572

Stevens, Mark W.
Conference on Boundaries of
Privacy in American Society
Woodrow Wilson School of Public
and International Affairs
November 18, 1971

FEDERAL STORAGE OF PERSONAL INFORMATION

I. STATEMENT OF THE PROBLEM

The problem of this paper is to investigate the Federal government's procedures for storing personal data, to determine actual and potential threats to privacy in these procedures, and to devise means to prevent the abuse of this stored information.

II. SUMMARY OF PRINCIPAL FINDINGS

1. Computers have changed and will continue to change the nature of information storage, probably in the direction of increased integration and centralization.

2. Individual agencies have wide lattitude in determining confidentiality standards and procedures, and in practice they vary considerably. Federal law has not succeeded in providing a uniform answer to the question of who has access to what information.

3. Federal laws encourage agencies to transfer information to each other. Laws protecting privacy in the transferral of information in effect provide little protection because of loopholes.

4. There are no government-wide laws encouraging the maintenance of accuracy in files, though a great deal of innacuracy exists. Granting each individual access to information about him is prohibitively expensive, but a limited access system is possible.

5. Laws governing obsolete information tend to prevent the destruction of records and contain no privacy oriented provisions.

6. A de facto data bank may develop without adequate supervision and threaten privacy.

7. A small, controlled, statistical data system offers a means to experiment with technological and legal safeguards in a centralized system.

8. Because of the difficulty of building legal controls into the Federal organization of information storage, controls over what may be put into the system provide the best protection for privacy.

9. No Federal definition of privacy exists with which to devise standards of access, tranferral of information and obsolescence.

575

III. POLICY RECOMMENDATIONS

1. Congress should pass a Constitutional amendment defining privacy as a fundamental right of all Americans.

A detailed definition seems unworkable. (pp. 18-19). On the other hand, an amendment, by tying in the courts, could go a long way in forcing agencies to decide whether or not the input of certain information is Constitutional and thereby indirectly guard against unwarranted access or transferral. (pp. 18-19, 26). An amendment would also place the burden on the government to prove that presently held information may be released or transferred. (pp. 18-19, 26).

2. Congress should pass legislation requiring agencies to permit access to or transferral of personal information only in aggregate form wherever possible except in national security cases, when transferral of data is necessary.

(See pp. 19a, 26)

3. Congress should establish an independent watchdog bureau, subject to Congress, which would permit release of individually identifiable data only after examining the privacy implications of the release, notifying the individual and receiving his consent. This bureau would also serve as an ombudsman, have the power to deny transferral of personal data, and rate the lifespan of personal data.

(See pp. 19a, 26, 31)

4. Congress should establish the right of an individual to examine his own file, with certain limitations.

(See pp. 27-27a)

5. Congress should appoint a Commission to analyze centralization trends, technological safeguards and the potential effect of the computer on Federal information storage with a view toward establishing a statistical national data

center on an experimental basis.

(See pp. 4-13, 32-35)

IV. TABLE OF CONTENTS

		Page
I.	STATEMENT OF THE PROBLEM	ii
II.	SUMMARY OF PRINCIPAL FINDINGS	iii
III.	POLICY RECOMMENDATIONS	iv
V.	DISCUSSION	1
	Computer Technology: The New Environment	1
	The National Data Bank Proposal	3
	Arguments for the center	4
	Arguments against a Data Center	10
	Access to Federal Information	13
	Present law and practice	13
	The Freedom of Information Act	16
	A potential solution to the information standards problem	18
	Transfer of Information Among Agencies	19
	Present law and practice	19
	Some possible non-solutions	24
	The Problem of Accuracy	26
	The Problem of Obsolete Information	29
	The Development of a De Facto National Data Bank	32
	Another look at the National Data Center	33
VI.	BIBLIOGRAPHY	36

V. DISCUSSION

COMPUTER TECHNOLOGY: THE NEW ENVIRONMENT

For man to "know himself and to make more rational, more predictable decisions about human affairs," vast amounts of information must be collected and correctly analyzed.[1] Until the advent of the computer, information collecting was spotty at best--decisions were made as much on hunches and feelings as facts. Now, however, the means to compare and store large amounts of information are available and computer sales, consequently, have skyrocketed. RCA, for example, has reported that there were 30,000 computers in the United States in 1967 and predicts 85,000 by 1975 and 150,000 by 1985.[2]

This growth of the computer industry complements the growth of government in the United States. As government has undertaken more and more social policies, its need for infor-

---

1 Alan Westin, quoted in 35th Report by the Committee on Government Operations, Privacy and the National Data Bank Concept, H.R. 1842, U.S. 90th Congress, 2nd Session, August 2, 1968, p. 11.

2 Glenn Seaborg, "Time, Leisure and the Computer: The Crisis of Modern Technology," speech reprinted for the Senate Committee on the Judiciary, March 14, 15, 1967, in U.S. 90th Congress, 1st Session, Subcommittee on Administrative Practice and Procedure, Computer Privacy: Appendix to Hearings, Part 1, p. 249. (These Senate Hearings are hereafter referred to as 'Senate Hearings.')

mation has increased. Conversely, as its information pool has increased, the government's potential for effective social action has expanded. Computers have made all this possible. Not only have these machines made information storage larger, more efficient, and more economical, they have also permitted man to develop programs which sort, collate and analyze information. The time of the manilla folder is ~~is~~ fast disappearing. Today, almost every agency in the Federal government uses computers to store and analyze data.

If there was an obvious danger to privacy in the old manilla folder methods of storing information, there is potentially a much greater danger to privacy in computerized storage, as well as some important, potential guarantees and safeguards. The problems of access and disclosure of information, of inaccuracies and leaks, of information transferral and obsolescence are still there, but the dimension of the problem is different now because a new variable has been added. Whereas 20 years ago, the threat to privacy came only from laws and people, today the threat comes almost as much from the potential of the machine.

> Our courts punish only individuals caught in the act of illegal behavior. The problem here is that our technological environment may make it difficult or impossible to apprehend certain types of criminals.[1]

Or, one might add, to protect privacy. Thus, in trying to

---

[1] Burton Squires, statement before House Committee on Government Operations, July 28, 1966, in U.S. 89th Congress, 2nd Session, Special Subcommittee on Invasion of Privacy, The Computer and Invasion of Privacy; Hearings, p. 138. (Hereafter referred to as House Hearings.)

come to grips with the danger to privacy in Federal storage practices, we will have to keep the computer continually in mind. Moreover, we must be careful not to think of solving problems once and for all, for computer technology is still in its infancy and provides a fluid environment and many options. As computers grow more and more sophisticated, so must our safeguards if we are to protect privacy.

THE NATIONAL DATA BANK PROPOSAL

Perhaps more than any one event, the National Data Bank proposal is responsible for today's heated interest in privacy. The proposal itself, however, arose very quietly, and without much public notice in 1965 when Richard Ruggles submitted a report to the Social Science Research Council. This report recommended the centralization of economic statistical data to aid policy making and scholarly research.[1] The Dunn report, commissioned by the Bureau of the Budget, followed in November of 1965, and also recommended the creation of a National Data Center in order to aid policy makers, scholars, and analysts in bringing together disparate pieces of information.[2]

During the compilation of the third and most important report--Carl Kaysen's Task Force on the data bank[3]--Repre-

---

[1] Report of the Committee on the Preservation and Use of Economic Data to the Social Science Research Council. Richard Ruggles, chairman, April, 1965. (available in Appendix to House Hearings, p. 195-253)

[2] U.S. Bureau of the Budget, Statistical Evaluation Report No. 6--Review of Proposal for a National Data Center, prepared by Edgar Dunn, Nov. 1965 (available in House Hearings, p. 254-84)

sentative Gallagher held hearings on the proposal.[1,2] The hearings aroused considerable public interest and press coverage when it was found that the Dunn and Ruggles reports hardly dealt with the privacy implications of the center at all. The Kaysen Task Force, caught off guard, added only a short Appendix to its report on the privacy implications.

It appears unlikely now that a data center will materialize in the near future. The arguments and response it generated, however, have crucial implications for the future protection of personal privacy. These arguments, for and against, provide us with a general framework of issues with which to examine federal storage of information.

Arguments for the center

The growth of the data bank concept was internally motivated, i.e., the idea sprung up in response to organizational problems within the government.[3] It was not by any means a proposal by the executive to allow the government to monitor the lives of citizens or keep a tighter watch over them. Rather, when the need for information increased as the government expanded, the need for statistics and correlations

---

1 U.S. Bureau of the Budget, Report of the Task Force on the Storage of and Access to Government Statistics, Carl Kaysen, chairman. October 1966. (Available in Senate Hearings, p. 25-46)

2 House Hearings, op. cit.

3 Carl Kaysen, interview, November 10, 1971, Princeton, New Jersey.

had also increased, as mentioned earlier. Long range samples, comparison of variables, and factually-based options had all become necessary. The present storage of government information, however, is highly decentralized; it is inefficient and uneconomical through duplication of effort and insufficient funding for small agencies. Much data cannot be utilized by more than one agency because of different and sometimes contradictory laws on confidentiality.[1] Furthermore, many individual agencies have differing classifying, coding, processing and correlation systems which resulted in the "inability to match all the relevant available information on a responding unit [individual] for analytical purposes."[2] In short, the organization of the government's statistics and information in myriad agencies with myriad practices reflected a time when information integration was, to a large extent, neither possible nor required.

The Kaysen Task Force proposed to establish a governmental apparatus capable of meeting current demands. A National Data Center would it said, largely solve the problems outlined above. The "risky potentials" of a data center

> are outweighed, on balance, by the real improvement in understanding of our economic and social processes this enterprise would make possible, with all the concomitant

---

[1] Task Force, p. 27-28

[2] Ibid., p. 28

gains in intelligent and effective public policy that such understanding could lead to.[1]

The proposed data center would, according to the Task Force, not collect information, but rather maintain an inventory of "all systematic, general purpose, large-scale quatitative information of a demographic, economic or social nature..." and establish "uniform disclosure standards so that legal requirements of confidentiality can be met with no unnecessary sacrifice of analytically useful information."[2] In addition to this, the data center would reduce duplication of effort and the public's burden of numerous questionnaires, preserve for continued use relevant information currently lost or irretrievable, provide improved analysis for government policy makers and scholars and make much information accessible which is now legal to request but too expensive and time consuming to find.[3]

Even though the privacy brouhaha caught the Kaysen Task Force largely by surprise, there are many arguments to the effect that a centralized data center would, in fact, improve the protection of privacy. These arguments, as presented in the House and Senate hearings, revolve around three major principles. First, the technical means are available to make the cost of illegal access or misuse of personal information so

---

[1] Carl Kaysen, testimony in Senate Hearings, p. 6

[2] Task Force, p. 30-31

[3] Ibid., p. 27-37

high that attempted break-ins or misuse would be unlikely. Second, by centralizing data and encouraging those interested in information to use the data center, not the agencies for their information, strict, uniform legislation may be set up governing disclosure and confidentiality. Third, an independent bureau could be set up to see that the center abided by legislative and technical safeguards. Fourth, the data center would be a statistical, not intelligence system.

The development of technological safeguards has the potential of developing as fast as the computer industry itself. It is not within the scope of this paper to examine in depth the various technical safeguards, but it should be noted that there are a large number of options. For example, one can program the computer to distinguish between proper and improper requests for access, to identify "trick" or "accidental" inquiries and to "scramble" the information so that a code is needed to find information.[1] Though proponents of a National Data Center all admit that individual identification of information is necessary for long range studies, they point out that one can make it difficult to identify who belongs to what information, for example, by using two decks of cards, one containing names and triggers and the other only information about an individual and the trigger

---

[1] Edgar Dunn, statement before House Hearings, p. 94

number but not his name. The name cards could be under intensive security.[1]

No technical system of safeguards is "foolproof" however, as almost every witness before the Congressional Hearings has admitted. It is obvious that because people are corruptable and can make mistakes, the machine too can be used for illicit purposes. Thus the Kaysen Task Force relied heavily on Congressional law to govern disclosure and confidentiality standards and the input of information. Confidentiality and disclosure standards will be discussed later in the paper, but suffice it to say now that Federal practice varies a great deal from agency to agency (more so in practice than in content) and that a uniform standard would enable legislators and bureaucrats to keep a closer, stricter watch over the handling of information. The Task Force was particularly impressed with the Census Bureau. The Census Bureau operates under law which does not permit disclosure of individualized information to either the public or other agencies, or permit collected information to be used for any law-enforcement, regulatory, or tax collection purposes with respect to individuals. Its record is superb, all observers say.

> The same statutory restraint could and should be extended to the data center, and the same results could be expected of it.[2]

---

[1] Wilely Branton, testimony before Senate Hearings, p. 311

[2] Carl Kaysen, statement before Senate Hearings, p. 8

As with the Census, Congress would be expected to review all input information placed into the Center.

Also proposed by some is an independent review board to watch over the activities of a National Data Center.[1] Arthur Miller particularly has been an advocate of this.

> The organization (the board) would operate the Center, regulate the nature of the information that can be recorded and stored, enforce the Congressional standard of care for insuring file accuracy, and protect the Center against breaches of security in accordance with the latest technology.[2]

Miller insists that the Board must lie outside the "existing administrative channels" to avoid becoming the captive of those it seeks to regulate.

Finally, by emphasizing the fact that this would be a statistical, not an intelligence system, advocates maintain that the center would not collect information, but only use information already available. In addition to this, the center would be interested solely in providing information in aggregate form, such as the Census Bureau provides.[3]

It is in the distinction between a statistical and an intelligence system however that the crux of the argument between proponents of data center and those hostile to it lies.

---

[1] Committee on Government Operations, H. R. 1842. op. cit., p. 8

[2] Arthur Miller, statement before Senate Hearings, p. 79

[3] Carl Kaysen, testimony before Senate Hearings, pp. 3-25

Arguments Against a Data Center

The arguments against a data center are not really arguments against a statistical system as such. Rather, they are arguments against the potential future of data storage. One of the most common metaphors used to describe the Kaysen proposal is that it is a "foot in the door." In other words, though critics don't doubt the potential for good in a statistical data center, they feel that such a system would plant the seed for a 1984 nightmare because of the need for some individual identification of information.

One of the major foundations of this belief is that controls--legislative or technical--will not succeed in preventing the transformance of a statistical system into an intelligence system. Though Edgar Dunn states "there is no doubt" that a statistical information system can be kept,[1] Arthur Miller in his statement in the Senate Hearings said the pressures for an intelligence or individually oriented system are mounting already for reasons of efficiency, economy, and increased capacity.

> The same efficiencies and economies that the Bureau of the Budget believes would be achieved with a statistical center, the same desire for an increased capacity to manipulate, collect, and record data by machine will eventually create pressure for an individualized system.[2]

---

[1] Edgar Dunn, statement before House Hearings, p. 93

[2] Arthur Miller, statement before Senate Hearings, p. 67

Moreover, as computers become more sophisticated in analysis and integration, the demand by bureaucrats and scholars for "valuable" correlations and collection will increase, thereby jeopardizing privacy. The line between beneficial public analysis and individual privacy, so the argument goes, will continue to blur as more and more chunks of information on an individual are "required" for comparison and public analysis.[1] That indivdual names or identifications must be kept even in a statistical system is seen as the first step towards a kind of creeping computerization of our lives.

Controls, say some other critics, would be ineffective because of the enormous temptation a data center would offer bureaucrats.

> When the details of our lives are fed into the central computer where they are constantly retrievable, we'll to some extent fall under the control of the machines' managers... The filekeepers of Washington have derogatory information of one sort or another on literally millions of citizens. The more such files are fed into files, the greater the hazard the information will become enormously tempting to use as a form of control.[2]

The temptation to sidestep the law would also increase because even though centralization in theory permits closer control over access, "the necessity of speedy national distribution... to authorized recipients requires that access be made available to persons at many locations remote from the central

---

[1] Vance Packard, testimony before House Hearings, pp. 7-22

[2] Ibid., p. 12

data bank.[1]

Moreover, some individuals believe legislative standards and controls are largely useless because rapid technological change is likely to outstrip the effectiveness of the slow legislative process. These people, such as Burton Squires, a professor of computer science, point out that in a centralized system, an entire file can be erased in seconds, making a human being in effect a non-person. Squires and others believe safeguards together with strict legislative standards are the best answer, yet maintain that even together, these controls are far from foolproof.

Lastly, critics argue that if we are to a large measure protected by the inefficiency and decentralization of the present system, why centralize? As Representative Gallagher says:

> Now he an individual has to go to 25 areas and he has to say, "May I have this information?" and he can be turned down at every one of them and I am sure he probably is turned down. Again, I say that you are placing an unbearable temptation before whoever is in control of this central data bank not to use it for non benevolent purposes.[2]

Information from agencies is currently so difficult to gather, that privacy is protected de ~~jure~~ facto, if not de ~~facto~~ jure. Even if tough technological and legal safeguards were established and the "price-per-unit-dirt" went up, that dirt

---

[1] Kenneth Karst, "The Files:Legal Controls over the Accuracy of Stored Personal Data," Law and Contemporarary Problems 31, (1966) p. 360

[2] Gallagher, statement before House Hearings, p.88

would be far more valuable and therefore worth the expense.[1]

Before discussing the relative merits of arguments for and against the data center, the current "inefficient and decentralized system" of information storage should be analyzed, keeping in mind the issues raised by the data center.

ACCESS TO FEDERAL INFORMATION

Present Law and Practice

In theory, the question of who has access to federal information is answered by confidentiality laws. In all there are about 100 Federal statutes providing protection from disclosure to the public of personal information. About 97% of all personal information, for example, is "given some measure of federal protection."[2] Leaving aside the large bundle of specific federal laws forbidding such things as public access to medical files and social security information, the directors of individual agencies in general determine their own disclosure and confidentiality standards. Title 44, section 3102, of the United States Code directs the head of each Federal agency to "establish and maintain...effective controls over the creation, maintenance, and use of records in the conduct of current business." For example, HEW and the Department of Labor may not disclose information "except as the Secretary

---

1 Arthur Miller, testimony before Senate Hearings, p.83

2 Note, "Privacy and Efficient Government: Proposals for a National Data Center," Harvard Law Review, 82 (Dec. 1968) p. 409

of Health, Education and Welfare, or the Secretary of Labor... may by regulation prescribe."[1]

The problem of unauthorized leaks of information could less dramatically be called the failure of disclosure laws. Federal law states that the "disclosure of confidential information generally" about an individual is punishable under law by a fine of not more that $1,000, or not more than a year in prison, or both. The employee who discloses confidential information must be fired.[2]

These confidentiality standards are the major legal means governing disclosure of information to the public. There are two major flaws from a privacy standpoint, however, in this system of confidentiality. First,

> The principle government technique of preserving confidentiality is the use of classification devices like "For Official Use Only;" even the Bureau of the Budget contends this has different meanings under different circumstances.[3]

Thus there is not in general a gradation of "confidentiality" with respect to access by Federal employes, since most information is labeled "For Official Use Only" and most Federal employees have only prove a "need to know" to gain access.[4]

---

1 United States Code, Title 44, section 1306

2 United States Code, Title 18, section 1905

4 According to Arthur Miller, its possible for a federal employee to even inspect IRS files if he submits a written application. See: Arthur Miller, The Assault on Privacy (Ann Arbor: University of Michigan Press, 1971) p., 145

3 Richard I. Miller, "Data Banks and Privacy," in Computers & the Law, report by Standing Committee on Law and Technology of the ABA, (Chicago: Commerce Clearing House, 1969) p. 160

Literally thousands of civil servants have access to scattered confidential information. For this reason, it would be difficult to trace the source of a leak. Given the willingness to spend money, there is little question that confidential information can be extracted from most agencies.[1] The second major flaw is that agency heads, as noted above, have wide lattitude in both establishing confidentiality standards and in maintaing control over access. This results in a great variety of different standards and access procedures which is neither fair nor justifiable if one is serious about establishing a uniform standard throughout the government with regard to both confidentiality and access control procedures.

Ironically, this decentralized nature of information storage which permits so many discrepancies in confidentiality standards is, in fact, the major protector of an individual's privacy.

> The prime protection in this area remained the inability of government agencies and private authorities to actually use the mountains of information they had received in anything like a centralized and efficient fashion.[2]

Moreover, the mingling morass of agency and federal confidentiality standards and practice combined with the topheavy prac-

---

[1] Vance Packard, The Naked Society (New York: David McKay Company, Inc., 1964)

[2] Alan Westin, "Legal Safeguards to Insure Privacy in a Computer Society," Communication of the ACM, 10, (Sept. 1967) Reprinted in Senate Hearings, p. 294-299

tices of agency heads who establish so many of the internal procedures combined once again with natural bureaucratic secretiveness encouraged, it seems likely, the low level bureaucrat to refuse requests for information in order to be safe and save the agency money. Furthermore, as mentioned earlier, the person who wants to assemble a full dossier on an individual has to go through so many agencies that the task is time consuming, expensive and probably, in most cases, not worth the effort.

It was because of this tendency to not release information that Congress passed the Freedom of Information Act--a measure which may substantially undercut this decentralized protection by broadening the right of access.

The Freedom of Information Act

Paradoxically, at the same time the interest in privacy was picking up steam, the demand for public access to government records was also increaing, due in large measure to Congressional and newspaper insistence that Federal agencies classified most everything, thereby hindering the public's right to watch over its government. Enacted in 1966, the Freedom of Information Act sought to place the burden on the agencies to prove that information witholding was warranted by actually listing the exemptions to free access, such as national security information. Most important for our purposes are the exemptions of "matters specifically exempted from disclosure by statute" and matters "the disclosure of

which would constitute a clearly unwarranted invasion of personal privacy."[1]

While seeming to protect the individual in these two last exemptions, the Act in effect raises several questions with respect to personal privacy.

> By establishing an across-the-board statutory policy directing the disclosure of governmental records, the Act reverses the traditional presumption in favor of a citizen's personal privacy, and places the burden on the information holding agency to find a specific statutory ground for refusing to honor a request for disclosure.[2]

Thus the Act tends to downgrade the de facto protection afforded by a plethora of regulations, statutes, and bureaucratic reluctance to search out information for an individual citizen. Furthermore, as Westin points out, it appoints the government

> ...the necessary champion of the citizen's right to privacy. There is no mechanism by which an individual can challenge in ~~the~~ court the willing release by a government agency to the public or to another agency of personal data collected from the individual.[3]

On the other hand, those denied access may seek a court order.

Furthermore, because the confidentiality standards are so decentralized and even contradictory, the individual agen-

---

1 United States Code, Title 5, Section 552

2 Miller, op. cit., p. 154

3 Alan F. Westin, Privacy and Freedom (New York: Atheneum, 1967) p. 387-88

cies are left with the words "clearly unwarranted invasion of privacy" with which to establish whether or not access to certain types of information outside of statutory regulations is in fact an invasion of privacy. Considering that government has yet to decide what privacy really is, that more and more varieties of information are being collected and that the agencies have the burden on them to define "privacy," it is not hard to see that in theory obtaining personal information is now a much simpler task. Clearly, the language begs the question and does nothing to help establish and effective, uniform policy.[1]

A Potential Solution to the Uniform Standards Problem

In the years ahead, the computer will inevitably increase the efficiency of information storage and reduce the cost of access. The need will be clearer then for a federal definition of "clearly unwarranted invasion of privacy" and the establishment of uniform confidentiality standards. Computers are not as inefficient as people and it is silly to continue to rely on the present system.

However, establishing a federal definition of privacy from which to draw uniform confidentiality standards is extremely difficult. Not only is the concept of "privacy" nebulous and obviously hard to translate into concrete legislation, but un-

---

[1] Miller calls the language "theatre of the absurd." See: Miller, op. cit., p. 156-57.

596

iform confidentiality standards themselves may not fill the ideal because of the scope and variety of data collection and storage and the need to weigh numerous factors to determine whether or not an individual's privacy is in fact being invaded if access is granted. (These factors include the context in which the data will be used, how it is used, the accuracy of the information, the completeness of the data, and a host of other elements.) Moreover, a definition is open-ended, i.e., by stating what information may not be released or collected, it leaves open what may be released and collected. Thus even if a tag such as the "clearly unwarranted" phrase was added the problem of numerous agency interpretations would remain.

A better solution might be to pass a constitutional amendment naming privacy as a fundamental right of Americans. Though this would not have the advantage of providing a blueprint with which to rate all information, it would nonetheless establish once and for all that privacy is indeed a right. Hopefully, this right, with the power of the courts behind it, would place the burden on the agencies once again to prove that releasing personal information is justified under the Constitution. Moreover, if the major solution to the problem of storage is the determination of <u>what</u> is collected (for only then are we certain that particular varieties of data will never be subject to access) then a constitutional amendment would force the government to judge whether certain <u>questions</u> violate an individual's privacy. This judgment could be challenged in court. In any case, an amendment would encourage harder hitting access

safeguards if courts granted the government the right to collect certain information only under severe limitations.

To further protect the individual, Congress could enact legislation requiring agencies to release personal information only in aggregate form wherever possible, and to permit release of personal information with names only when a federal board has 1) examined the privacy implications and notified the individual and 2) the individual has granted his consent. Hopefully, this would in large measure reduce the need to depend on such weak phrases as "clearly unwarranted invasion..." to protect personal privacy. Congress should also move toward establishing uniform technological safeguards limiting access through legislation ordering the Administrator of General Services to purchase computers with safeguard capabilities.[1] As far as I know, there is no federal law ordering agencies to even employ technological safeguards.

TRANSFER OF INFORMATION AMONG AGENCIES

Present Law and Practice

The central problem inherent in the transfer of in-

---

[1] Paul Baran, testimony before House Hearings, p. 126

formation among agencies is that data gathered for one purpose may not be sufficiently complete or truthful if used for a different purpose by another agency. Data-transferral may violate the privacy and trust of an individual by taking information he provided confidentially to one agency and using that information out of context for the different purposes of another agency. Sometimes, obviously, ~~inaccurate~~ information bred of a wrong context will harm an individual. Present law sidesteps this problem to a great extent.

Transfer of information between agencies is governed by the United States Code, title 45, section 3508. This law permits information transfer of personal information among agencies only in the form of statistical "tools or summaries", unless 1) "the information was not at the time of collection declared confidential", 2) the person consents to release of information to a second agency, or 3) the agency receiving the information has the authority, legally supported with criminal penalties for those refusing to supply information, to collect the information for itself. The same law also states that any information obtained in confidence by one agency carries all the unlawful disclosure rules and penalties with it when transferred to another agency.

By law, the director of the Bureau of the Budget directs the transfer of information. (He also is theoretically responsible for determining what information needs to be collected and methods of collecting it.) He may order one

Federal agency to make available to another Federal agency any personal information collected by another agency after December 24, 1942 and all agencies in general are directed "to cooperate to the fullest possible extent at all times in making information available to other agencies."[1] He is also ordered to "coordinate...the information collecting services of all agencies... using as far as practicable...files of information and existing facilites of the established Federal agencies."[2] Lastly, all Federal agencies are directed to tabulate their information in such a way as to "maximize the usefullness of the information to other Federal agencies and the public."[3]

It is clear that these laws were designed to facilitate the transfer of information and that only a cursory glance was thrown in the direction of protecting privacy. There are several loopholes in the laws which may result in unwarranted invasions of privacy. First, there is no provision establishing a "need to know" on the part of the receiving agencies.[4] Even if the director of the Bureau of the Budget were told to weigh all requests for transfer of information according to strict "need to know" standards, it is doubtful if he could throughly monitor the huge amounts of stored information. Moreover, there are no effective means to insure

---

1 United States Code, Title 44, Section 3507

2 United States Code, Title 44, Section 3503

3 United States Code, Title 44, Section 3501

4 Miller, op. cit., p. 143

that even the actual limitations on the director's power to transfer information are working.

> This need not and should not be the case. A computer system can be programmed to indicate the source of data that has been received from another agency.[1]

Second, once information has been transferred there is no guarantee that it will be either destroyed after use, or not eventually used again for other purposes which are far removed from the original purpose of the data gathering or transferred to a third agency. Third, it is perfectly legal to transfer information which an individual may have assumed was granted to only one agency so long as the agency receiving the information could collect the data itself.[2] Finally, it is conceivable under the law as written to transfer a certain type of information freely if at the time of collection it was not explicitly confidential. It should be mentioned also that the Freedom of Information Act may grant greater access to personal information to agencies, not just the public for the reasons previously mentioned.

As Miller points out, it is difficult "to obtain concrete information on the character of interagency exchanges of personal data."[3] It is difficult to know how widespread Vance Packard's horror stories of individuals who give the Federal Housing Authority information for a loan and then are refused

---

[1] Ibid. p.143

[2] Ibid. p. 143

[3] Ibid. p. 142

a job with a government contractor.[1] Part of the reason it is difficult to judge the extent of the problem is because it is doubtful if a man hurt in the way described above ever finds out the reason he lost his job. Nonetheless it seems safe to make several broad assumptions.

The law encourages transfer of data for efficiency reasons and there is little question that the capabilities of the computer will not only make that transfer easier, but will also create greater and greater temptations for "valuable" integration of information. This may create many future violations of an individual's assumption of confidentiality and individual agency use.

> The practice [inter-agency use of confidential data] is increasing, indeed is accelerated under the pressure of vast new Federal social welfare programs.[2]

Even if the exchange is currently mostly scientific and technical data[3] and even if confidentiality rules, national security requirements, agency jealously and the incompatibility of record-keeping techniques all hinder transfer of data, one can safely assume that pressures will mount for increased data exchange as the computer revolutionizes data storage.

Intelligence systems such as the F.B.I. have access to most government files and there is extensive exchange of information between the F.B.I., for example, and local law

---

[1] Vance Packard, testimony before House hearings, p. 11

[2] Albert Mindlin, "Confidentiality and Local Data Systems," statement presented to Senate hearings, p. 379

[3] Miller, op. cit., p. 141

enforcement agencies.

Some Possible Non-solutions

Given that information flow is necessary in our society, the question becomes how to limit access, as much as is consistent with privacy, to the group which gathered the information--and thereby reduce the risk of misunderstandings.

One potential solution is to make the principle of consent to transfer of personal data the controlling theme of all personal data transfers. This would also involve giving the individual the right to challenge all governmental release of personal data to the public.[1] There are several shortcomings to this, however. First, it is questionable whether low income persons applying for welfare or a loan, for example, will refuse consent if it is requested but not required, if only because they feel that refusing the request might jeopardize their chances.[2] Second, it is unlikely that Congress would pass such a measure, especially retroactively, simply because agencies would then have to duplicate their efforts to an enormous extent to arrive at correct statistics. Third, the consent principle does not distinguish between different kinds of information, though in theory this could be rectified. Fourth, by placing the burden on the individual, the

---

[1] Westin, Privacy and Freedom, p. 375

[2] This is similar to Karst's "implied coercion." See Karst, op. cit., p. 344-355

government essentially asks the individual to protect himself, i.e., this solution fails to come to grips with indivdual oversight or trust in the government.

Another potential solution is to treat personal information as a property right, with all the corrolary rights.[1] As Miller points out, this too begs the real question for it would not change the practice of government. A property right solution would just force people to be ever on the lookout when the government approached them.[2]

Finally, several persons have suggested an "aggregation scheme" whereby each piece of information would be rated on a confidentiality scale.[3] Thus, agencies would then be permitted to transfer information on the basis of some such confidentiality scale. However, the sensitivity of a given document "is not intrinsic, but varies with the relationship between the agency gathering the data and the agency receiving it."[4] If this intrinsic sensitivity is valid, then each agency would have to rate every type of information it possessed for every other bureau.[5] This is clearly an impossibly huge undertaking.

---

1 Charles Reich, testimony before House Hearings, p. 32-33

2 Miller, op. cit., p. 212

3 This was proposed with reference to the data center, but it can be expanded to cover the current government operation. See Committee on Government Operations, H.R. 1842, op. cit. p. 15

4 Ibid., p. 14, 15

5 Ibid., p. 14, 15

If none of these solutions are acceptable, what are we left with? Because no clear cut solution stands out, perhaps the best means of protecting privacy in this area is by approaching the problem from several different angles, hoping that the sum of approaches will equal better protection. Thus we could make use of all the methods outlined above, particularly consent, and appoint a bureau to rule on transfers. Crucial to the success of this bureau, however, would be Congressional legislation ~~ordering~~ permitting the transferral of information only in aggregate form wherever possible. If Congress did not pass such legislation, a bureau would be swamped if it had to rule on all data transferral. The idea of a bureau offers several advantages. First, it could be given the power to decide whether information should be transferred which is not in aggregate form. If permission is granted, the information should be transferred with the understanding that this data may only be used for the purposes outlined by the requesting agency--purposes which presumably would not include any possibility of information taken out of context or any other sort of abuse. Second, the bureau could serve as a watchdog, akin to the GAO, for privacy invasions growing out of access or transferral practices. Third, the bureau could be given ombudsman powers, and hear complaints from citizens whose privacy has been abused in the transfer of information.

Enacting a constitutional amendment might also serve some beneficial purposes by opening up the courts to complaints against the government. Once again, however, given the huge

supply of information, the real answer lies in what the government determines may be collected. In any case, the above suggestions would at least improve the present situation where the Bureau of the Budget is primarily concerned with easing transfer of information.

THE PROBLEM OF ACCURACY

There are several different kinds of inaccuracy. First, there is a potential for human mistakes at the gathering stage, at the clerical level and at the input level of the computer. Second, inaccuracy may occur by inserting "hearsay" or spottily investigated information into a file. Third, there is a potential for inaccuracy when facts are taken out of context.

As far as I can tell, there are no government wide regulations aiming to rectify human errors through review procedures or accuracy standards. This is understandable if not commendable because of the expense involved in reviewing files, and because most personal information is used for statistics. The scope of the problem is nonetheless illustrated by a stu-

dy which examined the Air Force personnel data system. It found an error rate of 5-6% in the files. Moreover, almost every file randomly chosen for investigation contained at least one error and most averaged three or four.[1]

The best means to rectify the error rate is to permit each individual access to his own file, unless the file concerns national security information.[2] There are two major objections to this. First, it might be prohibitively expensive. In order to ~~in fact~~ correct a sizeable percentage of the widespread inaccuracies in the files, the government would have to assume the burden of sending all information held on an individual to that individual and not rely solely on personal requests for information. Even if all information was centralized, this would be an expensive and time consuming task, particularly in terms of computer time. But given the decentralized nature of information storage, the collection of all personal information on an individual into one folder for mailing would be enormously time consuming and expensive. In any case, some kind of cost analysis on mailing information to 200,000,000 Americans should be made before any decision is taken. Considering this vast amount of decentralized information, the best solution might be to identify certain files or types of files as worthy of the expense of government notification and to allow individuals to request access to their file.[3] Notification, in particular, should be required if con-

---

[1] Cited in Committee on Government Operations, op. cit., p. 32

fidential information is to be either released or transferred. The second major objection to a system of individual access is that it would encourage hundreds of thousands of "petty squabbles."[4] The conditions of access described above should reduce the petty squabbles. But even so, petty squabbles are a price that should be paid to insure accuracy....

---

[2] This has been suggested by countless people.

[3] Miller, testimony before Senate Hearings, p. 77

[4] Karst, op. cit., p. 358-9

608

~~petty squabbles. But even so, petty squabbles are a price that should be paid to insure accuracy.~~

Given the need to store some "soft" data on some individuals, such as Federal employee ratings, national security information and information on criminals, the best protection to the privacy of the individual would be a combination of extremely strict safeguards on this information, the right of an individual to see his file if derogatory information is to be used against him, and some sort of watchdog bureau over the information gathering processes of intelligence systems such as the FBI. A bill now before the Senate--S. 1438--goes far to standardize what information may be gathered on Federal employees though it excludes the FBI.

Finally, innacuracy may occur when information is taken out of context--the "evaluative fact." People, in other words, can make wrong judgments if they don't know the history of a piece of information.

This problem of the evaluative fact is likely to increase as integration and analysis of information comes more and more into demand. Thus the problem of facts taken out of context really becomes somewhat similar to the problem of the transferrance of data, and the same solutions suggest themselves. In particular, however, to decrease the likelihood of innacurate contexts, information should be transferred, as a general principle only in aggregate form or with the names deleted.[1] Several agencies, such as the Social Security Ad-

---

[1] This is another idea applied to the data center which could be applied to the current system.

ministration, generally, do this on their own. The principle of consent, thoughed flawed, should also help reduce the problem by giving the individual the right to designate whether data gathered for one purpose may be used for another. Most important, however, is giving the individual access to his file if data in it is used in such a way as to directly affect him.[1]

The problem of the evaluative fact applies particularly to old information. There are countless horror stories of a man being denied a job or credit because he shoplifted a candy bar when he was fourteen. As usual, it is difficult to determine the extent of the obsolescent information problem because different agencies have different practices and the man denied a job because he stole a candy bar usually isn't told why he didn't make the grade.

THE PROBLEM OF OBSOLETE INFORMATION

Destroying Federal records is a complicated process, and a marvelous example of red tape in operation. Agencies are required to destroy after a minimum retention period records "not having sufficient value to justify their further retention. The agency itself designates classes of information "of continuing value."[2] However, before actually destroying the records, the head of each agency must give the Admin-

---

[1] Karst, op. cit., p. 359

[2] Code of Federal Regulations, Title 41, Sections 101-11, 401-4 (a).

istrator of General Services recommended lists of records for disposal which will not have "sufficient administrative, legal, research, or other value to warrant their further preservation..."[1] The Administrator after reviewing the proposals must submit a list of recommended disposals to a joint Congressional committee which then lets the Administraor know what records to destroy.[2]

It is obvious from reading these laws that their purpose is to prevent the destruction of records which might someday be valuable. I was unable to find any reference to privacy in the laws (except in unusual agencies such as the Census, which nonetheless has very modest disposal practices). In a sense these laws fit Max Weber's notion of a bureaucracy as self-perpetuating. ~~Only here~~ Here, it is information which is self-perpetuating, and it will probably become more so as computers lessen the cost of storage. The wording of the law is so vague, moreover, that almost any piece of information could probably be found to be of "continuing value" if the agency wanted to keep it. Determining what information to keep is therefore not subject to any concrete or uniform standards--the individual agency head generally determines what, in his own judgment he should destroy since it is he, after all, who provides the original lists to the Administrator. Furthermore, the distance between a small piece of information in one of the

---

[1] United States Code, Title 44, Section 3303

[2] Ibid.

hundreds of bureaus and final destruction is so long in terms of red tape that the process seems literally incapable of responding to the demands of personal privacy.

On the surface the solution seems to be to establish a retention standard which leaves the burden on the agency to prove that the information must be kept, rather than the current practice of proving it must be destroyed. The problem, once again, is that to do this one must have a detailed standard with which to rate the information's life span--and there are so many types of information that the task seems almost impossible at present. Not all cases are as clear cut as the one cited above and there is a real question whether derogatory information, if true, should be destroyed. Another solution, that of periodically re-evaluating all personal information, seems possible but equally difficult to accomplish given the vast amount of information in the Federal system.[1]

Despite the problems mentioned above, placing the burden on the agency and devising some sort of standard can be commenced. A start might be the following: First, and most importantly, data's life-span should be rated at the collection stage by a group independent of the agency in consultation with the agency. This group must make sure that the life-span of data must correspond to its declared purpose during consultation. Once made, it should be difficult to reverse a decision. If a long-term study is made on sensitive infor-

---

[1] Miller, op. cit., p. 250-51

mation, the two sets of cards should be employed (p. 7, 8) and individual consent required. Second, juvenile court records should be "sealed" and opened only in case of further trial.[1] The same should hold true for people acquitted in court. Third, all information gathered on an individual when he applies for a government job should, in general, be destroyed when the job is over or the government refuses employment. Fourth, some sort of rolling sensitivity scheme which makes information more protected as it grows older should be adopted.

THE DEVELOPMENT OF A DE FACTO NATIONAL DATA BANK

Like the telephone, telegraph and railroad before it, the development of the computer began as a decentralized process. And like the telephone, telegraph and railroad, computers too seem headed toward increased centralization.[2]

> Today we can see the independent, private automated information systems being inter-connected to form larger, growing systems. The direction of growth is clear.
> My thesis is this: Today we are already building the bits and pieces of separated automated information systems in both the private and government sectors...a de facto version of the system you are now pondering is already into the construction phase. It is in many ways more dangerous than the single data bank now being considered.[3]

Baran goes on to say that independently developed systems

---

1 Karst, op. cit., p. 358

2 Lance J. Hoffman, "Computers and Privacy: A Survey," Computing Surveys, 1,(June, 1969), p. 85-86.

3 Paul Baran, testimony before House Hearings, p120-21

will merge by necessity for economic and efficiency reasons. Already, about 20 Federal agencies, bureaus, and departments "operate line-sharing systems or are in the process of installing them."[1] Moreover, federal laws now on the books encourage the centralization of information storage. Title 40, section 759 of the United States Code urges the Administrator of General Services to transfer computers from agency to agency and to utilize them jointly in order to save money. As already indicated in the chapter on information transferral, the Director of the Bureau of the Budget is supposed to coordinate information collecting and encourage agencies to use files already available to them as much as possible.

Clearly, it is highly possible that under existing laws and practice, the pressure form "valuable" integration of information may create a national data bank in fact if not in name while nobody is really looking.

Another Look at the National Data Center

In one way, the original House Hearings on the National Data Center were beneficial, for they brought the privacy implications of a data bank before the public for the first time. On the other hand, this very fact of catching the data center advocates off-guard on the privacy issue created a certain tone in the Hearings "of beating up the eggheads" as Kaysen puts it.[2] Given the press coverage, and sarcastic questions

---

[1] Miller, op. cit., p. 60

[2] Interview, op. cit.

in the Hearings, the data center was irrevocably linked with 1984. There is a great danger in this, for by raising the spirit of Orwell, one tends to immediately deny it, whether in fact there is any foundation to the belief or not. The data center was killed, but in the process Congress may have ignored the really big question of coming to grips with technology in the present and in the future. Instead of seriously reviewing the issues, the committee tended to scream for the headlines.

The data center proposal as presented offered Congress a chance to rationally and calmly come to grips with the problems of government storage. It offered a serious opportunity to attempt to balance the relative benefits of centralization and decentralization. It gave Congress the opportunity to experiment with and study various technological safeguards in a system which was to be small and tightly controlled. The argument that no controls will work in a centralized system represents little more than a simplistic evasion of the entire issue. Uniform controls, obviously, must be found and developed for the entire system. The idea of an independent review board, as suggested by Miller, could also have been tried out.[1] A standardized means of disclosing only aggregate and not personal information could have been tested in a centralized area, with particular reference to its potential in the rest of the government. The list could go on and on. At the risk of be-

---

[1] Miller, testimony before Senate Hearings, p. 71

coming redundant, it is not really the data center which threatens us but rather lack of study of the problem.

The strongest argument of those opposing a National Data Center is that, in the end, it might prove to be a great temptation for anti-democratic forces in the United States. Given the seemingly inevitable integration of storage systems, the best protection to fall back on once again is control over the input of information. If Congress eventually establishes close control over the creation of questionnaires by developing a ~~comprehensive definition of privacy~~ Constitutional amendment, we will all be safer. But a new system of controls within the governmental organization of storage, geared to this ~~definition~~ amendment, are also needed to see that the laws are acted upon. An independent bureau, as suggested several times in the body of this paper, might be a good start.

Mark Stahl

VI BIBLIOGRAPHY

Books

Miller, Arthur R. Assault on Privacy. Ann Arbor, University of Michigan Press, 1971.
The most current, perceptive, and most complete book to date on the question of privacy.

Miller, Richard I. "Data Banks and Privacy." In Bigelow, Robert P., ed. Computers and the Law. 2nd ed. Chicago, Commerce Clearing House, 1969.

Packard, Vance. The Naked Society. New York, David McKay Company Inc., 1964.
A good if not serious expose.

Weston, Alan F. Privacy and Freedom. New York, Athenium, 1967.
Good comprehensive survey.

Wheeler, Stanton, ed. On Record: Files and Dossiers. New York, Russell Sage Foundation, 1969.
Fair, but not very hard hitting , review of American record-keeping organizations.

Periodicals

Hoffman, Lance J. "Computers and Privacy: A Survey." Computing Surveys. 1: 85-103. June, 1969.

Karst, Kenneth L. "The Files: Legal Controls over the Accuracy and Access of Stored Personal Data." Law and Contemporary Problems. 31: 342-372, 1966

Kaysen, Carl. "Data Banks and Dossiers.: The Public Interest. Spring, 1967. (reprint from Appendix of Senate Hearings, part 1, Computer Privacy. March 15, 1967, pp. 265-269. See under Government Publications)

Miller, Arthur R. "The National Data Center and Personal Privacy." Atlantic Monthly. 220: 53-58. November, 1967.

Note. "Privacy and Efficient Government Proposals for a National Data Center." Harvard Law Review. 82: 400-417. December, 1968.

Seaborg, Dr.. Glenn T. "Time, Leisure, and the Computer." (speech reprinted in Appendix of Senate Hearings, Computer Privacy. March 15, 1967, pp. 248-256. See under Government Publications.)

Weston, Alan F. "Legal Safeguards to Insure Privacy in a Computer Society."

Periodicals(cont'd)

Communication of the ACM. September, 1967. (reprint from Senate Hearings part 1, Computer Privacy. March 15, 1967. 10: 294-299. See under Government Publications.)

Government Publications

Report of the Commission on the Preservation and Use of Economic Data to the Social Science Research Council. Richard Ruggles, Chairman. April, 1965. (available in House Hearings, Computer Privacy. pp. 195-253.)

U.S. Bureau of the Budget. Report of the Task Force on the Storage and Access to Government Statistics. Carl Kaysen, Chairman. Washington, Government Printing Office, 1966.

U.S. Bureau of the Budget. Statistical Evaluation Report No.6-- Review of the Proposal for a National Data Center. Edgar Dunn, Chairman. Washington, Government Printing Office, 1965.

U.S. 89th Congress, 2nd session. House Committee on Government Operations before the Special Subcommittee on the Invasion of Privacy. The Computer and the Invasion of Privacy. Hearings...Washington, Government Printing Office, 1966.

U.S. 90th Congress, 1st session. Senate Committee on the Judiciary. Computer Privacy. Hearings before the Subcommittee on Administrative Practice and Procedure... on S. Res. 25. Washington, Government Printing Office, 1967.
All the arguments and principles of the controversy are available in these House and Senate Hearings.

U.S. 90th Congress, 2nd session. Senate Committee on the Judiciary. Computer Privacy. Hearings before the Subcommittee on Administrative practice and procedure... on S. Res. 25. Washington, Government Printing Office, 1968.

U.S. 90th Congress, 2nd session. Committee on Government Operations. "Privacy and the National Data Bank Concept." Hearings... on H.R. 1842. Washington, Government Printing Office, 1968.

Interviews

Kaysen, Carl. November 10, 1971. Princeton, New Jersey.