driver's status.

# Federal Advisory Committee Act

## 5 U.S.C. App. 1

Public Law 92-463

An act to authorize the establishment of a system governing the creation and operation of advisory committees in the executive branch of the Federal Government and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled, That:

### Section 2. Findings and Purposes

(a) The Congress finds that there are numerous committees, boards, commissions, councils, and similar groups which have been established to advise officers and agencies in the executive branch of the Federal Government and that they are frequently a useful and beneficial means of furnishing expert advice, ideas, and diverse opinions to the Federal Government.

(b) The Congress finds and declares that --

(1) the need for many existing advisory committees has not been adequately reviewed;

(2) new advisory committees should be established only when they are determined to be essential and their number should be kept to the minimum necessary;

(3) advisory committees should be terminated when they are no longer carrying out the purposes for which they were established;

(4) standards and uniform procedures should govern the establishment, operation, administration, and duration of advisory committees;

(5) the Congress and the public should be kept informed with respect to the number, purpose, membership, activities, and cost of advisory committees; and

(6) the function of advisory committees should be advisory only, and that all matters under their consideration should be determined, in accordance with law, by the official, agency, or officer involved.

### Section 3. Definitions

For the purpose of this Act --

(1) The term "Administrator" means the Administrator of the General Services Administration.

(2) The term "advisory committee" means any committee, board, commission, council, conference, panel, task force, or other similar group, or any subcommittee or other sub-group thereof (hereafter in this paragraph referred to as "committee"), which is --

(A) established by statute or organization plan, or

(B) established or utilized by the President; or

(C) established or utilized by one or more agencies;

in the interest of obtaining advise and recommendations for the President or one or more agencies or offices of the Federal Government, except that such term excludes

(i) any committee that is composed wholly of full-time, or permanent part-time, officers or employees of the Federal Government, and

(ii) any committee that is created by the National Academy of Sciences of the National Academy of Public Administration.

**Section 4. Applicability**

(a) The provisions of this Act or of any rule, order, or regulation promulgated under the Act shall apply to each advisory committee except to the extent that any Act of Congress establishing any such advisory committee specifically provides otherwise.

(b) Nothing in this Act shall be construed to apply to any advisory committee established or utilized by --

(1) the Central Intelligence Agency; or

(2) the Federal Reserve System.

(c) Nothing in this Act shall be construed to apply to any local civic group whose primary function is that of rendering a public service with respect to a Federal program, or any State of local committee, council, board, commission, or similar group established to advise or make recommendations to State of local officials or agencies.

**Section 5. Responsibilities of Congressional Committees**

(a) In the exercise of its legislative review functions, each standing committee of the Senate and the House of Representatives, shall make a continuing review of the activities of each advisory committee under its jurisdiction to determine whether such advisory committee should be abolished or merged with any other advisory committee, whether the responsibilities of each advisory committee should be revised, and whether such advisory committee performs a necessary function not already being performed. Each such standing committee shall take appropriate action to obtain the enactment of legislation necessary to carry out the purpose of this subsection.

(b) In considering legislation establishing, or authorizing the establishment of any advisory committee, each standing committee of the Senate and the House of Representatives, shall determine, and report such determination to the Senate or to the House of Representatives, as the case may be, whether the functions of the proposed advisory committee are being or could be performed by one or more agencies or by an advisory committee already in existence, or by enlarging the mandate of an existing advisory committee. Any such legislation shall --

(1) contain a clearly defined purpose for the advisory committee;

(2) require the membership of the advisory committee to be fairly balanced in terms of the points of view represented and the functions to be performed by the advisory committee;

(3) contain appropriate provisions to assure that the advice and recommendations of the advisory committee will not be inappropriately influenced by the appointing authority or by any special interest, but will instead be the result of the advisory committee's independent judgment;

(4) contain provisions dealing with authorization of appropriations, the date for submission of reports (if any), the duration of the advisory committee, and the publication of reports and other materials, to the extent that the standing committee determines the provisions of section 10 of this Act to be inadequate; and

(5) contain provisions which will assure that the advisory committee will have adequate staff (either supplied by an agency or employed by it), will be provided adequate quarters, and will have funds available to meet its other necessary expenses.

(c) To the extent they are applicable, the guidelines set out in subsection (b) of this section shall be followed by the President, agency heads, or other Federal officials in creating an advisory committee.

**Section 6. Responsibilities of the President**

(a) The President may delegate responsibility for evaluating and taking action, where appropriate, with respect to all public recommendations made to him by Presidential advisory committees.

(b) Within one year after a Presidential advisory committee has submitted a public report to the President, the President or his delegate shall make a report to the Congress stating either his proposals for action or his reasons for inaction, with respect to the recommendations contained in the public report.

(c) The President shall, not later than December 31 of each year, make an annual report to the Congress on the activities, status, and changes in the composition of advisory committees in existence during the proceeding fiscal year. The report shall contain the name of every advisory committee, the date of and the authority for its creation, its termination date or the date it is to make a report, its functions, a reference to the reports it has submitted, a statement of whether it is an ad hoc or continuing body, the dates of its meetings, the names and occupations of its current members, and the total estimated annual cost to the United States to fund, service, supply, and maintain such committee. Such report shall include a list of those advisory committees abolished by the President,

and in the case of advisory committees established by statute, a list of those advisory committees which the President recommends be abolished together with his reasons therefor. The President shall exclude from this report any information which, in his judgment, should be withheld for reasons of national security, and he shall include in such report a statement that such information is excluded.

**Section 7. Responsibilities of the Administrator of General Services, Committee Management Secretariat, establishment; review; recommendations to the President and Congress; agency cooperation; performance guidelines; uniform pay guidelines; travel expenses; expense recommendations**

(a) The Administrator shall establish and maintain within the General Services Administration a Committee Management Secretariat, which shall be responsible for all matters relating to advisory committees.

(b) The Administrator shall, immediately after October 6, 1972, institute a comprehensive review of the activities and responsibilities of each advisory committee to determine --

(1) whether such committee is carrying out its purpose;

(2) whether, consistent with the provisions of applicable statutes, the responsibilities assigned to it should be revised;

(3) whether it should be merged with other advisory committees; or

(4) whether it should be abolished.

The Administrator may from time to time request such information as he deems necessary to carry out his functions under this subsection. Upon the completion of the Administrator's review he shall make recommendations to the President and to either the agency head or the Congress with respect to actions he believes should be taken. Thereafter, the Administrator shall carry out a similar review annually. Agency heads shall cooperate with the Administrator in making the reviews required by this subsection.

(c) The Administrator shall prescribe administrative guidelines and management controls applicable to advisory committees, and, to the maximum extent feasible, provide advice, assistance, and guidance to advisory committees to improve their performance. In carrying out his functions under this subsection, the Administrator shall consider the recommendations of each agency head with respect to means of improving the performance of advisory committees whose duties are related to such agency.

(d)(1) The Administrator, after study and consultation with the Director of the Office of Personnel Management, shall establish guidelines with respect to uniform fair rates of pay for comparable services by members, staffs, and consultants for advisory committees in a manner which gives appropriate recognition to the responsibilities and qualifications required and other relevant factors. Such regulations shall provide that --

(A) no member of an advisory committee or of the staff of any advisory committee shall receive compensation at a rate in excess of the rate specified for GS-18 of the General Schedule under section 5332 of Title 5, United States Code;

(B) such members, while engaged in the performance of their duties away from their homes or regular places of business, may be allowed travel expenses, including per diem in lieu of subsistence, as authorized by section 5703 of Title 5, United States Code, for persons employed intermittently in the Government service; and

(C) such members --

(i) who are blind or deaf or who otherwise qualify as handicapped individuals (within the meaning of section 501 of the Rehabilitation Act of 1973 (29 U.S.C. §794), and

(ii) who do not otherwise qualify under section 3102 of Title 5, United States Code, by reason of being an employee of an agency (within the meaning of section 3102(a)(1) of such Title 5), may be provided services pursuant to section 3102 of such Title 5 while in performance of their advisory committee duties.

(2) Nothing in this subsection shall prevent --

(A) an individual who (without regard to his service with an advisory committee) is a full-time employee of the United States, or

(B) an individual who immediately before his service with an advisory committee was such an employee, from receiving compensation at the rate at which he otherwise would be compensated (or was compensated) as a full-time employee of the United States.

(e) The Administrator shall include in budget recommendations a summary of the amounts he deems necessary for the expenses of advisory committees, including the expenses for publication of reports where appropriate.

**Section 8. Responsibilities of Agency Heads; Advisory Committee Management Officer; designation**

(a) Each agency head shall establish uniform guidelines and management controls for advisory committees established by that agency, which shall be consistent with directives of the Administrator under section 7 and section 10. Each agency shall maintain systematic information on the nature, functions, and operations of each advisory committee within its jurisdiction.

(b) The head of each agency which has an advisory committee shall designate an Advisory Committee Management Officer who shall --

(1) exercise control and supervision over the establishment, procedure, and accomplishments of advisory committees established by the agency;

(2) assemble and maintain the reports, records, and other papers of any such committee during its existence; and

(3) carry out, on behalf of that agency, the provisions of section 552 of Title 5, United States Code,

with respect to such reports, records, and other papers.

**Section 9. Establishment and purpose of advisory committee; publication in Federal Register; charter; filing, contents, copy**

(a) No advisory committee shall be established unless such establishment is --

(1) specifically authorized by statute or by the President; or

(2) determined as a matter of formal record, by the head of the agency involved after consultation with the Administrator, with timely notice published in the Federal Register, to be in the public interest in connection with the performance of duties imposed on that agency by law.

(b) Unless otherwise specifically provided by statute or Presidential directive, advisory committees shall be utilized solely for advisory functions. Determinations of action to be taken and policy to be expressed with respect to matters upon which an advisory committee reports or makes recommendations shall be made solely by the President or an officer of the Federal Government.

(c) No advisory committee shall meet or take any action until an advisory committee charter has been filed with (1) the Administrator, in the case of Presidential advisory committees, or (2) with the head of the agency to whom any advisory committee reports and with the standing committee of the Senate and of the House of Representatives having legislative jurisdiction of such agency. Such charter shall contain the following information:

(A) the committee's official designation;

(B) the committee's objectives and the scope of its activity;

(C) the period of time necessary for the committee to carry out its purpose;

(D) the agency or official to whom the committee reports;

(E) the agency responsible for providing the necessary support for the committee;

(F) a description of the duties for which the committee is responsible, and, if such duties are not solely advisory, a specification of the authority for such function;

(G) the estimated annual operating costs in dollars and man-years for such committee;

(H) the estimated number and frequency of committee meetings;

(I) the committee's termination date, if less than two years from the date of the committee's establishment; and

(J) the date the charter is filed.

A copy of any such charter shall also be furnished to the Library of Congress.

**Section 10. Advisory committee procedures; meetings; notice, publication in Federal Register;**

**regulations; minutes; certification; annual report; Federal officer or employee, attendance**

(a)(1) Each advisory committee meeting shall be open to the public.

(2) Except when the President determines otherwise for reasons of national security, timely notice of each such meeting shall be published in the Federal Register, and the Administrator shall prescribe regulations to provide for other types of public notice to insure that all interested persons are notified of such meeting prior thereto.

(3) Interested persons shall be permitted to attend, appear before, or file statements with any advisory committee, subject to such reasonable rules or regulations as the Administrator may prescribe.

(b) Subject to section 552 of Title 5, United States Code, the records, reports, transcripts, minutes, appendixes, working papers, drafts, studies, agenda, or other documents which were made available to or prepared for or by each advisory committee shall be available for public inspection and copying at a single location in the offices of the advisory committee or the agency to which the advisory committee reports until the advisory committee ceases to exist.

(c) Detailed minutes of each meeting of each advisory committee shall be kept and shall contain a record of the persons present, a complete and accurate description of matters discussed and conclusions reached, and copies of all reports received, issued, or approved by the advisory committee. The accuracy of all minutes shall be certified to by the chairman of the advisory committee.

(d) Subsections (a)(1) and (a)(3) of this section shall not apply to any portion of an advisory committee meeting where the President, or the head of the agency to which the advisory committee reports, determines that such portion of such meeting may be closed to the public in accordance with subsection (c) of section 552b of Title 5, United States Code. Any such determination shall be in writing and shall contain the reasons for such determination. If such a determination is made, the advisory committee shall issue a report at least annually setting forth a summary of its activities and such related matters as would be informative to the public consistent with the policy of section 552(b) of Title 5, United States Code.

(e) There shall be designated an officer or employee of the Federal Government to chair or attend each meeting of each advisory committee. The officer or employee so designated is authorized, whenever he determines it to be in the public interest, to adjourn any such meeting. No advisory committee shall conduct any meeting in the absence of that officer or employee.

(f) Advisory committees shall not hold any meetings except at the call of, or with the advance approval of, a designated officer or employee of the Federal Government, and in the case of advisory committees (other than Presidential advisory committees), with an agenda approved by such officer or employee.

**Section 11. Availability of transcripts; "agency proceeding"**

(a) Except where prohibited by contractual agreements entered into prior to the effective date of this Act, agencies and advisory committees shall make available to any person, at actual cost of duplication, copies of transcripts of agency proceedings or advisory committee meetings.

(b) As used in this section "agency proceeding" means any proceeding as defined in section 551(12) of Title 5, United States Code.

**Section 12. Fiscal and administrative provisions; record-keeping; audit;**

**agency support services**

(a) Each agency shall keep records as will fully disclose the disposition of any funds which may be at the disposal of its advisory committees and the nature and extent of their activities. The General Services Administration, or such other agency as the President may designate, shall maintain financial records with respect to Presidential advisory committees. The Comptroller General of the United States, or any of his authorized representatives, shall have access, for the purpose of audit and examination, to any such records.

(b) Each agency shall be responsible for providing support services for each advisory committee established or reporting to it unless the establishing authority provides otherwise. Where any such advisory committee reports to more than one agency, only one agency shall be responsible for support services at any one time. In the case of Presidential advisory committees, such services may be provided by the General Services Administration.

**Section 13. Responsibilities of Library of Congress; reports and background papers; depository**

Subject to section 552 of Title 5, United States Code, the Administrator shall provide for the filing with the Library of Congress of at least eight copies of each report made by every advisory committee and, where appropriate, background papers prepared by consultants. The Librarian of Congress shall establish a depository for such reports and papers where they shall be available to public inspection and use.

**Section 14. Termination of advisory committees; renewal; continuation**

(a)(1) Each advisory committee which is in existence on the effective date of this Act shall terminate not later than the expiration of the two-year period following such effective date unless --

(A) in the case of an advisory committee established by the President or an officer of the Federal Government, such advisory committee is renewed by the President or that officer by appropriate action prior to the expiration of such two-year period; or

(B) in the case of an advisory committee established by an Act of Congress, its duration is otherwise provided for by law.

(2) Each advisory committee established after such effective date shall terminate not later than the expiration of the two-year period beginning on the date of its establishment unless --

(A) in the case of an advisory committee established by the President or an officer of the Federal Government such advisory committee is renewed by the President or such officer by appropriate action prior to the end of such period; or

(B) in the case of an advisory committee established by an Act of Congress, its duration is otherwise

provided for by law.

(b)(1) Upon the renewal of any advisory committee, such advisory committee shall file a charter in accordance with section 9(c).

(2) Any advisory committee established by an Act of Congress shall file a charter in accordance with such section upon the expiration of each successive two-year period following the date of enactment of the Act establishing such advisory committee.

(3) No advisory committee required under this subsection to file a charter shall take any action (other than preparation and filing of such charter) prior to the date on which such charter is filed.

(c) Any advisory committee which is renewed by the President or any officer of the Federal Government may be continued only for successive two-year periods by appropriate action taken by the President or such officer prior to the date on which such advisory committee would otherwise terminate.

**Section 15. Requirements Relating to the National Academy of Sciences and the**

**National Academy of Public Administration**

(a) An agency may not use any advice or recommendations provided by the National Academy of Sciences or National Academy of Public Administration that was developed by use of a committee created by that academy under an agreement with the agency, unless --

(1) the committee was not subject to any actual management or control by an agency or an officer of the Federal Government;

(2) in the case of a committee created after the date of the enactment of the Federal Advisory Committee Act Amendments of 1997, the membership of the committee was appointed in accordance with the requirements described in subsection (b)(1); and

(3) in developing the advice or recommendation, the academy complied with --

(A) subsection (b)(2) through (6), in the case of any advice or recommendation provided by the National Academy of Sciences, or

(B) subsection (b)(2) and (5), in the case of any advice or recommendation provided by the National Academy of Public Administration.

(b) The requirements referred to in subsection (a) are as follows:

(1) The Academy shall determine and provide public notice of the names and brief biographies of individuals that the Academy appoints or intends to appoint to serve on the committee. The Academy shall determine and provide a reasonable opportunity for the public to comment on such appointments before they are made or, if the Academy determines such prior comment is not practicable, in the period immediately following the appointments. The Academy shall make its best efforts to ensure that

(A) no individual appointed to serve on the committee has a conflict of interest that is relevant to the functions to be performed, unless such conflict is promptly and publicly disclosed and the Academy determines that the conflict is unavoidable,

(B) the committee membership is fairly balanced as determined by the Academy to be appropriate for the functions to be performed, and

(C) the final report of the Academy will be the result of the Academy's independent judgment. The Academy shall require that individuals that the Academy appoints or intends to appoint to serve on the committee inform the Academy of the individual's conflicts of interest that are relevant to the functions to be performed.

(2) The Academy shall determine and provide public notice of committee meetings that will be open to the public.

(3) The Academy shall ensure that meetings of the committee to gather data from individuals who are not officials, agents, or employees of the Academy are open to the public, unless the Academy determines that a meeting would disclose matters described in section 552(b) of Title 5, United States Code. The Academy shall make available to the public, at reasonable charge if appropriate, written materials presented to the committee by individuals who are not officials, agents, or employees of the Academy, unless the Academy determines that making material available would disclose matters described in that section.

(4) The Academy shall make available to the public as soon as practicable, at reasonable charge if appropriate, a brief summary of any committee meeting that is not a data gathering meeting, unless the Academy determines that the summary would disclose matters described in section 552(b) of Title 5, United States Code. The summary shall identify the committee members present, the topics discussed, materials made available to the committee, and such other matters that the Academy determines should be included.

(5) The Academy shall make available to the public its final report, at reasonable charge if appropriate, unless the Academy determines that the report would disclose matters described in section 552(b) of Title 5, United States Code. If the Academy determines that the report would disclose matters described in that section, the Academy shall make public an abbreviated version of the report that does not disclose those matters.

(6) After publication of the final report, the Academy shall make publicly available the names of the principal reviewers who reviewed the report in draft form and who are not officials, agents, or employees of the Academy.

(c) The Administrator of General Services may issue regulations implementing this section.

(d)(1) Except as provided in paragraph (2), this section and the amendments made by this section shall take effect on the date of the enactment of this Act.

(2) Subsection (a) and the amendments made by subsection (a) shall be effective as of October 6, 1972, except that they shall not apply with respect to or otherwise affect any particular advice or recommendations that are subject to any judicial action filed before the date of the enactment of this Act.

Not later than 1 year after the date of the enactment of this Act, the Administrator of General Services shall submit a report to the Congress on the implementation of and compliance with the amendments made by this Act.

**Section 16. Effective Date**

Except as provided in section 7(b), this Act shall become effective upon the expiration of ninety days following the date of enactment.

**Freedom of Information Act**

**5 U.S.C. § 552**

(a) Each agency shall make available to the public information as follows:

(1) Each agency separately state and currently publish in the Federal Register for the guidance of the public --

(A) descriptions of its central and field organizations and the established places at which, the employees (and in the case od a uniformed service, the members), from whom, and the methods whereby, the public may obtain information, make submittals or requests, or obtain decisions;

(B) statements of the general course and method by which its functions are channeled and determined, including the nature and requirements of all formal procedures available;

(C) rules of procedure, descriptions of forms available or the places at which forms may be obtained, and instructions as to the scope and contents of all papers, reports, or examinations;

(D) substantive rules of general applicability adopted as authorized by law, and statements of general policy or interpretations of general applicability formulated and adopted by the agency; and

(E) each amendment, revision, or repeal of the foregoing.

Except to the extent that a person has actual and timely notice of the terms thereof, a person may not in any manner be required to resort to, or be adversely affected by, a matter required to be published in the Federal Register and not so published. For the purpose of this paragraph, matter reasonably available to the class of persons affected thereby is deemed published in the Federal Register when incorporated by reference therein with the approval of the Director of the Federal Register.

(2) Each agency, in accordance with published rules, shall make available for public inspection and copying --

(A) final opinions, including concurring and dissenting opinions, as well as orders, made in the adjudication of cases;

(B) those statements of policy and interpretations which have been adopted by the agency and are not published in the Federal Register;

(C) administrative staff manuals and instructions to staff that affect a member of the public;

(D) copies of all records, regardless of form or format, which have been released to any person under paragraph (3) and which, because of the nature of the subject matter, the agency determines have

become or are likely to become the subject of subsequent requests for substantially the same records; and

(E) a general index of records referred to under subparagraph (D);

unless the materials are promptly published and copies offered for sale. For records created on or after November 1, 1996, within one year after such date, each agency shall make such records available, including by computer telecommunications or, if computer telecommunications means have not been established by the agency, by other electronic means. To the extent required to prevent a clearly unwarranted invasion of personal privacy, an agency may delete identifying details when it makes available or publishes an opinion, statement of policy, interpretation, staff manual, instruction, or copies of records referred to in subparagraph (D). However, in each case the justification for the deletion shall be explained fully in writing, and the extent of such deletion shall be indicated at the place in the record where the deletion was made. Each agency shall also maintain and make available for public inspection and copying, current indexes providing identifying information for the public as to any matter issued, adopted, or promulgated after July 4, 1967, and required by this paragraph to be made available or published. Each agency shall make the index referred to in subparagraph (E) available by computer telecommunications by December 31, 1999. Each agency shall promptly publish, quarterly or more frequently, and distribute (by sale or otherwise) copies of each index or supplements thereto unless it determines by order published in the Federal Register that the publication would be unnecessary and impracticable, in which case the agency shall nonetheless provide copies of such index on request at a cost not to exceed the direct cost of duplication. A final order, opinion, statement of policy, interpretation, staff manual or instruction that affects a member of the public may be relied on, used, or cited as precedent by an agency against a party other than an agency only if --

(i) it has been indexed and either made available or published as provided by this paragraph; or

(ii) the party has actual and timely notice of the terms thereof.

(3)(A) Except with respect to records made available under paragraphs (1) and (2) of this subsection, each agency, upon any request for records which (i) reasonably describes such records and (ii) is made in accordance with published rules stating the time, place, fees (if any), and procedures to be followed, shall make the records promptly available to any person.

(B) In making any record available to a person under this paragraph, an agency shall provide the record in the form or format requested by the person if the record is readily reproducible by the agency in that form or format. Each agency shall make reasonable efforts to maintain its records in forms or formats that are reproducible for purposes of this section.

(C) In responding under this paragraph to a request for records, an agency shall make reasonable efforts to search for the records in electronic form or format, except when such efforts would significantly interfere with the operation of the agency's automated information system.

(D) For purposes of this paragraph, the term "search" means to review, manually or by automated means, agency records for the purpose of locating those records that are responsive to the request.

(4)(A)(i) In order to carry out the provisions of this section, each agency shall promulgate regulations, pursuant to notice and receipt of public comment, specifying the schedule of fees applicable in the

processing of requests under this section and establishing procedures and guidelines for determining when such fees should be waived or reduced. Such schedule shall conform to the guidelines which shall be promulgated, pursuant to notice and receipt of public comment, by the Director of the Office of Management and Budget and which shall provide for a uniform schedule of fees for all agencies.

(ii) Such agency regulations shall provide that --

(I) fees shall be limited to reasonable standard charges for document search, duplication, and review, when records are requested for commercial use;

(II) fees shall be limited to reasonable standard charges for document duplication when records are not sought for commercial use and the request is made by an educational or noncommercial scientific institution, whose purpose is scholarly or scientific research; or a representative of the news media; and

(III) for any request not described in (I) or (II), fees shall be limited to reasonable standard charges for document search and duplication.

(iii) Documents shall be furnished without any charge or at a charge reduced below the fees established under clause (ii) if disclosure of the information is in the public interest because it is likely to contribute significantly to public understanding of the operations or activities of the government and is not primarily in the commercial interest of the requester.

(iv) Fee schedules shall provide for the recovery of only the direct costs of search, duplication, or review. Review costs shall include only the direct costs incurred during the initial examination of a document for the purposes of determining whether the documents must be disclosed under this section. Review costs may not include any costs incurred in resolving issues of law or policy that may be raised in the course of processing a request under this section. No fee may be charged by any agency under this section --

(I) if the costs of routine collection and processing of the fee are likely to equal or exceed the amount of the fee; or

(II) for any request described in clause (iii)(II) or (III) of this subparagraph for the first two hours of search time or for the first one hundred pages of duplication.

(v) No agency may require advance payment of any fee unless the requester has previously failed to pay fees in a timely fashion, or the agency has determined that the fee will exceed $250.

(vi) Nothing in this subparagraph shall supersede fees chargeable under a statute specifically providing for setting the level of fees for particular types of records.

(vii) In any action by a requester regarding the waiver of fees under this section, the court shall hear the matter de novo; Provided, That the court's review of the matter shall be limited to the record before the agency.

(B) On complaint, the district court of the United States in the district in which the complainant resides, or has his principal place of business, or in which the agency records are situated, or the District of Columbia, has jurisdiction to enjoin the agency from withholding agency records, and to

order the production of any agency records improperly withheld from the complainant. In such a case the court shall determine the matter de novo, and may examine the contents of such agency records in camera to determine whether such records or any part thereof shall be withheld under any of the exemptions set forth in subsection (b) of this section, and the burden is on the agency to sustain its action. In addition to any other matters to which a court accords substantial weight, a court shall accord substantial weight to an affidavit of an agency concerning the agency's determination as to technical feasibility under paragraph (2)(C) and subsection (b) and reproducibility under paragraph (3)(B).

(C) Notwithstanding any other provisions of law, the defendant shall serve an answer or otherwise plead to any complaint made under this subsection within thirty days after service upon the defendant of the pleading in which such complaint is made, unless the court otherwise directs for good cause shown.

(D) Except as to cases the court considers of greater importance, proceedings before the district court, as authorized by this subsection, and appeals therefrom, take precedence on the docket over all cases and shall be assigned for hearing and trial or for argument at the earliest practicable date and expedited in every way. **(Repealed by the *Federal Courts Civil Priorities Act*, P.L. 98-620, November 1984)**

(E) The court may assess against the United States reasonable attorney fees and other litigation costs reasonably incurred, in any case under this section in which the complainant has substantially prevailed.

(F) Whenever the court orders the production of any agency records improperly withheld from the complainant and assesses against the United States reasonable attorneys fees and other litigation costs, and the court additionally issues a written finding that the circumstances surrounding the withholding raise questions whether agency personnel acted arbitrarily or capriciously with respect to the withholding, the Special Counsel shall promptly initiate a proceeding to determine whether disciplinary action is warranted against the officer or employee who was primarily responsible for the withholding. The Special Counsel, after investigation and consideration of the evidence submitted, shall submit its findings and recommendations to the officer or employees or his representative. The administrative authority shall take the corrective action that the Special Counsel recommends.

(G) In the event of noncompliance with the order of the court, the district court may punish for contempt the responsible employee, and in the case of a uniformed service, the responsible member.

(5) Each agency having more than one member shall maintain and make available for public inspection a record of the final votes of each member in every agency proceeding.

(6)(A) Each agency, upon request for records made under paragraph (1), (2), or (3) of this subsection shall --

(i) determine within twenty days (excepting Saturdays, Sundays, and legal holidays) after the receipt of any such request whether to comply with such request and shall immediately notify the person making such request of such determination and the reasons therefor, and of the right of such person to appeal to the head of the agency any adverse determination; and

(ii) make a determination with respect to any appeal within twenty days (excepting Saturdays,

Sundays, and legal holidays) after the receipt of such appeal. If on appeal the denial of the request for records is in whole or part upheld, the agency shall notify the person making such request of the provisions for judicial review of that determination under paragraph (4) of this subsection.

(B)(i) In unusual circumstances as specified in this subparagraph, the time limits prescribed in either clause (i) or clause (ii) of subparagraph (A) may be extended by written notice to the person making such request setting forth the unusual circumstances for such extension and the date on which a determination is expected to be dispatched. No such notice shall specify a date that would result in an extension for more than ten working days, except as provided in clause (iii) of this subparagraph.

(ii) With respect to a request for which a written notice under clause (i) extends the time limits prescribed under clause (i) of subparagraph (A), the agency shall notify the person making the request if the request cannot be processed within the time limit specified in that clause and shall provide the person an opportunity to limit the scope of the request so that it may be processed within the time limit or an opportunity to arrange with the agency an alternative time frame for processing the request or modified request. Refusal by the person to reasonably modify the request or arrange such an alternative time frame be considered as a factor in determining whether exceptional circumstances exist for purposes of subparagraph (C).

(iii) As used in this subparagraph, "unusual circumstances" means, but only to the extent reasonably necessary to the proper processing of the particular requests --

(I) the need to search for and collect the requested records from field facilities or other establishments that are separate from the office processing the request;

(II) the need to search for, collect, and appropriately examine a voluminous amount of separate and district records which are demanded in a single request; or

(III) the need for consultation, which shall be conducted with all practicable speed, with another agency having a substantial interest in the determination of the request or among two or more components of the agency having substantial subject-matter interest therein.

(iv) Each agency may promulgate regulations, pursuant to notice and receipt of public comment, providing for the aggregation of certain requests by the same requester, or a group of requesters acting in concert, if the agency reasonably believes that such requests actually constitute a single request, which would otherwise satisfy the unusual circumstances specified in this subparagraph, and the requests involve clearly related matters. Multiple requests involving unrelated matters shall not be aggregated.

(C)(i) Any person making a request for any agency for records under paragraph (1), (2), or (3) of this subsection shall be deemed to have exhausted his administrative remedies with respect to such request if the agency fails to comply with the applicable time limit provisions of this paragraph. If the Government can show exceptional circumstances exist and that the agency is exercising due diligence in responding to the request, the court may retain jurisdiction and allow the agency additional time to complete its review of the records. Upon any determination by an agency to comply with a request for records, the records shall be made promptly available to such person making such request. Any notification of denial of any request for records under this subsection shall set forth the names and titles or positions of each person responsible for the denial of such request.

(ii) For purposes of this subparagraph, the term "exceptional circumstances" does not include a delay that results from a predictable agency workload of requests under this section.

(iii) Refusal by a person to reasonably modify the scope of a request or arrange an alternative time frame for processing a request (or a modified request) under clause (ii) after being given an opportunity to do so by the agency to whom the person made the request shall be considered as a factor in determining whether exceptional circumstances exist for purposes of this subparagraph.

(iv) In determining whether exceptional circumstances exist, a court shall consider the efforts by an agency to reduce the number of pending requests under this section.

(D)(i) Each agency may promulgate regulations, pursuant to notice and receipt of public comment, providing for multitrack processing of requests for records based on the amount of work or time (or both) involved in processing requests.

(ii) Regulations under this subparagraph may provide a person making a request that does not qualify for the fastest multitrack processing an opportunity to limit the scope of the request in order to qualify for faster processing.

(iii) This subparagraph shall not be considered to affect the requirement under subparagraph (C) to exercise due diligence.

(E)(i) Each agency shall promulgate regulations, pursuant to notice and receipt of public comment, providing for expedited processing of request for records --

(I) in cases in which the person requesting the records demonstrates a compelling need; and

(II) in other cases determined by the agency.

(ii) Notwithstanding subparagraph (A)(i), regulations under this subparagraph must ensure --

(I) that a determination of whether to provide expedited processing shall be made, and notice of the determination shall be provided to the person making the request, within 10 days after the date of the request; and

(II) expeditious consideration of administrative appeals of such determinations of whether to provide expedited processing;

(iii) An agency shall process as soon as practicable any request for records to which the agency has granted expedited processing under this subparagraph. Agency action to deny or affirm denial of a request for expedited processing pursuant to this subparagraph, and failure by an agency to respond timely to such a request shall be subject to judicial review under paragraph (4), except that the judicial review shall be based on the record before the agency at the time of the determination.

(iv) A district court of the United States shall not have jurisdiction to review an agency denial of expedited processing of a request for records after the agency has provided a complete response to the request.

(v) For purposes of this subparagraph, the term "compelling" means --

(I) that a failure to obtain requested records on an expedited basis under this paragraph could reasonably be expected to pose an imminent threat to the life or physical safety of an individual; or

(II) with respect to a request made by a person primarily engaged in disseminating information, urgency to inform the public concerning actual or alleged Federal Government activity.

(vi) A demonstration of compelling need by a person making a request for expedited processing shall be made by a statement certified by such person to be true and correct to the best of such person's knowledge and belief.

(F) In denying a request for records, in whole or in part, an agency shall make a reasonable effort to estimate the volume of any requested matter the provision of which is denied, and shall provide any such estimate to the person making the request, unless providing such estimate would harm an interest protected by the exemption in subsection (b) pursuant to which the denial is made.

(b) This section does not apply to matters that are --

(1)(A) specifically authorized under criteria established by an Executive order to be kept secret in the interest of national defense or foreign policy and (B) are in fact properly classified pursuant to such Executive order;

(2) related solely to the internal personnel rules and practices of an agency;

(3) specifically exempted from disclosure by statute (other than section 552b of this title), provided that such statute --

(A) requires that the matter be withheld from the public in such a manner as to leave no discretion on the issue; or

(B) establishes particular criteria for withholding or refers to particular types of matters to be withheld;

(4) trade secrets and commercial or financial information obtained from a person and privileged or confidential;

(5) inter-agency or intra-agency memorandums or letters which would not be available by law to a party other than agency in litigation with the agency;

(6) personnel and medical files and similar files the disclosure of which would constitute a clearly unwarranted invasion of personal privacy;

(7) records or information compiled for law enforcement purposes, but only to the extent that the production of such law enforcement records or information --

(A) could reasonably be expected to interfere with enforcement proceedings;

(B) would deprive a person of a right to a fair trial or an impartial adjudication;

(C) could reasonably be expected to constitute an unwarranted invasion of person privacy;

(D) could reasonably be expected to disclose the identity of a confidential source, including a State, local or foreign agency or authority or any private institution which furnished information on a confidential basis, and, in the case of a record or information compiled by a criminal law enforcement authority in the course of a criminal investigation or by an agency conducting a lawful national security intelligence investigation, information furnished by a confidential source;

(E) would disclose techniques and procedures for law enforcement investigations or prosecutions, or would disclose guidelines for law enforcement investigations or prosecutions if such disclosure could reasonably be expected to risk circumvention of the law; or

(F) could reasonably be expected to endanger the life or physical safety of any individual.

(8) contained in or related to examination, operating, or condition reports prepared by, on behalf of, or for the use of an agency responsible for the regulation or supervision of financial institutions; or

(9) geological and geophysical information and data, including maps, concerning wells.

Any reasonably segregable portion of a record shall be provided to any person requesting such record after deletion of the portions which are exempt under this subsection. The amount of information deleted shall be indicated on the released portion of the record, unless including that indication would harm an interest protected by the exemption in this subsection under which the deletion is made. If technically feasible, the amount of information deleted shall be indicated at the place in the record where such deletion is made.

(c)(1) Whenever a request is made which involves access to records described in subsection (b)(7)(A) and --

(A) the investigation or proceeding involves a possible violation of criminal law; and

(B) there is reason to believe that --

(i) the subject of the investigation or proceeding is not aware of the pendency; and

(ii) disclosure of the existence of the records could reasonably be expected to interfere with enforcement proceedings, the agency may, during only such time as that circumstance continues, treat the records as not subject to the requirements of this section.

(2) Whenever informant records maintained by a criminal law enforcement agency under an informant's name or personal identifier by a third party according to the informant's name or personal identifier, the agency may treat the records as not subject to the requirements of this section unless the informant's status as an informant has been officially confirmed.

(3) Whenever a request is made which involves access to records maintained by the Federal Bureau of Investigation pertaining to foreign intelligence or counterintelligence, or international terrorism, and the existence of the records is classified information as provided in subsection (b)(1), the Bureau may, as long as the existence of the records remains classified information, treat the records as not

subject to the requirements of this section.

(d) This section does not authorize the withholding of information or limit the availability of records to the public, except as specifically stated in this section. This section is not authority to withhold information from Congress.

(e)(1) On or before February 1 of each year, each agency shall submit to the Attorney General of the United States a report which shall cover the preceding fiscal year and which shall

include --

(A) the number of determinations made by such agency not to comply with requests for records made to such agency under subsection (a) and the reasons for each such determination;

(B)(i) the number of appeals made by persons under subsection (a)(6), the result of such appeals, and the reason for the action upon each appeal that results in a denial of information; and

(ii) a complete list of all statutes that the agency relies upon to authorize the agency to withhold information under subsection (b)(3), a description of whether a court has upheld the decision of the agency to withhold under each such statute, and a concise description of the scope of any information withheld;

(C) the number of requests for records pending before the agency as of September 30 of the preceding year, and the median number of days that such requests had been pending before the agency as of that date;

(D) the number of requests for records received by the agency and the number of requests which the agency processed;

(E) the median number of days taken by the agency to process different types of requests;

(F) the total amount of fees collected by the agency for processing requests; and

(G) the number of full-time staff of the agency devoted to processing requests for records under this section, and the total amount expended by the agency for processing such requests.

(2) Each agency shall make each such report available to the public including by computer telecommunications, or if computer telecommunications have not been established by the agency by other electronic means.

(3) The Attorney General of the United States shall make each report which has been made available by electronic means available at a single electronic access point. The Attorney General of the United States shall notify the Chairman and ranking minority member of the Committee on Government Reform and Oversight of the House of Representatives and the Chairman and ranking minority member of the Committees on Government Affairs and the Judiciary of the Senate, no later than April 1 of the year in which each such report is issued, that such reports are available by electronic means.

(4) The Attorney General of the United States, in consultation with the Director of the Office of

Management and Budget, shall develop reporting and performance guidelines in connection with reports required by this subsection by October 1, 1997, and may establish additional requirements for such reports as the Attorney General determines may be useful.

(5) The Attorney General of the United States shall submit an annual report on or before April 1 of each calendar year which shall include for the prior calendar year a listing of the number of cases arising under this section, the exemption involved in each case, the disposition of such case, and the cost, fees, and penalties assessed under subparagraphs (E), (F), and (G) of subsection (a)(4). Such report shall also include a description of the efforts undertaken by the Department of Justice to encourage agency compliance with this section.

(f) For purposes of this section, the term --

(1) "agency" as defined in section 551(1) of this title includes any executive department, military department, Government corporation, Government controlled corporation, or other establishment in the executive branch of Government (including the Executive Office of the President), or any independent regulatory agency; and

(2) "record" and any other term used in this section in reference to information includes any information that would be an agency record subject to the requirements of this section if maintained by an agency in any format, including an electronic format.

(g) The head of each agency shall prepare and make publicly available upon request, reference material or a guide for requesting records or information from the agency, subject to the exemptions in subsection (b), including --

(1) an index of all major information systems of the agency'

(2) a description of major information and record locator systems maintained by the agency; and

(3) a handbook for obtaining various types and categories of public information from the agency pursuant to chapter 35 of title 44, and under this section.

# Government in the Sunshine Act

## 5 U.S.C. 552b

### Open Meetings

(a) For purposes of this section --

(1) the term "agency" means any agency, as defined in section 552(e) of this title, headed by a collegial body composed of two or more individual members, a majority of whom are appointed to such position by the President with the advice and consent of the Senate, and any subdivision thereof authorized to act on behalf of the agency;

(2) the term "meeting" means the deliberations of at least the number of individual agency members required to take action on behalf of the agency where such deliberations determine or result in the joint conduct or disposition of official agency business, but does not include deliberations required or permitted by subsection (d) or (e); and

(3) the term "member" means an individual who belongs to a collegial body heading an agency.

(b) Members shall not jointly conduct or dispose of agency business other than in accordance with this section. Except as provided in subsection (c), every portion of every meeting of an agency shall be open to public observation.

(c) Except in a case where the agency finds that the public interest requires otherwise, the second sentence of subsection (b) shall not apply to any portion of an agency meeting, and the requirements of subsections (d) and (e) shall not apply to any information pertaining to such meeting otherwise required by this section to be disclosed to the public, where the agency properly determines that such portion or portions of its meeting or the disclosure of such information is likely to --

(1) disclose matters that are (A) specifically authorized under criteria established by an executive order to be kept secret in the interest of national defense and foreign policy and (B) in fact properly classified pursuant to such executive order;

(2) relate solely to the internal personnel rules and practices of an agency;

(3) disclose matters specifically exempted from disclosure by statute (other than 552 of this title), provided that such statute (A) requires that the matters to be withheld from the public in such a manner as to leave no discretion on the issue, or (B) establishes particular criteria for withholding or refers to particular types of matters to be withheld;

(4) disclose trade secrets and commercial or financial information obtained from a person and privileged or confidential;

(5) involve accusing any person of a crime, or formally censuring any person;

(6) disclose information of a personal nature where disclosure would constitute a clearly unwarranted invasion of personal privacy;

(7) disclose investigatory records compiled for law enforcement purposes, or information which if written would be contained in such records, but only to the extent that the production of such records or information would

(A) interfere with enforcement proceedings,

(B) deprive a person of a right to a fair trial or an impartial adjudication,

(C) constitute an unwarranted invasion of personal privacy,

(D) disclose the identity of a confidential source and, in the case of a record compiled by a criminal law enforcement authority in the course of a criminal investigation, or by an agency conducting a lawful national security intelligence investigation, confidential source information furnished only by the confidential source;

(E) disclose investigative techniques and procedures, or

(F) endanger the life or physical safety of law enforcement personnel;

(8) disclose information contained in or related to examination, operating, or condition reports prepared by, on behalf of, or for the use of an agency responsible for the regulation or supervision of financial institutions;

(9) disclose information the premature disclosure of which would --

(A) in the case of any agency which regulates currencies, securities, commodities, or financial institutions, be likely to

(i) lead to significant financial speculation in currencies, securities, or commodities, or

(ii) significantly endanger the stability of any financial institution; or

(B) in the case of any agency, be likely to significantly frustrate implementation of a proposed agency action;

except that subparagraph (B) shall not apply in any instance where the agency has already disclosed to the public the content or nature of its proposed action, or where the agency is required by law to make such disclosure on its own initiative prior to taking final agency action on such proposal; or

(10) specifically concern the agency's issuance of a subpoena, or the agency's participation in a civil

action or proceeding, an action in a foreign court or international tribunal, or an arbitration, or the initiation, conduct, or disposition by the agency of a particular case of formal agency adjudication pursuant to the procedures in section 554 of this title or otherwise involving a determination on the record after opportunity for a hearing.

**Vote to Close Meetings**

(d)(1) Action under subsection (c) shall be taken only when a majority of the entire membership of the agency (as defined in subsection (a)(1)) votes to take such action. A separate vote of the agency members shall be taken with respect to each agency meeting a portion or portions of which are proposed to be closed to the public pursuant to subsection (c) or with respect to any information which is proposed to be withheld under subsection (c). A single vote may be taken with respect to a series of meetings, a portion or portions of which are proposed to be closed to the public, or with respect to any information concerning such series of meetings, so long as each meeting in such series involves the same particular matters and is scheduled to be held no more than thirty days after the initial meeting in such series. The vote of each agency member participating in such vote shall be recorded and no proxies shall be allowed.

(2) Whenever any person whose interests may be directly affected by a portion of a meeting requests that the agency close such portion to the public for any of the reasons referred to in paragraph (5), (6), or (7) of subsection (c), the agency, upon request of any one of its members, shall vote by recorded vote whether to close such meeting.

(3) Within one day of any vote taken pursuant to paragraph (1) or (2), the agency shall make publicly available a written copy of such vote reflecting the vote of each member on the question. If a portion of a meeting is to be closed to the public, the agency shall, within one day of the vote taken pursuant to paragraph (1) or (2) of this subsection, make publicly available a full written explanation of its action closing the portion together with a list of all persons expected to attend the meeting and their affiliation.

(4) Any agency, a majority of whose meetings may properly be closed to the public pursuant to paragraph (4), (8), (9)(A), or (10) of subsection (c), or any combination thereof, may provide by regulation for the closing of such meetings or portions thereof in the event that a majority of the members of the agency votes by recorded vote at the beginning of such meeting, or portion thereof, to close the exempt portion or portions of the meeting, and a copy of such vote, reflecting the vote of each member on the question, is made available to the public. The provisions of paragraphs (1), (2), and (3) of this subsection and subsection (e) shall not apply to any portion of a meeting to which such regulations apply; *Provided*, That the agency shall, except to the extent that such information is exempt from disclosure under the provisions of subsection (c), provide the public with public announcement of the time, place, and subject matter of the meeting and of each portion thereof at the earliest practicable time.

**Notice of Meetings**

(e)(1) In the case of each meeting, the agency shall make public announcement, at least one week before the meeting, of the time, place, and subject matter of the meeting, whether it is to be open or closed to the public, and the name and phone number of the official designated by the agency to respond to requests for information about the meeting. Such announcement shall be made unless a majority of the members of the agency determines by a recorded vote that agency business requires

that such meeting be called at an earlier date, in which case the agency shall make public announcement of the time, place, and subject matter of such meeting, and whether open or closed to the public, at the earliest practicable time.

(2) The time or place of a meeting may be changed following the public announcement required by paragraph (1) only if the agency publicly announces such change at the earliest practicable time. The subject matter of a meeting. or the determination of the agency to open or close a meeting, or portion of a meeting, to the public, may be changed following the public announcement required by this subsection only if

(A) a majority of the entire membership of the agency determines by a recorded vote that agency business so requires and that no earlier announcement of the change was possible and

(B) the agency publicly announces such change and the vote of each member upon such change at the earliest practicable time.

(3) Immediately following each public announcement required by this subsection, notice of time, place, and subject matter of a meeting, whether the meeting is open or closed, any change in one of the preceding, and the name and phone number of the official designated by the agency to respond to requests for information about the meeting, shall be submitted for publication in the *Federal Register*.

**Closed Meetings**

(f)(1) For every meeting closed pursuant to paragraphs (1) through (10) of subsection (c), the General Counsel or chief legal officer of the agency shall publicly certify that, in his or her opinion, the meeting may be closed to the public and shall state each relevant exemption provision. A copy of such certification, together with a statement from the presiding officer of the meeting setting forth the time and place of the meeting, and the persons present, shall be retained by the agency. The agency shall maintain a complete transcript or electronic recording adequate to record fully the proceedings of each meeting, or portion of a meeting, closed to the public, except that in the case of a meeting, or portion of a meeting, closed to the public pursuant to paragraph (8), (9)(A), or (10) of subsection (c), the agency shall maintain either such a transcript or recording, or a set of minutes. Such minutes shall fully and clearly describe all matters discussed and shall provide a full and accurate summary of any action taken, and the reasons therefor, including a description of each of the views expressed on any item and the record of any roll call vote (reflecting the vote of each member on the question). All documents considered in connection with any action shall be identified in such minutes.

**Transcript of Closed Meetings**

(2) The agency shall make promptly available to the public, in a place easily accessible to the public, the transcript, electronic recording, or minutes (as required by paragraph (1)) of the discussion of any item on the agenda, or of any item of the testimony of any witness received at the meeting, except for such item or items of such discussions or testimony as the agency determines to contain information which may be withheld under subsection (c). Copies of such transcript, or minutes, or a transcription of such recording disclosing the identity of each speaker, shall be furnished to any person at the actual cost of duplication or transcription. The agency shall maintain a complete verbatim copy of the transcript, a complete copy of the minutes, or a complete electronic recording of each meeting, or portion of a meeting, closed to the public, for a period of at least two years after such meeting, or

until one year after the conclusion of any agency proceeding with respect to which the meeting or portion was held, whichever occurs later.

**Regulations**

(g) Each agency subject to the requirements of this section shall, within 180 days after the date of enactment of this section, following consultation with the Office of the Chairman of the Administrative Conference of the United States and published notice in the *Federal Register* of at least thirty days and opportunity for written comment by any person, promulgate regulations to implement the requirements of subsections (b) through (f) of this section. Any person may bring a proceeding in the United States District Court for the District of Columbia to require an agency to promulgate such regulations if such agency has not promulgated such regulations within the time period specified herein. Subject to any limitations of time provided by law, any person may bring a proceeding in the United States Court of Appeals for the District of Columbia to set aside agency regulations issued pursuant to this subsection that are not in accord with the requirement of subsections (b) through (f) of this section and to require the promulgation of regulations that are in accord with such subsection.

**Judicial Review**

(h)(1) The district courts of the United States shall have jurisdiction to enforce the requirements of subsections (b) through (f) of this section by declaratory judgment, injunctive relief, or other relief as may be appropriate. Such actions may be brought by any person against an agency prior to, or within sixty days after, the meeting out of which the violation of this section arises, except that if public announcement of such meeting is not initially provided by the agency in accordance with the requirements of this section, such action may be instituted pursuant to this section at any time prior to sixty days after any public announcement of such meeting. Such actions may be brought in the district court of the United States for the district in which the agency meeting is held or in which the agency in question has its headquarters, or in the District Court for the District of Columbia. In such actions a defendant shall serve his answer within thirty days after the service of the complaint. The burden is on the defendant to sustain his action. In deciding such cases, the court may examine in camera any portion of the transcript, electronic recording, or minutes of a meeting closed to the public, and may take such additional evidence as it deems necessary. The court, having due regard for orderly administration and the public interest, as well as the interests of the parties, may grant such equitable relief as it deems appropriate, including granting an injunction against future violations of this section or ordering the agency to make available to the public such portion of the transcript, recording, or minutes of a meeting as is not authorized to be withheld under subsection (c) of this section.

(2) Any federal court otherwise authorized by law to review agency action may, at the application of any person properly participating in the proceeding pursuant to other applicable law, inquire into violations by the agency of the requirements of this section and afford such relief as it deems appropriate. Nothing in this section authorizes any federal court having jurisdiction solely on the basis of paragraph (1) to set aside, enjoin, or invalidate any agency action (other than an action to close a meeting or to withhold information under this section) taken or discussed at any agency meeting out of which the violation of this section arose.

(i) The court may assess against any party reasonable attorney fees and other litigation costs reasonably incurred by any other party who substantially prevails in any action brought in accordance with the provisions of subsection (g) or (h) of this section, except that costs may be assessed against

the plaintiff only where the court finds that the suit was initiated by the plaintiff primarily for frivolous or dilatory purposes. In the case of assessment costs against an agency, the costs may be assessed by the court against the United States.

**Report to Congress**

(j) Each agency subject to the requirements of this section shall annually report to Congress regarding its compliance with such requirements, including a tabulation of the total number of agency meetings open to the public, the total number of meetings closed to the public, the reasons for closing such meetings, and a description of any litigation brought against the agency under this section, including any costs assessed against the agency in such litigation (whether or not paid by the agency).

(k) Nothing herein expands or limits the present rights of any person under section 552 of this title, except that the exemptions set forth in subsection (c) of this section shall govern in the case of any request made pursuant to section 552 to copy or inspect the transcripts, recordings, or minutes described in subsection (f) of this section. the requirements of chapter 33 of title 44, United States Code, shall not apply to the transcripts, recordings, and minutes described in subsection (f) of this section.

(l) This section does not constitute authority to withhold any information from Congress, and does not authorize the closing of any agency meeting or portion thereof required by any other provision of law to be open.

(m) Nothing in this section authorizes any agency to withhold from any individual any record, including transcripts, recordings, or minutes required by this section, which is otherwise accessible to such individual under section 552a of this title.

# The Privacy Act of 1974

## 5 U.S.C. 552a

### (a) Definitions

For purposes of this section:

(1) the term "agency" means agency as defined in section 552(e) of this title;

(2) the term "individual" means a citizen of the United States or an alien lawfully admitted for permanent residence;

(3) the term "maintain" includes maintain, collect, use or disseminate;

(4) the term "record" means any item, collection, or grouping of information about an individual that is maintained by an agency, including, but not limited to, his education, transactions, medical history, and criminal or employment history and that contains his name, or the identifying number, symbol, or other identifying particular assigned to the individual, such as a finger or voice print or a photograph;

(5) the term "system of records" means a group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying particular assigned to the individual;

(6) the term "statistical record" means a record in a system of records maintained for statistical research or reporting purposes only and not used in whole or in part in making any determination about an identifiable individual; except as provided in section 8 of Title 13;

(7) the term "routine use" means, with respect to the disclosure of a record, the use of such record for a purpose which is compatible with the purpose for which it was collected;

(8) the term "matching program" --

(A) means any computerized comparison of --

(i) two or more automated systems of records or a system of records with non-Federal records for the purpose of --

(I) establishing or verifying the eligibility of, or continuing compliance with statutory and regulatory requirements by, applicants for, recipients or beneficiaries of, participants in, or providers of services with respect to, cash or in-kind assistance or payments under Federal benefit programs; or

(II) recouping payments or delinquent debts under such Federal benefit programs, or

(ii) two or more automated Federal personnel or payroll systems of records or a system of Federal personnel or payroll records with non-Federal records,

(B) but does not include --

(i) matches performed to produce aggregate statistical data without personal identifiers;

(ii) matches performed to support any research or statistical project, the specific data of which may not be used to make decisions concerning the rights, benefits, or privileges of specific individuals;

(iii) matches performed, by an agency (or component thereof) which performs as its principal function any activity pertaining to the enforcement of criminal laws, subsequent to the initiation of a specific criminal or civil law enforcement investigation of a named person or persons for the purpose of gathering evidence against such person or persons;

(iv) matches of tax information

(I) pursuant to section 6103(d) of the Internal Revenue Code of 1986,

(II) for purposes of tax administration as defined in section 6103(b)(4) of such Code,

(III) for the purpose of intercepting a tax refund due an individual under authority granted by section 464 or 1137 of the Social Security Act, or

(IV) for the purpose of intercepting a tax refund due an individual under any other tax fund intercept program authorized by statute which has been determined by the Director of the Office of Management and Budget to contain verification, notice, and hearing requirements that are substantially similar to the procedures in section 1137 of the Social Security Act;

(v) matches --

(I) using records predominantly relating to Federal personnel, that are performed for routine administrative purposes (subject to guidance provided by the Director of the Office of Management and Budget pursuant to subsection (v); or

(II) conducted by an agency using only records from systems of records maintained by that agency;

If the purpose of the match is not to take any adverse financial, personnel, disciplinary, or other adverse action against Federal personnel; or

(vi) matches performed for foreign counterintelligence purposes or to produce background checks for security clearances of Federal personnel or Federal contractor personnel;

(9) the term "recipient agency" means any agency, or contractor thereof, receiving records contained in a system of records from a source agency for use in a matching program;

(10) the term "non-Federal agency" means any State or local government, or agency thereof, which receives records contained in a system of records from a source agency for use in a matching program;

(11) the term "source agency" means any agency which discloses records contained in a system of records to be used in a matching program, or any State of local government, or agency thereof, which discloses records to be used in a matching program;

(12) the term "Federal benefit program" means any program administered or funded by the Federal Government, or by any agency of a State on behalf of the Federal Government, providing cash or in-kind assistance in the form of payments, grants, loans, or loan guarantees to individuals; and

(13) the term "Federal personnel" means officers and employees of the Government of the United States, members of the uniformed services (including members of the Reserve Components), and individuals entitled to receive immediate or deferred retirement benefits under any retirement program of the Government of the United States (including survivor benefits).

**(b) Conditions of disclosure**

No agency shall disclose any record which is contained in a system of records by any means of communication to any person, or to another agency, except pursuant to a written request by, or with the prior consent of, the individual to whom the record pertains, unless disclosure of the record would be:

(1) to those officers and employees of the agency which maintains the record who have a need for the record in the performance of their duties;

(2) required under section 552 of this title;

(3) for a routine use as defined in subsection (a)(7) of this section and described under subsection (e)(4)(D) of this section;

(4) to the Bureau of Census for purposes of planning or carrying out a census or survey or related activity pursuant to the provisions of Title 13;

(5) to a recipient who has provided the agency with advance adequate written assurance that the records will be used solely as a statistical research or reporting record, and the record is transferred in a form that is not individually identifiable;

(6) to the National Archives of the United States as a record which has sufficient historical or other value to warrant its continued preservation by the United States Government, or for evaluation by the Administrator of General Services or his designee to determine whether the record has such value;

(7) to another agency or to an instrumentality of any governmental jurisdiction within or under the control of the United States for a civil or criminal law enforcement activity if the activity is authorized by law, and if the head of the agency or instrumentality has made a written request to the agency which maintains the record specifying the particular portion desired and the law enforcement activity for which the record is sought;

(8) to a person pursuant to a showing of compelling circumstances affecting the health or safety of an individual if upon such disclosure notification is transmitted to the last known address of such individual;

(9) to either House of Congress, or to the extent of matter within its jurisdiction, any committee or subcommittee thereof, any joint committee of Congress, or subcommittee of any joint committee;

(10) to the Comptroller General, or any of his authorized representatives, in the course of the performance of the duties of the General Accounting Office;

(11) pursuant to the order of a court of competent jurisdiction;

(12) to a consumer reporting agency in accordance with section 3711(f) of Title 31.

**(c) Accounting for certain disclosures**

Each agency, with respect to each system of records under its control shall:

(1) except for disclosures made under subsections (b)(1) or (b)(2) of this section keep an accurate accounting of:

(A) the date, nature, and purpose of each disclosure of a record to any person or to another agency made under subsection (b); and

(B) the name and address of the person or agency to whom the disclosure is made;

(2) retain the accounting made under paragraph (1) of this subsection for at least five years or the life of the record, whichever is longer, after the disclosure for which the accounting is made;

(3) except for disclosures made under subsection (b)(7) of this section, make the accounting under paragraph (1) of this subsection available to the individual named in the record at his request; and

(4) inform any person or other agency about any correction or notation of dispute made by the agency in accordance with subsection (d) of this section of any record that has been disclosed to the person or agency if an accounting of the disclosure was made.

**(d) Access to records**

Each agency that maintains a system of records shall:

(1) upon request by any individual to gain access to his record or to any information pertaining to him which is contained in the system, permit him and, upon his request, a person of his choosing to accompany him, to review the record and have a copy made of all or any portion thereof in a form comprehensible to him, except that the agency may require the individual to furnish a written statement authorizing discussion of that individual's record in the accompanying person's presence;

(2) permit the individual to request amendment of a record pertaining to him and:

(A) not later than 10 days (excluding Saturdays, Sundays and legal public holidays) after the date of receipt of such request, acknowledge in writing such receipt; and

(B) promptly, either:

(i) make any correction of any portion thereof which the individual believes is not accurate, relevant, timely, or complete; or

(ii) inform the individual of its refusal to amend the record in accordance with his request, the reason for the refusal, the procedures established by the agency for the individual to request a review of that refusal by the head of the agency, and the name and business address of that official;

(3) permit the individual who disagrees with the refusal of the agency to amend his record to request a review of such refusal, and not later than 30 days (excluding Saturdays, Sundays, and legal holidays) from the date on which the individual requests such review, complete such review and make a final determination unless, for good cause shown, the head of the agency extends such 30-day period; and if, after his review, the reviewing official also refuses to amend the record in accordance with the request, permit the individual to file with the agency a concise statement setting forth the reasons for his disagreement with the refusal of the agency, and notify the individual of the provisions for judicial review of the reviewing official's determination under subsection (g)(1)(A) of this section;

(4) in any disclosure, containing information about which the individual has filed a statement of disagreement, occurring after the filing of the statement under paragraph (3) of this subsection, clearly note any portion of the record which is disputed and provide copies of the statement and, if the agency deems it appropriate, copies of a concise statement of reasons of the agency for not making the amendments requested, to persons or other agencies to whom the disputed record has been disclosed; and

(5) nothing in this section shall allow an individual access to any information compiled in reasonable anticipation of a civil action or proceeding.

**(e) Agency requirements**

Each agency that maintains a system of records shall:

(1) maintain in its records only such information about an individual as is relevant and necessary to accomplish a purpose of the agency required to be accomplished by statute or by executive order of the President;

(2) collect information to the greatest extent practicable directly from the subject individual when the information may result in adverse determinations about an individual's rights, benefits, and privileges under Federal programs;

(3) inform each individual whom it asks to supply information, on the form which it uses to collect the information or a separate form that can be retained by the individual:

(A) the authority (whether granted by statute, or by executive order of the President) which authorizes the solicitation of the information and whether disclosure of such information is mandatory or voluntary;

(B) the principal purpose for which the information is intended to be used;

(C) the routine uses which may be made of the information, as published pursuant to paragraph (4) (D) of this subsection; and

(D) the effects on him, if any, of not providing all or any part of the requested information;

(4) subject to the provisions of paragraph (1) of this subsection, publish in the Federal Register upon establishment or revision a notice of the existence and character of the system of records, which notice shall include:

(A) the name and location of the system;

(B) the categories of individuals on whom records are maintained in the system;

(C) the categories of records maintained in the system;

(D) each routine use of the records contained in the system, including the categories of users and purpose of such use;

(E) the policies and practices of the agency regarding storage, retrievability, access controls, retention, and disposal of the records;

(F) the title and business address of the agency official who is responsible for the system of records;

(G) the agency procedures whereby an individual can be notified at his request if the system of records contains a record pertaining to him;

(H) the agency procedures whereby an individual can be notified at his request how he can gain access to any record pertaining to him contained in the system of records, and how he can contest its contents; and

(I) the categories of sources of records in the system;

(5) maintain all records which are used by an agency in making any determination about any individual with such accuracy, relevance, timeliness, and completeness as is reasonably necessary to assure fairness to the individual in the determination;

(6) prior to disseminating any record about an individual to any person other than an agency, unless the dissemination is made pursuant to subsection (b)(2) of this section, make reasonable efforts to assure that such records are accurate, complete, timely, and relevant for agency purposes;

(7) maintain no record describing how any individual exercises rights guaranteed by the First Amendment unless expressly authorized by statute or by the individual about whom the record is maintained or unless pertinent to and within the scope of an authorized law enforcement activity;

(8) make reasonable efforts to serve notice on an individual when any record on such individual is made available to any person under compulsory legal process when such process becomes a matter of public record;

(9) establish rules of conduct for persons involved in the design, development, operation, or

maintenance of any system of records, or in maintaining any record, and instruct each such person with respect to such rules and the requirements of this section, including any other rules and procedures adopted pursuant to this section and the penalties for noncompliance;

(10) establish appropriate administrative, technical and physical safeguards to insure the security and confidentiality of records and to protect against any anticipated threats or hazards to their security or integrity which could result in substantial harm, embarrassment, inconvenience, or unfairness to any individual on whom information is maintained;

(11) at least 30 days prior to publication of information under paragraph (4)(D) of this subsection, publish in the Federal Register notice of any new use or intended use of the information in the system, and provide an opportunity for interested persons to submit written data, view, or arguments to the agency; and

(12) if such agency is a recipient agency or a source agency in a matching program with a non-Federal agency, with respect to any establishment or revision of a matching program, at least 30 days prior to conducting such a program, publish in the Federal Register notice of such establishment or revision.

**(f) Agency rules**

In order to carry out the provisions of this section, each agency that maintains a system of records shall promulgate rules, in accordance with the requirements (including general notice) of section 553 of this title, which shall:

(1) establish procedures whereby an individual can be notified in response to his request if any system of records named by the individual contains a record pertaining to him;

(2) define reasonable times, places, and requirements for identifying an individual who requests his record or information pertaining to him before the agency shall make the record or information available to the individual;

(3) establish procedures for the disclosure to an individual upon his request of his record or information pertaining to him, including special procedures, if deemed necessary, for the disclosure to an individual of medical records, including psychological records pertaining to him;

(4) establish procedures for reviewing a request from an individual concerning the amendment of any record or information pertaining to the individual, for making a determination on the request, for an appeal within the agency of an initial adverse agency determination, and for whatever additional means may be necessary for each individual to be able to exercise fully his rights under this section; and

(5) establish fees to be charged, if any, to any individual for making copies of his record, excluding the cost of any search for and review of the record.

The Office of the Federal Register shall biennially compile and publish the rules promulgated under this subsection and agency notices published under subsection (e)(4) of this section in a form available to the public at low cost.

**(g) Civil remedies**

(1) Whenever any agency:

(A) makes a determination under subsection (d)(3) of this section not to amend an individual's record in accordance with his request, or fails to make such review in conformity with that subsection;

(B) refuses to comply with an individual request under subsection (d)(1) of this section;

(C) fails to maintain any record concerning any individual with such accuracy, relevance, timeliness, and completeness as in necessary to assure fairness in any determination relating to the qualifications, character, rights, or opportunities of, or benefits to the individual that may be made on the basis of such record, and consequently a determination is made which is adverse to the individual; or

(D) fails to comply with any other provisions of this section, or any other rule promulgated thereunder, in such a way as to have an adverse effect on an individual,

the individual may bring a civil action against the agency, and the district courts of the United States shall have jurisdiction in the matters under the provisions of this subsection.

(2)(A) In any suit brought under the provisions of subsection (g)(1)(A) of this section, the court may order the agency to amend the individual's record in accordance with his request or in such other way as the court may direct. In such a case the court shall determine the matter de novo.

(B) The court may assess against the United States reasonable attorney fees and other litigation costs reasonably incurred in any case under this paragraph in which the complainant has substantially prevailed.

(3)(A) In any suit brought under the provisions of subsection (g)(1)(B) of this section, the court may enjoin the agency from withholding the records and order the production to the complainant of any agency records improperly withheld from him. In such a case the court shall determine the matter de novo, and may examine the contents of any agency records in camera to determine whether the records or any portion thereof may be withheld under any of the exemptions set for in subsection (k) of this section, and the burden is on the agency to sustain its action.

(4) In any suit brought under the provisions of subsection (g)(1)(C) or (D) of this section in which the court determines that the agency acted in a manner which was intentional or willful, the United States shall be liable to the individual in an amount equal to the sum of:

(A) actual damages sustained by the individual as a result of the refusal or failure, but in no case shall a person entitled to recovery receive less than the sum of $1,000; and

(B) the costs of the action together with reasonable attorney fees as determined by the court.

(5) An action to enforce any liability created under this section may be brought in the district court of the United States in the district in which the complainant resides, or has his principal place of business, or in which the agency records are situated, or in the District of Columbia, without regard to the amount in controversy, within two years from the date on which the cause of action arises, except that where an agency has materially and willfully misrepresented any information required under this section to be disclosed to an individual and the information so misrepresented is material to

establishment of the liability of the agency to the individual under this section, the action may be brought at any time within two years after discovery by the individual of the misrepresentation. Nothing in this section shall be construed to authorize any civil action by reason of any injury sustained as the result of a disclosure of a record prior to September 27, 1975.

**(h) Rights of legal guardians**

For purposes of this section, the parent of any minor, or the legal guardian of any individual who has been declared incompetent due to physical or mental incapacity or age by a court of competent jurisdiction, may act on behalf of the individual.

**(i) Criminal penalties**

(1) Any officer or employee of an agency, who by virtue of his employment or official position, has possession of, or access to, agency records which contain individually identifiable information the disclosure of which is prohibited by this section or by rules or regulations established thereunder, and who knowing that disclosure of the specific material is prohibited, willfully discloses the material in any manner to any person or agency not entitled to receive it, shall be guilty of a misdemeanor and fined not more than $5,000.

(2) Any officer or employee of any agency who willfully maintains a system of records without meeting the notice requirements of subsection (e)(4) of this section shall be guilty of a misdemeanor and fined not more than $5,000.

(3) Any person who knowingly and willfully requests or obtains any record concerning an individual from an agency under false pretenses shall be guilty of a misdemeanor and fined not more than $5,000.

**(j) General exemptions**

The head of any agency may promulgate rules, in accordance with the requirements (including general notice) of sections 553(b)(1), (2), and (3), (c) and (e) of this title, to exempt any system of records within the agency from any part of this section, except, subsections (b), (c)(1) and (2), (e)(4) (A) through (F), (e)(6), (7), (9), (10) and (11), and (i) if the system of records is:

(1) maintained by the Central Intelligence Agency; or

(2) maintained by an agency or component thereof which performs as its principal function any activity pertaining to the enforcement of criminal laws, including police efforts to prevent, control, or reduce crime or to apprehend criminals, and the activities of prosecutors, courts, correctional, probation, pardon, or parole authorities, and which consists of

(A) information compiled for the purpose of identifying individual criminal offenders and alleged offenders and consisting only of identifying data and notations of arrests, the nature and disposition of criminal charges, sentencing, confinement, release, and parole and probation status;

(B) information compiled for the purpose of a criminal investigation, including reports of informants and investigators, and associated with an identifiable individual; or

(C) reports identifiable to an individual compiled at any stage of the process of enforcement of the criminal laws from arrest or indictment through release from supervision.

At the time rules are adopted under this subsection, the agency shall include in the statement required under section 553(c) of this title, the reasons why the system of records is to be exempted from a provision of this section.

**(k) Specific exemptions**

The head of any agency may promulgate rules, in accordance with the requirements (including general notice) of sections 553b(1), (2), and (3), (c) and (e) of this title, to exempt any system of records within the agency from subsections (c)(3), (d), (e)(1), (e)(4)(G), (H), and (I) and (f) of this section if the system of records is:

(1) subject to the provisions of section 552b(1) of this title;

(2) investigatory material compiled for law enforcement purposes, other than material within the scope of subsection (j)(2) of this section; Provided, however, That if any individual is denied any right, privilege, or benefit that he would otherwise be entitled to by Federal law, or for which he would otherwise be eligible, as a result of the maintenance of such material, such material shall be provided to such individual, except to the extent that the disclosure of such material would reveal the identity of a source who furnished information to the Government under an express promise that the identity of the source would be held in confidence, or, prior to the effective date of this section, under an implied promise that the identity of the source would be held in confidence;

(3) maintained in connection with providing protective services to the President of the United States or other individuals pursuant to section 3058 of Title 18;

(4) required by statute to be maintained and used solely as statistical records;

(5) investigatory material compiled solely for the purpose of determining suitability, eligibility, or qualifications for Federal civilian employment, military service, Federal contracts, or access to classified information, but only to the extent that the disclosure of such material would reveal the identity of a source who furnished information to the Government under an express promise that the identity of the source would be held in confidence, or, prior to the effective date of this section, under an implied promise that the identity of the source would be held in confidence;

(6) testing or examination material used solely to determine individual qualifications for appointment or promotion in the Federal service the disclosure of which would compromise the objectivity or fairness of the testing or examination process; or

(7) evaluation material used to determine potential for promotion in the armed services, but only to the extent that the disclosure of such material would reveal the identity of a source who furnished information to the Government under an express promise that the identity of the source would be held in confidence, or, prior to the effective date of this section, under an implied promise that the identity of the source would be held in confidence.

At the time rules are adopted under this subsection, the agency shall include in the statement required under section 553(c) of this title, the reasons why the system of records is to be exempted from a

provision of this section.

**(l) Archival records**

(1) Each agency record which is accepted by the Administrator of General Services for storage, processing, and servicing in accordance with section 3103 of Title 44 shall, for the purposes of this section, be considered to be maintained by the agency which deposited the record and shall be subject to the provisions of this section. The Administrator of General Services shall not disclose the record except to the agency which maintains the record, or under rules established by that agency which are not inconsistent with the provisions of this section.

(2) Each agency record pertaining to an identifiable individual which was transferred to the National Archives of the United States as a record which has sufficient historical or other value to warrant its continued preservation by the United States Government, prior to the effective date of this section, shall, for the purposes of this section, be considered to be maintained by the National Archives and shall not be subject to the provisions of this section, except that a statement generally describing such records (modeled after the requirements relating to records subject to subsections (e)(4)(A) through (G) of this section) shall be published in the Federal Register.

(3) Each agency record pertaining to an identifiable individual which is transferred to the National Archives of the United States as a record which has sufficient historical or other value to warrant its continued preservation by the United States Government, on, or after the effective date of this section, shall, for purposes of this section, be considered to be maintained by the National Archives and shall be exempt from the requirements of this section except subsections (e)(4)(A) through (G) and (e)(9) of this section.

**(m) Government contractors**

(1) When an agency provides by a contract for the operation by or on behalf of the agency of a system of records to accomplish an agency function, the agency shall, consistent with its authority, cause the requirements of this section to be applied to such system. For purposes of subsection (i) of this section any contractor and any employee of such contractor, if such contract is agreed to on or after the effective date of this section, shall be considered to be an employee of an agency.

(2) A consumer reporting agency to which a record is disclosed under 3711(f) of Title 31 shall not be considered a contractor for purposes of this section.

**(n) Mailing lists**

An individual's name and address may not be sold or rented by an agency unless such action is specifically authorized by law. This provision shall not be construed to require the withholding of names and addresses otherwise permitted to be made public.

**(o) Matching agreements**

(1) No record which is contained in a system of records may be disclosed to a recipient agency or non-Federal agency for use in a computer matching program except pursuant to a written agreement between the source agency and the recipient agency or non-Federal agency specifying --

(A) the purpose and legal authority for conducting the program;

(B) the justification for the program and the anticipated results, including a specific estimate of any savings;

(C) a description of the records that will be matched, including each data element that will be used, the approximate number of records that will be matched, and the projected starting and completion dates of the matching program;

(D) procedures for providing individualized notice at the time of application, and notice periodically thereafter as directed by the Data Integrity Board of such agency (subject to guidance provided by the Director of the Office of Management and Budget pursuant to subsection (v), to --

(i) applicants for and recipients of financial assistance or payments under Federal benefit programs, and

(ii) applicants for and holders of positions as Federal personnel, that any information provided by such applicants, recipients, holders, and individuals may be subject to verification through matching programs;

(E) procedures for verifying information produced in such matching programs as required under subsection (p);

(F) procedures for the retention and timely destruction of identifiable records created by a recipient agency or non-Federal agency in such matching program;

(G) procedures for ensuring the administrative, technical, and physical security of the records matched and the results of such programs;

(H) prohibitions on duplication and redisclosure of records provided by the source agency within or outside the recipient agency or the non-Federal agency, except where required by law or essential to the conduct of the matching program;

(I) procedures governing the use by a recipient agency or non-Federal agency of records provided in a matching program by a source agency, including procedures governing return of the records to the source agency or destruction of records used in such program;

(J) information on assessments that have been made on the accuracy of the records that will be used in such matching program; and

(K) that the Comptroller General may have access to all records of a recipient agency or non-Federal agency that the Comptroller General deems necessary in order to monitor or verify compliance with the agreement.

(2)(A) A copy of each agreement entered into pursuant to paragraph (1) shall --

(i) be transmitted to the Committee on Governmental Affairs of the Senate and the Committee on Government Operations of the House of Representatives; and

(ii) be available upon request to the public.

(B) No such agreement shall be effective until 30 days after the date on which such a copy is transmitted pursuant to subparagraph (A)(i);

(C) Such an agreement shall remain in effect only for such period, not to exceed 18 months, as the Data Integrity Board of the agency determines is appropriate in light of the purposes, and length of time necessary for the conduct, of the matching program;

(D) Within 3 months prior to the expiration of such an agreement pursuant to subparagraph (C), the Data Integrity Board of the agency may, without additional review, renew the matching agreement for a current, ongoing matching program for not more than one additional year if --

(i) such program will be conducted without any change, and

(ii) each party to the agreement certifies to the Board in writing that the program has been conducted in compliance with the agreement.

**(p) Verification and opportunity to contest findings**

(1) In order to protect any individual whose records are used in a matching program, no recipient agency, non-Federal agency, or source agency may suspend, terminate, reduce, or make a final denial of any financial assistance or payment under a Federal benefit program to such individual, or take other adverse action against such individual, as a result of information produced by such matching program, until --

(A)(i) the agency has independently verified the information; or

(ii) the Data Integrity Board of the agency, or in the case of a non-Federal agency the Data Integrity Board of the source agency, determines in accordance with guidance issued by the Director of the Office of Management and Budget that --

(I) the information is limited to identification and amount of benefits paid by the source agency under a Federal benefit program; and

(II) there is a high degree of confidence that the information provided to the recipient agency is accurate;

(B) the individual receives a notice from the agency containing a statement of its findings and informing the individual of the opportunity to contest such findings; and

(C)(i) the expiration of any time period established for the program by statute or regulation for the individual to respond to that notice, or

(ii) in the case of a program for which no such period is established, the end of the 30-day period beginning on the date on which notice under subparagraph (B) is mailed or otherwise provided to the individual.

(2) Independent verification referred to in paragraph (1) requires investigation and confirmation of

specific information relating to an individual that is used as a basis for an adverse action against the individual, including where applicable investigation and confirmation of --

(A) the amount of any asset or income involved;

(B) whether such individual actually has or had access to such asset or income for such individual's own use; and

(C) the period or periods when the individual actually had such asset or income.

(3) Notwithstanding paragraph (1), an agency may take any appropriate action otherwise prohibited by such paragraph, if the agency determines that the public health or public safety may be adversely affected or significantly threatened during the notice period required by such paragraph.

**(q) Sanctions**

(1) Notwithstanding any other provisions of law, no source agency may disclose any record which is contained in a system of records to a recipient agency or non-Federal agency for a matching program if such source agency has reason to believe that the requirements of subsection (p), or any matching agreement entered into pursuant to subsection (o), or both, are not being met by such recipient agency.

(2) No source agency may renew a matching agreement unless --

(A) the recipient agency or non-Federal agency has certified that it has complied with the provisions of that agreement, and

(B) the source agency has no reason to believe that the certification is inaccurate.

**(r) Report on new systems and matching programs**

Each agency that proposes to establish or make a significant change in a system of records or a matching program shall provide adequate advance notice of any such proposal (in duplicate) to the Committee on Government Operations of the House of Representatives, the Committee on Governmental Affairs of the Senate, and the Office of Management and Budget in order to permit an evaluation of the probable or potential effect of such proposal on the privacy or other rights of individuals.

**(s) Biennial report**

The President shall biennially submit to the Speaker of the House of Representatives and the President pro tempore of the Senate a report --

(1) describing the actions of the Director of Management and Budget pursuant to Section 6 of the Privacy Act of 1974 during the preceding 2 years;

(2) describing the exercise of individual rights of access and amendment under this section during such years;

(3) identifying changes in or additions to systems of records;

(4) containing other such information concerning administration of this section as may be necessary

or useful to the Congress in reviewing the effectiveness of this section in carrying out the purposes of the Privacy Act of 1974.

**(t) Effect of other laws**

Relationship of the Privacy Act to the Freedom of Information Act.

(1) No agency shall rely on any exemption contained in section 552 of this title to withhold from an individual any record which is otherwise accessible to such individual under the provisions of this section.

(2) No agency shall rely on any exemption in this section to withhold from an individual any record which is otherwise accessible to such individual under the provisions of section 552 of this title.

**(u) Data Integrity Boards**

(1) Every agency conducting or participating in a matching program shall establish a Data Integrity Board to oversee and coordinate among the various components of such agency the agency's implementation of this section.

(2) Each Data Integrity Board shall consist of senior officials designated by the head of the agency, and shall include any senior official designated by the head of the agency as responsible for implementation of this section, and the Inspector General of the agency, if any. The Inspector General shall not serve as the chairman of the Data Integrity Board.

(3) Each Data Integrity Board --

(A) Shall review, approve, and maintain all written agreements for receipt or disclosure of agency records for matching programs to ensure compliance with subsection (o), and all relevant statutes, regulations, and guidelines;

(B) shall review all matching programs in which the agency has participated during the year, either as a source agency or recipient agency, determine compliance with applicable laws, regulations, and agency agreements, and assess the cost and benefits of such programs;

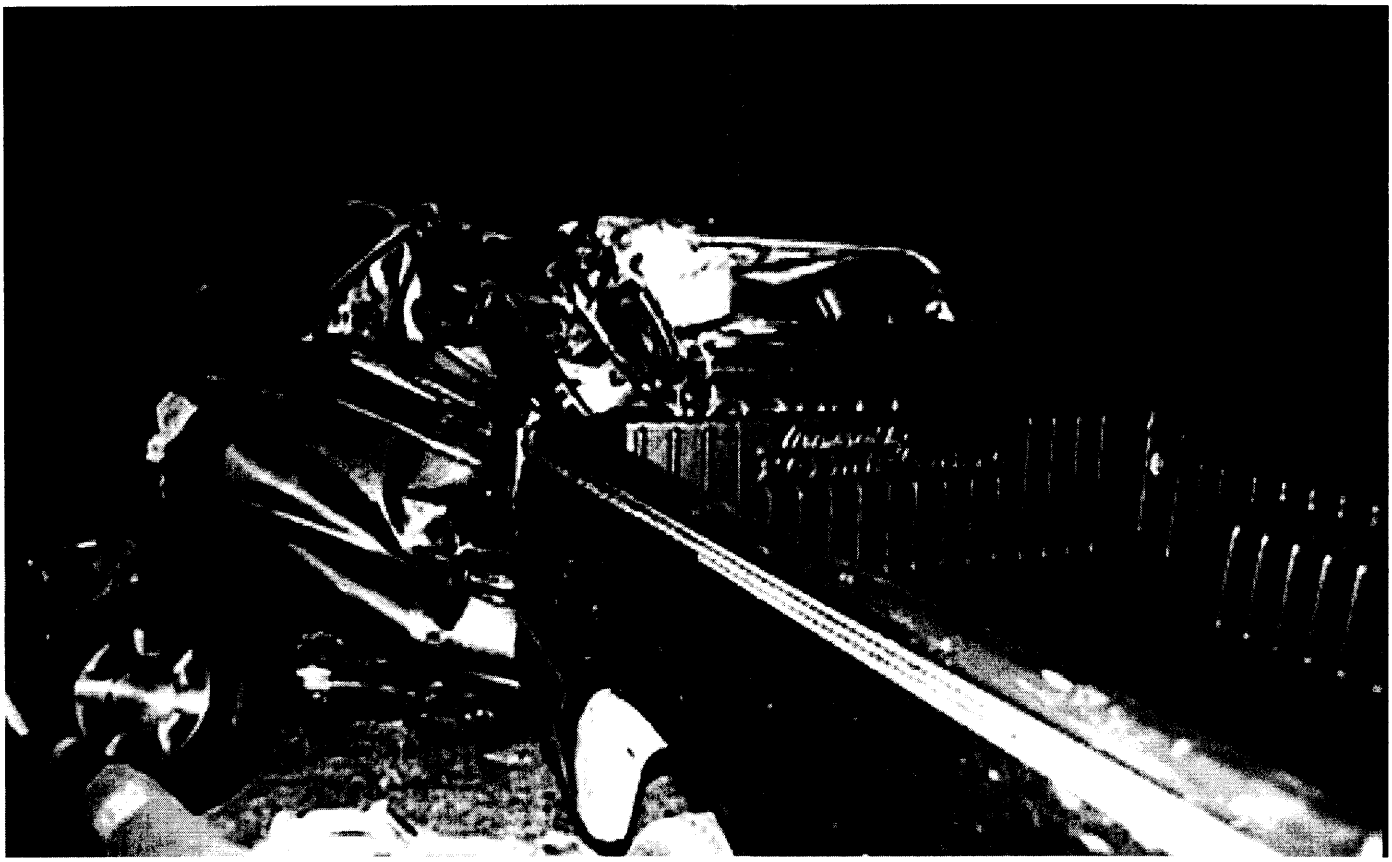
(C) shall review all recurring matching programs in which the agency has participated during the year, either as a source agency or recipient agency, for continued justification for such disclosures;

(D) shall compile an annual report, which shall be submitted to the head of the agency and the Office of Management and Budget and made available to the public on request, describing the matching activities of the agency, including --

(i) matching programs in which the agency has participated as a source agency or recipient agency;

(ii) matching agreements proposed under subsection (o) that were disapproved by the Board;

# Attachments



*56% of fatal crashes involved only one vehicle, compared to 28% of both injury crashes and property-damage-only crashes.*

Photographer: Frank Staples
© 1999 Click, Incorporated, Transportation Safety Technologies

**Rules of construction**

Nothing in the amendments made by the Computer Matching and Privacy Protection Act shall be construed to authorize --

(1) the establishment or maintenance by any agency of a national data bank that combines, merges, or links information on individuals maintained in systems of records maintained by other Federal agencies;

(2) the direct linking of computerized systems of records maintained by Federal agencies;

(3) the computer matching of records not otherwise authorized by law; or

(4) the disclosure of records for computer matching except to a Federal, State, or local agency.

(iii) the matching program is in the public interest.

(C) The decision of the Director to approve a matching agreement shall not take effect until 30 days after it is reported to committees described in subparagraph (A).

(D) If the Data Integrity Board and the Director of the Office of Management and Budget disapprove a matching program proposed by the Inspector General of an agency, the Inspector General may report the disapproval to the head of the agency and to Congress.

(6) The Director of the Office of Management and Budget shall, annually during the first 3 years after the date of enactment of this subsection and biennially thereafter, consolidate in a report to the Congress the information contained in the reports from the various Data Integrity Boards under paragraph (3)(D). Such report shall include detailed information about costs and benefits of matching programs that are conducted during the period covered by such consolidated report, and shall identify each waiver granted by a Data Integrity Board of the requirement for completion and submission of a cost-benefit analysis and the reasons for granting the waiver.

(7) In the reports required by paragraphs (3)(D) and (6), agency matching activities that are not matching programs may be reported on an aggregate basis, to the extent necessary to protect ongoing law enforcement or counterintelligence investigations.

**(v) Office of Management and Budget responsibilities**

The Director of the Office of Management and Budget shall --

(1) develop and, after notice and opportunity for public comment, prescribe guidelines and regulations for the use of agencies in implementing the provisions of this section; and

(2) provide continuing assistance to and oversight of the implementation of this section by agencies.

**Section 7**

(a)(1) It shall be unlawful for any Federal, State, or local government agency to deny to any individual any right, benefit, or privilege provided by law because of such individual's refusal to disclose his social security account number.

(2) the provisions of paragraph (1) of this subsection shall not apply with respect to --

(A) any disclosure which is required by Federal statute, or

(B) the disclosure of a social security number to any Federal, State or local agency maintaining a system of records in existence and operating before January 1, 1975, if such disclosure was required under statute or regulation adopted prior to such date to verify the identity of an individual.

(b) Any Federal, State or local government agency which requests an individual to disclose his social security account number shall inform that individual whether that disclosure is mandatory or voluntary, by what statutory or other authority such number is solicited, and what uses will be made of it.

(iv) the reasons for any waiver of the requirement in paragraph (4) of this section for completion and submission of a cost-benefit analysis prior to the approval of a matching program;

(v) any violations of matching agreements that have been alleged or identified and any corrective action taken; and

(vi) any other information required by the Director of the Office of Management and Budget to be included in such report;

(E) shall serve as a clearinghouse for receiving and providing information on the accuracy, completeness, and reliability of records used in matching programs;

(F) shall provide interpretation and guidance to agency components and personnel on the requirements of this section for matching programs;

(G) shall review agency recordkeeping and disposal policies and practices for matching programs to assure compliance with this section; and

(H) may review and report on any agency matching activities that are not matching programs.

(4)(A) Except as provided in subparagraphs (B) and (C), a Data Integrity Board shall not approve any written agreement for a matching program unless the agency has completed and submitted to such Board a cost-benefit analysis of the proposed program and such analysis demonstrates that the program is likely to be cost effective.

(B) The Board may waive the requirements of subparagraph (A) of this paragraph if it determines in writing, in accordance with guidelines prescribed by the Director of the Office of Management and Budget, that a cost-benefit analysis is not required.

(C) A cost-benefit analysis shall not be required under subparagraph (A) prior to the initial approval of a written agreement for a matching program that is specifically required by statute. Any subsequent written agreement for such a program shall not be approved by the Data Integrity Board unless the agency has submitted a cost-benefit analysis of the program under the preceding approval of such agreement.

(5)(A) If a matching agreement is disapproved by a Data Integrity Board, any party to such an agreement may appeal the disapproval to the Director of the Office of Management and Budget. Timely notice of the filing of such an appeal shall be provided by the Director of the Office of Management and Budget to the Committee on Governmental Affairs of the Senate and the Committee on Government Operations of the House of Representatives.

(B) The Director of the Office of Management and Budget may approve a matching agreement notwithstanding the disapproval of a Data Integrity Board in the Director determines that --

(i) the matching program will be consistent with all applicable legal, regulatory, and policy requirements;

(ii) there is adequate evidence that the matching agreement will be cost-effective; and

# GM installs 'black box' on its autos

**Device offers potential for safety gains, privacy disputes**

ASSOCIATED PRESS

WASHINGTON, June 1 — General Motors Corp. has a device in many of its new cars that functions like the black box recorder in airplanes: It collects data as a car crashes. Doctors and government officials say that information can help them better understand how the human body tolerates car crashes. It could then be applied to construct safer cars, improve the treatment of crash victims and write government auto safety standards that would better protect crash victims.

COMPLETE STORY

TECHNOLOGY BBS: WHAT'S YOUR OPINION?
BBS Weigh in on the "black box"

ADVERTISING ON MSNBC

The data includes

THE EXISTENCE OF THE SO-CALLED auto black box system also is raising sensitive privacy

- the speed of the car, whether the driver was wearing a seat belt, when an air bag deployed and whether the driver used the brakes.

questions about whether such information can be used in litigation.

The most sophisticated version of GM's device, known formally as a sensing and diagnostic module, is in hundreds of thousands of GM cars from the 1999 model year, GM says. It is part of the air bag sensing system on the 1999 Buick Century, Park Avenue and Regal, the Cadillac Eldorado, DeVille and SeVille, the Chevrolet Camaro and Corvette and the Pontiac Firebird.

The module will be in almost all GM vehicles within the next few years, the company says.

The module stores information in the instants before a car sensor identifies a crash and fires the air bags. The data includes the speed of the car, whether the driver was wearing a seat belt, when an air bag deployed and whether the driver used the brakes. It can also determine whether a warning light was illuminated on the dashboard telling an owner to service an air bag.

GM has quietly installed different versions of the sensing system on some cars throughout the 1990s, but the modules have become more sophisticated over time. Their existence became public in a paper written by GM and government engineers and presented at a conference last month.

Up until now, government crash investigators could only take an educated guess at the speed of a car involved in an accident based on evidence at the crash scene.

"Technology allowing vehicle safety researchers to collect objective data would open the door to a new generation of understanding," the paper said.

GM is currently the only automaker that makes such data and the tools to recover it available to researchers, the paper said.

Bob Lange, director of engineering safety for GM, said he wanted to use the information to better understand the injuries of people of all ages in crashes so that autos could be designed to "reduce the likelihood of injuries."

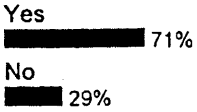
GM has been using the technology on Indy race cars since 1992 and it has led to better crash protection for drivers, Lange said.

"There's an incredible opportunity to improve safety," said Dr. Jeffrey Augenstein of the Crash Injury Research and Engineering Network. Augenstein said if doctors know more about crashes, they can target their treatment of patients, in some cases including checks for serious injuries they might have missed.

John Hinch, a research engineer at the National Highway Traffic Safety Administration and one of the authors of the paper, said he saw "lots of potential" in using the module's data. GM hopes to have laptops available so government crash investigators can download data independently of the company by the end of the year.

**Is the automobile "black box" a good idea?**

• 2996 responses




MSNBCINTERACTIVE

Survey results tallied every 60 seconds. Live Votes reflect respondents' views and are not scientifically valid surveys.

"If we can understand crashes better, we can have better sensors (in automobiles), better air bags," Hinch said. "NHTSA can build better (safety) rules and have better information for consumers."

## Problems, solutions, products

- **Full coverage:** The millennium bug
- **Y2K Toolkit:** Downloads, resources, tests, more
- **Bug of the Day:** Get the latest glitch, plus a month of alerts and three months of fixes
- **Tools and Toys:** Products for productivity and fun
- **Tech BBS:** Discuss technology issues on the Technology Bulletin Board
- **Enter the ZDNet Zone:** Reviews, downloads, tips and more from ZDNet

Insurers also seem to favor so-called black boxes for cars, in part because it would help them determine who is at fault in accidents. But they say courts will first have to sort through how such devices could be used in litigation and whether they are reliable if contradicted by eyewitness accounts.

Norman Jolly, an attorney who has litigated auto cases, said he has already seen auto companies try to use air bag deployment information stored on a car computer chip as a defense in lawsuits.

He believes companies will not be able to keep such information private. "They're going to know if your case has merit, and vice versa," Jolly said.

Ford Motor Co. said a more limited version of the module was on all its 1999 vehicles, but the company is unable to retrieve the data for customers.

*© 1999 Associated Press. All rights reserved. material may not be published, broadcast, rewritten or redistributed.*

MSNBC VIEWER'S TOP 10

Would you recommend this story to other viewers?

not at all 1 - 2 - 3 - 4 - 5 - 6 - 7 highly

BACK TO TOP

MSNBC is optimized for
- Microsoft Internet Explorer
- Windows Media Player

- MSNBC Terms, Conditions and Privacy © 1999

NON-PERSONAL IDENTIFIERS

SEX _____

AGE _____

YRS
DRIVING _____

# OF CRASHES
AS OCCUPANT _____

# OF CRASHES
AS DRIVER _____

PERCENTAGE
OF SEAT BELT
USAGE _____

WAS SEATBELT
USED DURING
CRASH? _____

WERE YOU
INJURED BY BELT? _____

WERE YOU INJURED BY
AIRBAG? _____

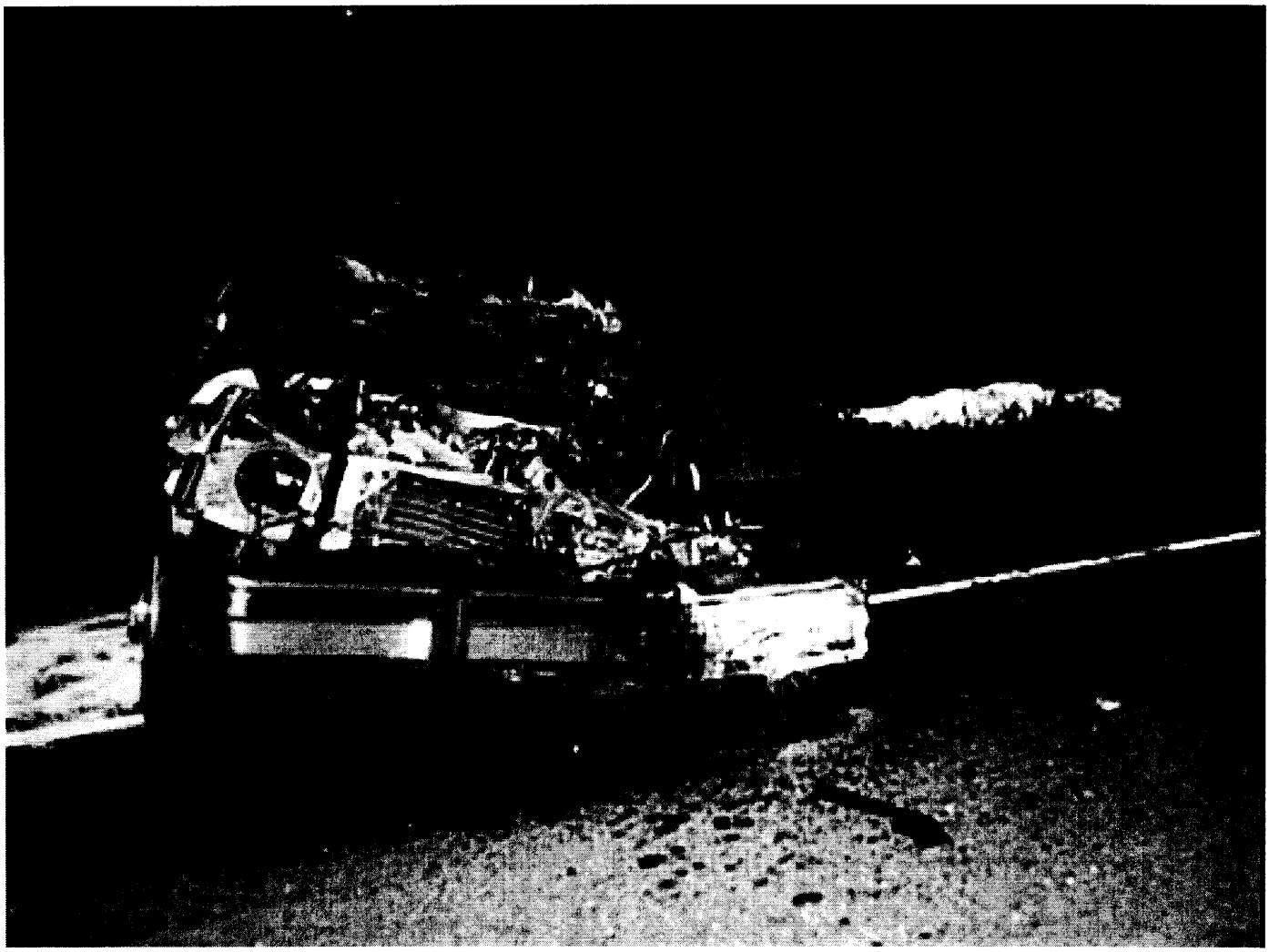
DO YOU TRUST
AIRBAG? _____

COUNTY _____

## TRANSPORTATION RECORDER PERCEPTIONS

| POSITIVE | NEGATIVE |
|---|---|
| 1) _____ | 1) _____ |
| 2) _____ | 2) _____ |
| 3) _____ | 3) _____ |
| 4) _____ | 4) _____ |
| 5) _____ | 5) _____ |
| 6) _____ | 6) _____ |
| 7) _____ | 7) _____ |
| 8) _____ | 8) _____ |
| 9) _____ | 9) _____ |
| 10) _____ | 10) _____ |

# Resources



*Collision with another motor vehicle in transport was the most common first harmful event for fatal, injury, and property-damage-only crashes.*

Photographer: Frank Staples
© 1999 Click, Incorporated, Transportation Safety Technologies

1. American Civil Liberties Union (ACLU). (1999). Washington, DC. Legislative Briefing Kit on Electronic Monitoring at http://www.aclu.org/library/pbr2.html

2. ____. ACLU White Paper: Big Brother in the Wires; Wiretapping in the Digital Age. March 1998.

3. Blincoe, L.J. 1996. The Economic Cost of Motor Vehicle Crashes. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

4. Campbell, B.J., et al (1986). Seat Belt Law Experience in Four Foreign Countries Compared to the United States (Falls Church, Va.: University of North Carolina: AAA Foundation for Traffic Safety).

5. Clinton, William J (1997). Executive Order 13043 - Increasing Seat Belt Use in the United States, Presidential Documents, FR Doc. 97-10331 Filed 4-17-97, The White House, Washington, DC.

6. C. R. Drane and C. A. Scott, "Positioning Systems and Privacy," 2nd International Congress on Intelligent Transport Systems Yokohama, Japan, November 1995.

7. Volvo Technical Report: (1997) DARR- Applying Digital Technology to Accident Research at http://www.volvo.se/rt/trmag/vtr972/article2/a2no297.html

8. Evans, L., Traffic Safety and the Driver. New York, NY: Van Nostrand Reinhold, 1991.

9. Background Information About Crash Data at http://www.nhtsa.gov/people/ncsa/codes/Mindata/Background.html

10. Federal Register, Volume 63, No. 181 / / Proposed Rules. Pg. 49958. Washington, DC. Friday, September 18, 1998.

11. Highway Occupant Protection at http://www.ntsb.gov/Recs/highway_occupant_protection.htm#SafetyIssue

12. Harris, Louis., "A Survey of the Attitudes of the American People on Highway and Auto Safety," April 1998.

13. Insurance Institute for Highway Safety (IIHS). Q & A: Speed Limits. 1998. Available at http://www.hwysafety.org/

14. ITS America. 1999. Washington, DC. ITS Fair Information and Privacy Principles at http://www.itsa.org/

15. ____. Archived Data User Server Addendum to the ITS Program Plan. September 1999.

16. ____. The National Intelligent Transportation Systems Deployment Strategy. March 1999.

17. Kullgren, Andrews., Validity and Reliability of Vehicle Collision Data: Crash Pulse Recorders for Impact Severity and Injury Risk Assessments in Real-Life Frontal Impacts, Thesis for Doctor in Medical Science, Karolinska Institute, Stockholm, Sweden, 1998.

18. National Highway Traffic Safety Administration web page at http://www.nhtsa.gov/

19. National Information Infrastructure: An Overview at http://www.brint.com/papers/nii/overview.htm

20. National Transportation Safety Board web page at http://www.ntsb.gov/

21. National Safety Council. (1998). Accident Facts, 1998 Edition. Itasca, IL NSC web page at http://www.nsc.org/

22. Phen, Dowdy, Ebbeler, Kim, Moore, and Van Zandt; Advanced Air Bag Technology Assessment; JPL Publication 98-3; April 1998. The report can be found on the JPL web site - http://csmt.jpl.nasa.gov/airbag/contents.html

23. Office of Technology Assessment. Washington, DC. 1975. Automobile Collision Data: An Assessment of Needs and Methods of Acquisition.

24. Office of Technology Assessment. Washington, DC. 1982. Mandatory Passive Restraint Systems in Automobiles: Issues and Evidence.

25. Office of Technology Assessment. Washington, DC. 1986. Electronic Record Systems and Individual Privacy

26. Office of Technology Assessment. Washington, DC. 1995. Electronic Surveillance in a Digital Age

27. Office of Technology Assessment. Washington, DC. 1993. Advanced Network Technology.

28. Office of Technology Assessment. Washington, DC. 1993. Data Format Standards for Civilian Remote Sensing Satellites.

29. Office of Technology Assessment. Washington, DC. 1995. Advanced Automotive Technology: Visions of a Super-Efficient Family Car.

30. Office of Technology Assessment. Washington, DC. 1995. Wireless Technologies and the National Information Infrastructure.

31. Presidential Initiative for Increasing Seat Belt Usage (1997) at http://www.nhtsa.dot.gov/people/injury/airbags/presbelt/

32. Society of Automotive Engineers. (1999) Intelligent Vehicle Technologies at http://www.sae.org/

33. Telestrategies, Inc. (1996). The Emerging Market for Wireless Location-Based Products and Services (Washington, D.C.: Telestrategies, Inc.).

34. Transportation Statistics Annual Report 1998. Washington, DC.

35. UR WWW: Wire Report 1212e: Claim Fraud in One-Third of Auto BI Claims at http://www.uwreport.com/wire/news1296/1212aut.htm

36. U.S. Department of Transportation (USDOT), 1997. Strategic Plan 1997-2002. 30 September.

37. U.S. Department of Transportation (USDOT), Bureau of Transportation Statistics (BTS). 1998. National Transportation Statistics 1998. Available at http://www.bts.gov/ntda/nts/

38. U.S. Department of Transportation (USDOT), National Highway Traffic Safety Administration (NHTSA). 1997. Traffic Safety Facts 1996. Washington, DC.

39. ____. Report to the Congress on the National Highway Traffic Safety Administration ITS Program: Program Progress During 1992-1996 and Strategic Plan for 1997-2002. January 1997.

40. ____. Motor Vehicle Crashes as a Leading Cause of Death in the U.S., 1994, Technical Report DOT H5 808 687. Washington, DC.

41. ____. Intelligent Vehicle Initiative Business Plan at http://www.its.dot.gov/ivi/bp114b.htm

42. ____. Federal Motor Vehicle Safety Standards; Occupant Crash Protection, 49 CFR Part 571, Docket No. 74-14; Notice 113 at http://www.nhtsa.dot.gov/cars/rules/rulings/OccCProt.html

43. ____. Traffic Safety Facts 1997. Washington, DC.

44. U.S. Government Printing Office. Washington, DC. The Privacy Act of 1974, Government in the Sunshine Act, Federal Advisory Committee Act, Freedom of Information Act (FOIA), and Drivers Privacy Protection Act.

# Photo Credits

*All states, the District of Columbia, and Puerto Rico have laws requiring children of certain ages to be restrained in child safety seats.*

Photographer: Frank Staples
© 1999 Click, Incorporated, Transportation Safety Technologies

Frank Staples is an amateur photographer residing in Southern Pines, Moore County, North Carolina. He is an independent small business owner and a volunteer fire fighter.

For more than a decade, Frank has been taking pictures of crashes and fires. His photos are often published in the local newspaper, *The Pilot.*

The photos included in this research project were taken within a thirty-mile radius of Southern Pines over a period of ten years. Frank graciously contributes them (as an advocate of highway safety) and permits their publication and distribution solely within the guidelines of this research project.

# Conclusion

*Nearly 39% of the persons who were killed in traffic crashes in 1997 died in alcohol-related crashes.*

Photographer: Frank Staples
© 1999 Click, Incorporated, Transportation Safety Technologies

Approximately one hundred people were involved in this research project. Throughout the research, it has been my objective to create a dialogue in an area usually lacking consensus – automobile safety. Dialogue is two or more people talking about the same issue in a rational manner. This is not as simple as it sounds. At times, our discussions were heated, with zealots on opposing sides, each trying to persuade one another of the truth, logic, and virtue of their views and the falsity, irrationality, and downright evil of the others'. Many students were experts at employing vague, emotion-laden terms. Few had the ability to define these terms when challenged. However, once students understood that they would be held responsible for everything said in a public forum, they developed a conviction that they could and would learn from the other side. Obviously, this stopped none from clinging to their cherished viewpoints, but more than a few widened their views to include new knowledge. Their original thinking was enriched and this in turn enhanced their public understanding. By voicing themselves in a public forum – a community college classroom – they all gained from the experience.

Some of the vital issues and critical questions of the group concerned a general mistrust for the government, the automobile industry, and technological change. An even deeper concern was the fear that personal liberties and privacy would be encroached. One of the interesting footnotes to this research project was the discovery of a local amateur photographer who was willing to contribute pictures.

It was a challenging task to define the issue at the outset with enough specificity to make sure that both sides would be talking about the same area of controversy. This was in fact so difficult that after several failed attempts, it became clear that the group's most valuable contribution would be in citing the positives and negatives. Several students wanted to complete half the assignment and were dismayed when told they would receive partial credit. This forced everyone to see both sides and thus created the elusive "dialogue on automobile safety," which is perhaps the overall value of this research study.

Professor T. Kowalick, Southern Pines, North Carolina, August 1999