Privacy Act and the Data Base: Implementation of the Privacy Act

William B. Camm, Staff Assistant for Tests, Technical Information Division, US Army Research Institute for the Behavioral and Social Sciences, Alexandria, Virginia 22333

The legal constraints of the Privacy Act and the increased legislative pressure to handle greater amounts of personal data faster and better pose many new problems for social science research in the US Government. The immediate, intermediate, and long-range solutions to the dilemma can be achieved through the development of a precautionary, systematic set of collection and storage procedures; full use of comprehensive data bases; and the insulation of each contributing set of data.

This paper is offered as a contribution to the current discussion on the legal constraints of the Privacy Act of 1974 (P. L. 93-579) on the pressing requirement that the Department of Defense handle greater amounts of complex personnel research data faster, better, and at minimal cost. The mandate to minimize the cost of collecting, maintaining, and using personal data and, at the same time, maximize the utility of the collected data is articulated in the Paperwork Reduction Act of 1980 (P. L. 96-511). The act became effective 1 April 1981.

The technology capable of cogent management of information resources is here, and we already know quite a bit about how to apply it in terms of cutting costs, enhancing usefulness, coordinating and sharing common procedures, and improving service to management and the user.

The problem, however, is to insure that the collection, maintenance, use, and dissemination of personal data and information are consistent with the Privacy Act. The restrictions and limitations imposed by the Privacy Act loom large as a potential hindrance to effective information resource management in terms of information control, resource contraints, cost of data, added hardware, and special computer programs. To date, neither the impact of the new wave of information resource management on the Privacy Act nor the constraints of the Privacy Act on the Paperwork Reduction Act has been sorted out. The Office of Management and Budget (OMB) has been assigned the responsibility for providing overall direction in the development and implementation of policies, principles, standards, and guidelines in all areas of P. L. 96-511. Privacy Act enhancement is on the OMB agenda for April 1983.

All Federal agencies are moving ahead with their own interpretation of P. L. 96-511 with the expectation that the resulting implementation will be found acceptable. The Office of the Assistant Secretary of Defense, Information Control Division, has established DOD-wide policy to insure compliance with P. L. 96-511. The Army has established an Information Management Office under the Office of the Chief of Staff to define, develop, and manage the Army information resources program.

Organizations that deal with personal information have reached informal agreements on most aspects of the program but have avoided special problems with regard to processing and maintaining personal data--especially with the concept of an integrated data base. Nevertheless, the problems of effective information resource management and protection of individual privacy are quite real and very pressing and cannot be ignored.

THE PRIVACY ACT

The Privacy Act is imposed on executive departments, military departments, Government and Government-controlled corporations, other establishments in the executive branch including the Office of the President, and independent regulatory agencies. Congress and its agencies (e.g., GAO) are exempt. So are Federal Courts. It limits the manner in which they collect, use and disclose information about people. The act was codified as 5 USC 552a in 1976. The act gives the individual the right to be protected

against the power of officials with access to data banks. There are three related aspects to the Privacy Act rights:

> Personal autonomy--the right to make a choice about personal behavior and lifestyle.
>
> Freedom from outside interference--the right to be left alone.
>
> Protection of private information--the right to control where and how information about oneself is communicated to others.

This portion of the paper focuses on the third point, control and protection of personal information.

Since 1965, personal privacy has become an important social value, covered under tort laws, in the United States. Privacy is related to personal freedom and, although rights of privacy are not expressly mentioned in the Constitution, is supported by the Supreme Court's language used in most of its important decisions. The constitutional amendments most commonly cited in this regard by the Supreme Court are the first, third, fourth, fifth, ninth, and fourteenth.

In 1976, an inventory of Federal data systems revealed that 97 agencies had a total of 7,000 records systems containing nearly 4 billion dossiers. The Department of Defense alone had 2,219 systems with 321 million different names and records (OMB, 1976). Most of the records systems at that time were not a matter of public record. The Privacy Act prohibits secret files and further states that individuals should be able to find out what information about them is contained in Federal records and how that information is used. For example, a person is able to prevent personal information that was given for one specific purpose from being used for another purpose without his or her consent. Provisions will be made for the individual to correct and amend personal records in possession of the government.

Government agencies handling identifiable personal data should show that such data are reliable and current and take positive steps to prevent their misuse. Collected data should also be safeguarded and securely stored if they contain identifiable information. For research use, the connection between the names and data should be destroyed when no longer needed. Code numbers and code words can be used if several sets of data are collected on the same person. A number of methods for storing personal data are described in the literature (Boruch, 1971a, b). If knowledge of illegal activities is requested, anonymity should be guaranteed so that the data cannot be subpoenaed in legal proceedings. Insulated data banks might be considered. Research data are not automatically privileged information. There are a few exceptions, such as data regarding drug research. Congress and courts may, and often do, subpoena such data.

The success of the Privacy Act is hard to measure objectively. The enforcement of data protection regulations and the supervision and control of the collection and storage of information about individuals depend, for the most part, on the good faith of the agencies and legal action by individuals. Congress believed that self-regulation was the best initial

method for control because it eliminated the need for an additional government agency and, at the same time, would aid the necessary advance in the technology of information collection and storage. Control agencies were not to be considered unless the agencies themselves proved that self-regulation had failed. The potential for frustration of the law is so great that the Privacy Protection Study Commission (1971) recommended that the Privacy Act be broadened to include all items that an agency can readily identify in all of its systems to insure compliance with the Privacy Act. So far, Congress has not implemented the Commission's recommendation.

Violations of the Privacy Act are misdemeanors subject to a maximum fine of $5,000. Unlike damage actions brought against an agency, criminal penalities are imposed on the person who committed the crime. The punishment, if there is a conviction, is applied to any

> Agency officer or employee who knowingly or willfully makes improper disclosures of information pertaining to an individual.
>
> Agency officer or employee who willfully maintains records without meeting Notice Requirements Requests (i.e., maintains a secret system of records).
>
> Person who knowingly and willfully requests or obtains individual records from an agency under false pretenses.

If the court finds that an agency (its officers or employees) acted in an intentional or willful manner, the complainant may receive actual damages ($1,000 minimum). But it is difficult, in most cases, for the complainant to show proof of intentional and willful agency misconduct. The complainant must also show that the conduct was greater than "gross negligence"; "ordinary negligence" on the part of the agency does not meet requirements of the law as it is written. In addition, the complainant must also prove actual damages by establishing that the agency's action had a direct adverse impact upon him or her. Finally, if the individual wins, the U.S. Treasury (not the agency or its members) is liable for the actual damages, court costs, and attorney fees. This situation tends to dampen the deterrent effect that civil actions may have upon data collection practices of agencies (Bushkin & Schaen, 1975).

THE DATA BASE

A data base may be viewed as a digital computer version of a manual file system. The manual file system comprises file folders identified by a name or number. The computer file consists of records, each identified by a primary key and secondary keys, for example, name, age, rank, and Social Security number. At this point, the computerized record system departs from the manual system. Access to the items in the computerized system can be made through the primary or any secondary key, or through any other indicator in the individual record. Users of computerized records systems are often in remote locations, and restrictions, like code names for the primary key or identification tab of a single system, no longer exist.

The recent trend, under the impetus of the Paperwork Reduction Act, is toward integrated data bases where a collection of data or records is linked together using a common identification key. The reason for the innovation is related to a greater need for individualized information and a growing proficiency in processing and interpreting data. Also, as expected, data collected for one purpose is frequently useful for related purposes.

At this point, the distinction between records that relate to individuals for the purpose of taking some sort of action concerning that individual and records that are collected and maintained for the purpose of planning and policy decisions should be made. The former, in the strict sense, is termed a system of records; the latter is statistical record. However, most records are mixed, and it is rare to find a true statistical record in either Government or academic research. A true statistical data base cannot contain information that can be related to an identified individual, and no individual contributing to the data base should be identified with it.[1]. The Army Research Institute for the Behavioral and Social Sciences (ARI) collects and maintains systems of records until such time as the data are edited, coded, stripped of the personal identification, and entered into the data base. The ARI Systems Notice (ARI, 1980) covers, at this writing, all ARI systems of records of the moment and the future, provided the data collection effort remains within the operational confines of the public notice. If a new and different system of records is contemplated, then an additional notice, or modification of the current notice, will be required. The new notice must be published in the Federal Register at least 90 days prior to any data collection for the new system of records.

DATA COLLECTION AND STORAGE PROCEDURES

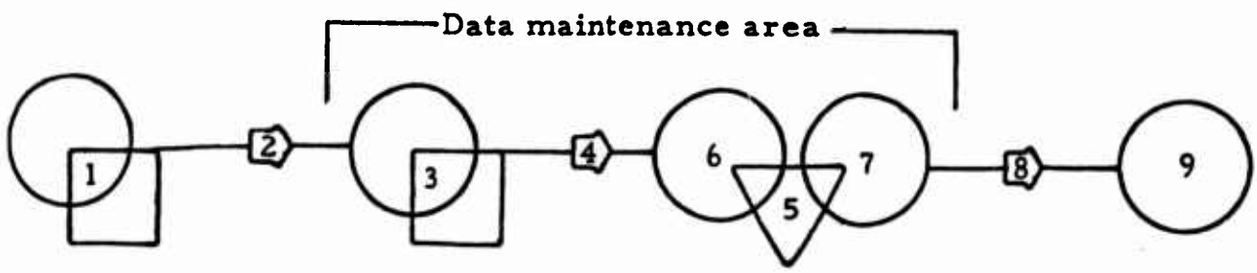
The procedure involved from the start of the data gathering through the final destruction of the system of records (i.e., removal of personal identifiers) and the publication of the results for the various users may theoretically be compromised at a number of points during the collection, transmission, storage, and processing of the data. Nine arbitrary points are conceptualized here for the purpose of illustration in Figure 1.

The data collection point 1 surveys, questionnaires, tests, interviews, or ratings is obvious and frequently overlooked despite the Privacy Act Statement at that point stating that "Full confidentiality of the responses will be maintained in the processing of the data. . . ." (DA Form 4368-R). The Privacy Act requires that all agencies involved in data collection--in the development of a data base there may be several--provide

---

[1]Insofar as the Privacy Act is concerned, however, the only operative criterion is whether or not the agency does in practice retrieve the information by reference to some personal identifier.

appropriate administrative, technical, and physical safeguards. The common threat to personal information at point 1 is the person who is authorized to have access to the information for one purpose but who misuses that same information for an unauthorized purpose. The entire data collection operation, if possible, should remain under a single work group. It is tempting for the researcher to ignore most of the problems at point 1 and go on to the second potential compromise point (transmission). The personal information is easiest to protect in the computer-based area. The transmission of the data (point 2) may be by messenger, mail, telephone, or microwave and is subject to compromise during transmission and upon receipt. Any privacy compromise here is seldom intended and is most likely the result of careless handling. Security compromise during transmission is not specifically treated in this paper. The editing and coding process (point 3) is the first step in preparing the data for the computer and is the time to check for accuracy, relevance, timeliness, and completeness. It is also a good time to remove the personal identification in preparation for linkage with additional information in the integrated data base unless that linkage is necessary for the subsequent interpretation of the data. Data transmission (point 4) to the computer area is usually less of a risk than point 2. However, the information must be checked, edited, and sorted and may easily be identified by resourceful people. Points 5, 6, and 7 involve checking the processing of the data being edited and stored. During format checks of tables, graphs, and the like, careless handling may result in compromise. Error listings are another source of compromise at this point. The location of each item of information should be recorded and confined to the computer area; extraneous data should be destroyed when no longer useful. Finally, point 8 is transmission of the data (the report) to the user, point 9. Exploitation can occur when common and unique properties of individuals are displayed in the reports. It is then a simple matter to sort, count, and identify individuals and/or groups from the final report. For example, tabulation of results may yield grade level, age, sex, location, and other properties that with cross-tabulations identify individuals and/or groups.



1. Data Collection
2. Data Transmission
3. Editing and Coding
4. Data Transmission
5. Data Preparation
6. Computer Processing and Storage
7. Tabulation and Display of Results
8. Data Transmission
9. Report: The results to user

Figure 1. Flow from personnel information initial collection to statistical record to final report. Numbers represent potential compromise points.

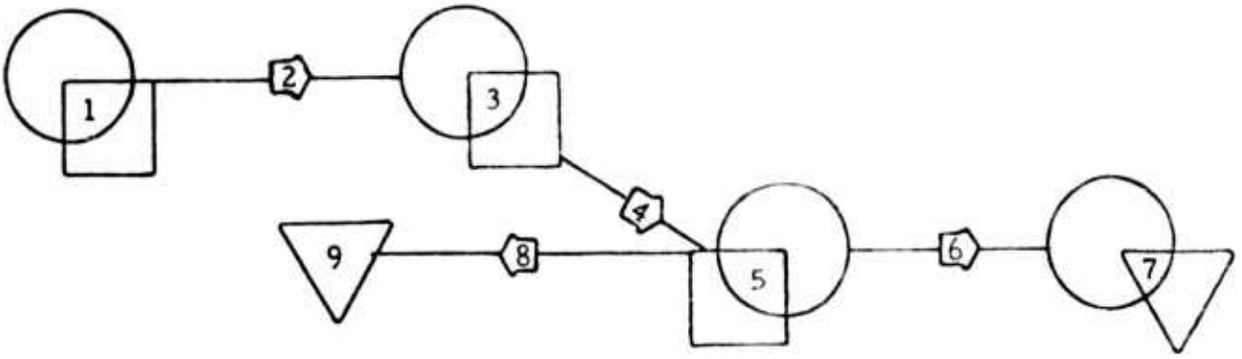
In practice, the situation is more complicated. Longitudinal studies which involve collecting and maintaining information over a period of time may present problems. A statistical data base of this sort needs an insulated method of linking recent data with data already stored. To complicate matters, a secondary user or users are often involved. And most problems arise--at least insofar as privacy safeguards are concerned--when the primary user establishes the data base for administrative purposes and the secondary user is more interested in research, or vice versa. Often, there is no relationship of purpose between the records system of one user and the established data base of another.

PRIVACY SAFEGUARDS

Privacy safeguards for data bases are similar to those required for most records systems. Certain data bases, for example those concerned with current sensitive issues, such as medical histories, performance by ethnic groups, illegal actions or country of origin (Barnes, 1979), are subject to intentional invasion for several reasons by individuals whose interests range from apprehension concerning possible misuse, real or imagined, of the information contained in the data base to intelligence-gathering activities of foreign governments. Added precautions might be considered.

For example, the data from MILPERCEN's proposed data base are coded with a cryptographic code known only to MILPERCEN. The coded data plus identifying information are sent to ARI to merge with ASVAB data, which is also coded using the identifying information to link with the MILPERCEN record (Figure 2). The identification is then deleted. The merged file is given to ARI's Personnel Utilization Technical Area. MILPERCEN cannot obtain anything other than their own data from the file, and ARI cannot meaningfully identify data from MILPERCEN but will have the necessary information for a validation of ASVAB. The same scheme can be used in longitudinal studies with different independent groups. The code linkage can either be destroyed or stored in a safe place beyond the reach of all but extraordinary requests.

Assuming reasonable precaution in data collection, maintenance, storage, and reporting, the insulated data base with its disposable code links and the resulting statistical record will easily meet future requirements for privacy protection of ARI integrated data bases. There are many other effective methods to insulate and link record systems. There is no one best way to protect personal information. The point is that such protection can and should be provided.



1. MILPERCEN Data--Coded
2. Data Transmission
3. ARI ASVAB Data--Coded
4. Data Transmission
5. Merge Data--Edit and Match Codes
6. Data Transmission of All Coded and Merged Data
7. The Data Base--Statistical Records only
8. Transmission of Cryptographic Key
9. Safe Storage of Code Key

Figure 2. Schematic flow and proposed development of one insulated data base.

Army Research Institute. ARI Systems Notice. A1306.01 DAPE. Federal Register, 45FR No. 223, 75736. Nov. 17, 1980.

Barnes, J. A. Who Should Know What? Cambridge, England: Cambridge University Press, 1979.

Boruch, R. F. Assuring confidentiality of responses in social research: A note on strategies. American Sociologist, 1971a, 6, 308-311.

Boruch, R. F. Maintaining confidentiality of data in education research: A systematic analysis. American Psychologist, 1971b, 26, 413-430.

Bushkin, A. A., & Schaen, S. I. The Privacy Act of 1974. McLean, Va.,: System Development Corporation, 1975.

Office of Management and Budget. The first annual report to the President for CY 1975. Washington, D.C.: U.S. Government Printing Office, 1976.

Privacy Protection Commission. Personal privacy in an information society. Washington, D.C.: U.S. Goverment Printing Office, 1971.