

DOCUMENT RESUME

ED 161 785

95

SO 011 235

AUTHOR Boruch, Robert F.; Cecil, Joseph S.
 TITLE Report from the United States: The Privacy Act of 1974 and the Social Sciences Need for Access to Data.
 INSTITUTION Northwestern Univ., Evanston, Ill. Dept. of Psychology.
 SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.; National Science Foundation, Washington, D.C.
 REPORT NO A-46
 PUB DATE Oct 78
 CONTRACT NIE-C-74-0115
 GRANT APR-77-00349
 NOTE 23p.; Paper presented at the International Conference on Emerging Data Protection and the Social Sciences Need for Access to Data (Cologne, West Germany, August 9-11, 1978)

EDRS PRICE MF-\$0.83 HC-\$1.67 Plus Postage.
 DESCRIPTORS *Confidentiality; Confidential Records; *Data; Data Collection; Disclosure; Ethics; *Laws; Legal Problems; Privacy; *Research Needs; *Social Science Research; Statistical Data
 IDENTIFIERS *Privacy Act of 1974

ABSTRACT

This paper examines the effects of Privacy laws on the conduct of legitimate social research. The report focuses on national legislation dealing with individual privacy, in particular the Privacy Act of 1974, and considers several new statutes which concern the privacy of participants in research. A large portion of the paper discusses (1) regulations concerning access to administrative records and data, statistical microdata, and archived data; and (2) regulations on data handling within social science research. The Privacy Act of 1974 permits individuals to gain access to personal records maintained by federal agencies and to challenge incorrect information. The law restricts the kinds of identifiable information that an agency may collect, and regulates the disclosure of personal information by federal agencies. However disclosure of anonymous records for research purposes is expressly permitted by the law. Following a 1977 meeting between social scientists and the Privacy Protection Study Commission, the Commission recommended disclosure of administrative or research records in identifiable form, for purposes of legitimate research, provided that certain conditions are met. The conditions are generally reasonable, e.g., the disclosure of identifiable records must be necessary to accomplish the research or statistical purpose. (RM)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

ED161785

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Robert F. Boruch

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC) AND
USERS OF THE ERIC SYSTEM."

Report from the United States:

The Privacy Act of 1974 and the Social
Sciences Need for Access to Data¹

Robert F. Boruch and Joseph S. Cecil
Northwestern University
Evanston, Illinois, U.S.A. 60201

Division of Methodology and Evaluation Research

Department of Psychology

Report No. A-46 October, 1978

Presented at: The International Conference on Emerging Data
Protection and the Social Sciences Need for Access
to Data, Sponsored by the Committee of European
Social Science Data Archives (CESSDA) and the
International Federation of Data Organizations (IFDO)
at Cologne, West Germany, August 9-11, 1978.

Supported by: National Institute of Education Contract NIE-C-74-0115
and National Science Foundation Grant APR-77-00349.

SP011235

October, 1978

For CESSDA

A-46

Report from the United States:

The Privacy Act of 1974, and the Social
Sciences Need for Access to Data¹

Robert F. Boruch and Joseph S. Cecil
Northwestern University
Evanston, Illinois, U.S.A. 60201

This report focuses generally on national legislation dealing with individual privacy, notably the Privacy Act of 1974. It also considers briefly several new statutes which deal specifically with the privacy of participants in research. Our primary purpose is to outline the probable and actual effects of these laws on the conduct of legitimate social research. We adhere to an outline which Paul Muller (1978) developed in order to structure the presentation and to facilitate comparison of privacy laws in developed countries.

A. General Characteristics of National Privacy Legislation:
The Privacy Act of 1974

The Privacy Act of 1974 attempts to protect a citizen's right to privacy by regulating the collection, management, and disclosure of personal information maintained by governmental agencies.² It permits individuals to learn of personal information maintained by federal agencies, to gain access to these records, and to challenge information which may be incorrect. It also restricts the kinds of identifiable information that an agency may collect and maintain. Finally, and perhaps most important, the Privacy Act regulates the disclosure of personal information by federal agencies.

The Act is the first attempt by Congress to provide comprehensive protection of an individual's right to privacy at least with respect to federal record systems. Before the Privacy Act, federal data management policy was embodied partly in the Federal Reports Act: record sharing among agencies was encouraged to reduce the burden and expense of reporting. This policy was restricted by statutes or regulations providing for confidentiality of certain record systems, e.g., Title 13 of the U.S. Code, prohibits disclosure of Census records. The Privacy Act makes explicit a new general policy, recognizing the right of an individual to control dissemination of information, he or she provides to the government.

In the United States, as in Sweden, Germany, and other countries, the origins of the legislation are mixed. It stems from vague public fears of computers, numbers, and depersonalization. It stems from concrete verifiable abuses of administrative records by law enforcement authorities and others. It has been stimulated by both liberal and conservative concerns about the amount of information maintained by government and about the absence of general law restricting the disclosure of records. And like all dramatic legislation, it is the product of work by good and conscientious leaders as well as from the work of less conscientious ones.

Unlike recent and pending legislation in other countries, the Act also created a Privacy Protection Study Commission to assess implementation of the law, to identify its shortcomings, and to determine how privacy legislation might be extended to areas beyond federal record systems. The establishment of a temporary Commission in this instance is a remarkable illustration of legislative willingness to anticipate weaknesses in the law, and to create a mechanism for identifying deficiencies and methods of correction.

The origin, structure, and content of the law make clear Congress's strong interest in administrative uses of administrative record systems. Indeed, in examining the legislative hearings leading up to the Privacy Act, we were unable to discover a single reference to researchers' abuse of administrative records or of research records to which they were provided access.

1. The Distinction between Statistical, and Administrative Data

Section 552(a) of the Privacy Act enumerates definitions and distinguishes between "record" and "statistical record." The definition of statistical record hinges on use: it is a record maintained only for statistical research purposes, which is not used for making a determination about an individual. "Record" on the other hand is defined in terms of content: "an item...including but not limited to education, financial transactions..." and containing individual identifying information. The definition of record implies that "record" means both administrative and statistical records. There is no special provision or recognition of the phrases statistical data and administrative data in the current Privacy Act. The report of the Privacy Protection Study Commission should help to remedy these shortcomings.

The Commission report takes pains to recognize the "functional distinction between research records and administrative records. An administrative record is defined in terms of its use for arriving at an administrative decision about an identified individual; a research record is defined as any form of record used solely for research or statistical purpose. The same distinction is made palpable in the Commission's dedication of a major chapter in their report to confidentiality issues and the researcher. The chapter is one of sixteen which deal with topics ranging from privacy problems in employment settings to those in law enforcement agencies. The functional distinction used in the Commission report is nearly identical to one proposed earlier by Boruch (1971).

Special statutes designed to protect confidentiality of social research data, which we discuss later, also define statistical record and research record primarily in terms of their use (i.e., no determinations about individuals) but do not define administrative records explicitly.

2. Definition of Anonymous and Identifiable Records

The Privacy Act bears primarily on "identifiable" records and uses the word identifiable in prescribing conditions under which records may be released, for example. But it provides no explicit definition of the term. The Privacy Protection Study Commission report recognized this weakness and

3

offered a definition of information in individually identified form: "any material that could probably be uniquely associated with the identity of the individual to whom it pertains" (p. 572).

Mindful of the administrative burden on agencies which would result if coverage concerned all identifiable information, Congress restricted the coverage of the Act to records which are contained in a "system of records," further defined as "any group of records from which information is retrieved by the name of the individual or by some number, symbol, or other identifying particular assigned to the individual." This definition encompasses a vast number of the identifiable records maintained by federal agencies, but it does not extend to those record systems in which the information is not actually retrieved by individual identifier.

Merely removing the name or individual identifier from the record usually will be a sufficient guarantee of anonymity to permit disclosure of records under the Privacy Act's exemption for statistical research. However, this may be inadequate if public records are available to aid in identifying the anonymous data. The exemption states that disclosure is permitted only if the record is "in a form which is not individually identifiable." The Office of Management and Budget Guidelines for the Privacy Act interpret this part of the law to mean that the identity of the individual cannot be deduced from tabulations or other presentations of the information. Where such deductive disclosure is a threat, the agency will have to take precautions beyond the simple deletion of individual identifiers if it wishes to disclose records.

The OMB guidelines, which recognize the possibility of deductive disclosure, are an institutional interpretation of the Privacy Act. In this sense, the legislation can be said to recognize the concepts of anonymous and identifiable record; the recognition is imperfect in part because very little is known about the likelihood of deductive disclosure based on (ostensibly) anonymous statistical records or tables.

To accommodate this problem, the Federal Committee on Statistical Methodology, an interagency group, created a Subcommittee on Disclosure Avoidance Techniques and their report outlines definitions for exact deductive disclosure, approximate and probabilistic disclosure. It also describes techniques for dealing with the problem. It extends but fails to recognize earlier work in Sweden by the Central Bureau of Statistics, in Germany by Schlörer and others, and in the United States by Hoffman, Turn, and others in academic research. The Privacy Protection Study Commission references the Federal Committee's work and refrains from any specific statements about deductive disclosure.

3. Provisions for Research or Statistical Utilization

The Privacy Act of 1974 has no detailed provisions for utilization of data for research. However, disclosure of anonymous records for research purpose is expressly permitted by the law. While no similar provision exists to permit access to identifiable information for research, the Act implicitly recognizes the researcher's need for such access (see Section C, below).

Partly because research uses of records were understated in the Privacy Act and ignored entirely in other privacy related legislation, such as the Tax Reform Act of 1976, Boruch (1977) made the following proposal to the Privacy Protection Study Commission. In order to avoid the kinds of unnecessary statistical research related problems engendered by the Privacy Act, by the Tax Reform Act, and by other legislation, bills which may bear on statistical research should be routinely reviewed for their impact on that research. It's not clear, for example, that the 20 or so major medical studies which capitalized on address lists maintained by the Internal Revenue Service over the past 10 years (without negative consequence for the addressee) were even considered in the Tax Reform Act hearings. (See Section C.1). Yet the routine use of lists for worthy socio-medical research has been terminated under the Act accidentally and essentially without a fair hearing. A mandatory "statistical research impact statement" or similar evaluation device should be a part of the routine legislative review process. Moreover, it is not only government statisticians who ought to be involved in the process. Epidemiologists, psychologists, economists, and others outside government must be given a better opportunity to participate in the process than now exists.

The Privacy Protection Study Commission did not exploit the suggestion. It did, however, capitalize on other material from that testimony, on testimony by social scientists such as Flaherty, Gastwirth, and Demareth, and by medical researchers in considering research. The Commission report puts notable stress on the "citizen as research participant," the value of and need for research, the importance of researchers' use of administrative records for research purposes, and the secondary analysis of research information. So, for example, Commission Recommendation #7 provides for disclosure of administrative or research records in identifiable form, for purposes of legitimate research, provided that certain conditions are met. The conditions are generally reasonable, e.g., the disclosure of identifiable records must be necessary to accomplish the research or statistical purpose.

B. Regulations Concerning Data Collection

1. General Regulations

To guard against the promiscuous collection of information, the Act requires that agencies shall maintain "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 executive order of the President." In the context of record systems used for research, this means that the records must be linked to such authorization or the record system will be in violation of the Act. Since much of social research is exploratory in nature, the narrow interpretation and strict enforcement of this requirement could sharply limit the collection of social science data.

Each federal agency is responsible for interpreting and applying this section of the Act to its own record systems. Unlike Sweden's Privacy Act, the initiation of a new data collection effort need not be reviewed by a central Data Inspection Board. Within an agency such as the U. S. Office of Education, however, other no less influential committees of independent

reviewers, set up under other laws, may review the need for a new research project. In addition, peer review of research proposals submitted for funding to the federal government are routinely reviewed for their importance.

The Privacy Protection Study Commission found that this constraint on information collection has had very little effect on maintenance of agency records. It found that, in general, agency practices have not changed and there has been no great purging of existing record systems. However, there are indications that the restrictions may have had some effect by discouraging agencies from collecting new information. If new information collection programs are being discouraged, this could be especially harmful to research activities. It's not clear that the information available to the Privacy Commission was adequate to determine the restriction's effect. No time series data are presented; no case studies of any import are discussed; no competing explanations for the cancellation of a record system are identified; no sampling scheme for the Commission's survey of record systems is established. It's possible that the Commission's investigations on relative stability of record systems were premature.

Five factors help to mitigate the potentially harsh effect of this requirement applied to research record systems. First, there is no central review board of laymen to register judgments about whether scientific data is necessary or relevant. Such a board would doubtless be as disruptive in the United States as it has been in Sweden. Second, review bodies which include both researchers and the lay public already review research proposals in the special areas governed by each federal agency. Though they are imperfect, though their performance is not always well documented, the boards have had sufficient experience and influence to make additional reviews unnecessarily redundant. Third, some research projects have received explicit statutory authorization, or implicit authorization in the statutory statement of the agency purpose. Such record systems clearly meet the statutory requirements of the Act. For example, the National Institute of Education is mandated to support educational research, the U.S. Office of Education is often required to evaluate educational programs, the National Center for Educational Statistics is dedicated to compilation of statistical data on the educational state of the nation. All these missions are predicated on the assumption that some information is essential for improving the quality of education in the United States. Future record systems are more likely to receive explicit authorization now that this Privacy Act requirement has been enacted. The fourth influential factor is that the head of an agency may exempt from this requirement those record systems "required by statute to be maintained and used as statistical records." (emphasis added). This statutory language has been interpreted by some agencies to protect only those systems of records which are prohibited by statute from being used to make a determination about an individual. Not only must the record be restricted by statute from use for administrative purposes, it must also be protected from compulsory legal process. There are very few statutes which offer such absolute protection to research records, and these are discussed in Section F, below. Even these research record systems have been exempted only from the notice, access, and correction requirements

of the Act and must still comply with the provision requiring information to be relevant and necessary.

The second major general requirement of the Privacy Act is that an agency gathering research information directly from individuals must inform each individual of the authority which authorizes the solicitation of the information, the purposes for which the information is to be used, whether disclosure of such information is mandatory or voluntary, and the consequences to the individual of refusing to provide the requested information. These disclosures, commonly referred to as the "Privacy Act Statement," are in response to past complaints that agencies have used pressure, deceit, and coercion to gain personal information. While these provisions will encourage the collection of information with the knowing and informed consent of the individual respondents, the effects of such information on cooperation rate in research have only recently become a topic of methodological investigation. Singer (1978), for example, has initiated very nice experiments to assay the effect of amount and type of information on response rate and candor.

2. Cross-sectional, Longitudinal, and Observational Studies

No provisions of the Privacy Act or of any other privacy legislation distinguish among cross-section, longitudinal, and observational studies. The Privacy Protection Study Commission Report defines "longitudinal research" though no other distinctions are drawn. Their definition hinges on establishing "how the state of a group varies and the average relation between states" over time, and closely follows a definition in Dalenius and Klevmarcken (1976).

C. Regulations on Access to Administrative Records

1. Access for Sampling Purposes

The Privacy Act's prohibition on disclosure of identifiable information without the prior written consent of the individual could sharply restrict the use of administrative and other records for sampling purposes. In particular, researchers occasionally seek access to agency record systems to obtain a sample of individuals for anticipated research or to supplement existing research information. The consent requirement could interfere with both of these activities.

When seeking to supplement existing research data with information from agency records, researchers will often be able to anticipate this need and obtain the informed consent of their research participants at the time the research information was collected. However, if the need for access to agency records was not anticipated, the researcher may have to recontact the individual to obtain consent. Even if the individual agreed to subsequent release of agency information when the research information was initially collected, if the researcher seeks access to research records which did not exist at the time consent was obtained, the earlier consent may be invalid and it may be necessary to recontact the individual to obtain proper consent. Recontacting a participant in an earlier research study

7
imposes special difficulties. Some target populations are highly mobile, so that addresses and telephone numbers obtained at the initial encounter may be outdated. Some target populations are difficult to recruit for research, so that subsequent attempts to obtain consent to release agency information will likely be expensive and subject to biases in self-selection.

More formidable obstacles are faced by researchers who wish to use agency records to generate a sample of identified individuals to be contacted for participation in anticipated research. Since the purpose of obtaining access is to obtain a list of names and addresses of individuals, the research will be unable to obtain consent for release of this information. Researchers employed by the agency maintaining the records may avoid the consent requirements by demonstrating a need for the record in the performance of their duties. But researchers outside the agency have found the consent requirement to be a frustrating hurdle. For example, the Medical Follow-Up Study, sponsored by the National Academy of Sciences, demonstrates the difficulties the consent requirement poses for non-federal researchers who must rely upon agency records to generate a research sample. The Medical Follow-Up Study is a longitudinal research program on the health and medical status of U. S. military veterans. The program, which began in 1946, was designed to capitalize on the enormous pool of baseline medical data collected as part of the normal military processing of servicemen. Those data have been used to track the natural history of a variety of diseases, especially those diseases which occur too infrequently to be studied through customary sampling procedures. Since 1973, the name and addresses of veterans selected for the Follow-Up Studies were obtained from the Taxpayer Address Request Program of the Internal Revenue Service. With current addresses, researchers then contacted the veterans to explain the purpose of the research and request their consent for release of medical information. In April 1976, the IRS suspended this service, ruling that the Privacy Act prohibits the disclosure of current addresses without the consent of the individual. Without current addresses, these extensive follow-up studies will be terminated.

Because of such problems, the Privacy Protection Study Commission dedicated special attention to researcher use of administrative records, including use of sampling (p. 588). Several recommendations are offered to resolve problems: Recommendation #7 permits automatic disclosure of administrative records, without individual authorization, but under certain conditions, for research purposes; Recommendation #9 prohibits redisclosure of the information by the researcher unless the same conditions are met.

Regardless of the Commission's recommendations, it is sometimes possible to employ special procedural strategies to circumvent prohibitions against disclosure of individual identification. For example, the archive may be asked, by the researcher to sample the records in accordance with instructions and to forward questionnaires to the individuals and to ask individuals to return questionnaires to the researcher. There is a wide array of such procedures, but many increase the cost or difficulty of research, and may not be feasible in the particular research setting.

2. Access as a Data Base

In general, such access to identifiable records is permitted for outside researchers only if the individual provides consent for disclosure of the record. Disclosure of records by one agency to another for research may be permissible under the "routine uses" provisions of the Act. The latter exemption is ambiguous, however, and a novel research use may be refused at the discretion of the agency.

The changes in the Privacy Act proposed by the Privacy Protection Study Commission would acknowledge the propriety of access to identifiable records for research purposes, permitting broader access to and adopting formal measures to insure the consent of the individual who provides the information. A major factor in the Commission's willingness to permit greater access for research purposes is the protection which would be accorded research records if the Commission's recommendations for a functional separation of research and statistical records are adopted. Since such a separation would insulate the research record from uses which would be detrimental to the individual, the Commission was willing to go quite far in permitting access to identifiable records for research purposes. However, the Commission would permit access to identifiable information from administrative files, for research purposes, only when it is absolutely necessary (Recommendation #7), would require interagency agreements on the matter (#8), and would prohibit redisclosure (#9).

3. Access to Administrative Data for Linkage.

A Privacy Act exemption which may assist the social scientist seeking identifiable information permits the transfer of identifiable agency records to the Bureau of the Census for planning, or carrying out a census, survey, or other related activity. The law will permit linkages between agency files when conducted by the Bureau of the Census for some purpose such as to establish the credibility of these alternative sources of information. Furthermore, this exemption would seem to allow the Bureau to accept identifiable agency records to perform certain statistical analyses for researchers outside the Bureau who are unable to gain access to these records. Presumably such analyses are a "related activity" under the exemption, and the Bureau will continue to act as a brokerage agent to perform such research.

Aside from this special provision, the Privacy Act dedicates no attention to the need for record linkage or its implementation. Because record linkage, between administrative and research systems or between two research systems, can be important in research and because the Act's inattention to the matter provokes confusion, the Privacy Protection Study Commission discussed the matter. It refers specifically to data sets which are "interlinked" for research purposes (p. 586) and the existence of procedures for linkage, such as brokerage or insulated file linkage, which accomplish research goals without breaking privacy. Again, to clear the way for linkage related uses of administrative records, the Commission's recommendations on disclosure (#7), interagency agreement (#8), and redisclosure (#9) should be helpful to researchers both within and outside government.

Irrespective of the Commission recommendations, the procedural approaches to linking records from different archives without breaching privacy regulations governing disclosure still appear appropriate. See Section F.2 below and Boruch and Cecil (1978).

4. A General Exemption which may Permit Access for Sampling, Data Bases, and Linkage.

A major Privacy Act exemption which might benefit social scientists permits disclosure of an identifiable record for a "routine use" of such a record. "Routine use" is defined as a use "for a purpose which is compatible with the purpose for which (the record) was collected." Such ambiguity in statutory language suggests that an agency might be free to define "statistical analysis" as a routine use of all or a selected portion of agency record systems, permitting researchers outside the agency to have access to identifiable records without gaining the consent of the individuals to whom the records pertain. In fact, almost all agency notices allow for disclosure involving statistical research programs as a routine use. If this designation is extended to the record system of interest to the social scientist, the researcher may have access to the identifiable agency record system for research purposes without the consent of the individuals. Such a special exemption may be helpful to social science research, but the legality of such an exemption may be questioned. A House Committee Report (H. R. Rep No 93-1416, 93rd Cong., 2nd Sess. 12 (1974) recognized the potential for abuse of the routine use exemption, and promised vigorous oversight of agency practices in this area. If this designation has been extended to a sufficient number of files to result in a general exemption for social scientists, it may be tested in court and found in violation of the Privacy Act.

5. Summary on Access for Sampling, Data Bases, and Linkage

The restrictions of the Privacy Act are sufficient to interfere with access to agency records for subsequent research activities. Unless the researcher can invoke one of the exceptions to the Privacy Act, access to identifiable agency records is forbidden unless the consent of the individual is obtained. While exceptions to the Privacy Act may restrict protection of confidential research records, they also represent the only means by which researchers may obtain access to regulated record systems. The exception most likely to aid researchers permits agencies to disclose information for purposes of statistical research if the record is transferred in a form which is not individually identifiable. In fact, records which are not individually identifiable are not regulated by the Privacy Act. This suggests that the statistical and procedural techniques such as insulated file linkage, which permit meaningful analysis of data while preserving the anonymity of the respondents, can be alternative analytical techniques when the Privacy Act restricts the release of identifiable records.

D. Regulations Concerning Access to Statistical Microdata

1. Public Use Samples

There are no special provisions in the Privacy Act which encourage or require production of public-use samples. The Act does permit public use samples in the general sense that disclosure of anonymous records and statistical data for research are expressly permitted. Further, records which are used solely for statistical research purposes are exempted from many portions of the Act, notably those which give individuals the right to access their own record.⁴

2. Integration of Microdata Files

The Act's main focus is on preventing unnecessary transfer of identifiable administrative records across agency boundaries. The Act may impede some linkages among research data archives.

Disclosure of statistical records is expressly permitted under the Act whether the records stem from administrative archives or research archives. The main qualification on disclosure is that the records contain no individual identification (see Section A.2). As a matter of practice then, this means that statistical records may be transferred from one government agency to another, or from government to the independent research. Once the transfer is made, limited integration of different microdata files is possible. The integration is limited simply because the records contain no individual identifiers. For example, government agency A or its contractors may possess information on education and employment of individuals in specific school districts during 1960-1970. Agency B maintains similar information for 1970-1980. The integration of statistical data based on school districts, rather than individuals within school district, is permissible and possible. But the resultant integrated file limits analyses to those characteristic of work on microaggregated samples.

Integration of anonymous records on the same or different individuals is also permissible under this disclosure provision of the Act. Such integration has a relatively recent history however. The statistical research on synthetic matching and on methods for linking anonymous records on the same individuals is still developing. The techniques are complicated and generally produce a kind of analogue or imperfect surrogate for an integrated record based on clear identification. See, for example, Moore, Kuhns, Trefftz, and Montgomery (1977).

Privacy Act provisions which permit linkage of records for routine use and the special exemptions for transfers of records to the Census Bureau facilitate some efforts to integrate microdata. In fact, because identifiers can be used as a basis for integration in some of these cases, the process is no more than record linkage for research purposes. Jabine (1975), for example, observes that integration of Social Security data with data held by Census Bureau, by SSA contractors, and other projects can be accomplished legally through disclosure of SSA records (e.g., identifiers only or identified information). The Social Security research on quality and character of reporting requires linkage in these and other projects, and such research is routine. However, Jabine expresses some concern that some integration projects may be canceled if "deemed not directly related to administration of the Social Security Act".

E. Regulations on Preservation and Access to Archived Data

1. Public Archives

The Privacy Act refers to transfer of identifiable records to the National Archives only if the record is of sufficient historical merit. Disclosure rules applying to the originating agency then must apply to the National Archives. The Act's provisions for disclosure of statistical data apply to the National Archives as they apply to any other agency or individual.

The National Archives has recently undertaken to acquire and store statistical (anonymous) records stemming from federally supported social program evaluations. Its enabling legislation gives the National Archives the right to do so, but no national legislation, nor provision of the Privacy Act, requires the National Archives to store such research data. Routine storage will of course facilitate secondary analysis of statistical data.

Other archival sources of statistical data exist, notably the public use data tapes made available by the U. S. Census Bureau, the Social Security Administration, National Center for Health Statistics, and other public and private agencies. See Hedrick, Boruch, Ross (1978) and the references therein. The Privacy Act appears to be relevant to these only in the sense of generally permitting disclosure when identifiers are excluded.

2. Conditions of Access

Generally speaking, access to statistical data or anonymous records maintained by the National Archives and others mentioned above is conditional on written request and nominal payment for costs of tape reproduction or file documentation. Where identifiable records are at issue, the rules for researchers gaining access are determined by the agency maintaining the records.

Where staff of an archive believes that disclosure of anonymous records or a sample of anonymous records will jeopardize individual privacy, then the Privacy Act will not require release of the information.

F. Regulations on Data Handling Within Social Science Research

1. Confidentiality Assurances

The Privacy Act of 1974 makes no unusual demands of social scientist in handling research data. To rectify the inattention given to research records, the Privacy Protection Study Commission offered several recommendations on confidentiality assurance. The most important (Recommendation #1) require that information collected for research purposes be used only for that purpose and not for making determinations about individuals. Recommendation #2 buttresses this, suggesting that legislation be created to assure that research records cannot be appropriated for nonresearch use. Recommendations #4 and #5 put time constraints on maintenance and require the research to set up physical confidentiality assurances.

Recommendations #1 and #2 of the Commission report have precedent in new law which extends special legal protection to the social research and the research participant. We describe several statutes which provide testimonial privilege to the researcher below, adapting description from Boruch and Cecil (1978):

The statute most frequently cited as providing protection for non-governmental social science research is found in §303 (a) of the Public Health Services Act (Pub. L. No. 91-513) as amended in 1974 (Pub. L. No. 43-282, encoded as 42 U.S.C.A. §242a). This statute permits the Secretary of Health, Education and Welfare to:

authorize persons engaged in research on mental health including research on the use and effect of alcohol and other psychoactive drugs, to protect the privacy of individuals who are the subject of such research by withholding from all persons not connected with the conduct of such research the names or other identifying characteristics of such individuals. Persons so authorized to protect the privacy of such individuals may not be compelled in any Federal, State, or local civil, criminal, administrative, legislative, or other proceedings to identify such individuals.

The scope of protection offered by this statute is still open to interpretation. The proposed federal regulations (40 Federal Register 234 (Dec. 4, 1975, pp. 56692-56695) do not define "mental health research," but do define "research" broadly to include behavioral science studies, surveys, and evaluations. Within these areas, unless the participant agrees to the release of the information or release is required by a medical emergency, a researcher who has received the authorization of the Secretary of Health, Education and Welfare, may refuse to comply with a subpoena issued by any branch of government.

For the past several years, however, the Attorney General has exercised a similar discretionary authority to extend an evidentiary privilege to drug research projects. Section 502 (c) of the Controlled Substances Act (Pub. L. No. 91-513, encoded as 21 U.S.C.A. 872 (c)), empowered the Attorney General to:

... authorize persons engaged in research to withhold the names and other identifying characteristics of persons who are the subjects of such drug abuse research. Persons who obtain this authorization may not be compelled in any Federal, State, or local civil, criminal, administrative, legislative, or other proceeding to identify the subjects of research for which such authorization was obtained.

This statute is very similar to the HEW statute. Fortunately, the procedures for awarding this privilege have been in existence long enough to permit some general observations about the manner in which this administrative discretion has been exercised. (see Hedrick, Nelson, Cecil, 1978).

Research data concerning criminal activities are especially vulnerable to forced disclosure by court-ordered subpoena. The unwillingness of the courts to exercise their discretion to bar researchers' evidence of criminal activity has resulted in federal statutory protection for criminal justice research. Section 524(a) of the Crime Control Act of 1973 (Pub. L. 93-83) provides that identifiable research and statistical information may only be used for the purpose for which it was obtained, and without the consent of the individual cannot be admitted as evidence or used for any purpose in any judicial or administrative proceeding. The statute provides:

Except as provided by Federal law other than this chapter, no officer or employee of the Federal Government, nor any recipient of assistance under the provisions of this chapter shall use or reveal any research or statistical information furnished under this chapter by any person and identifiable

to any private person for any purpose other than the purpose for which it was obtained in accordance with this chapter. Copies of such information shall be immune from legal process, and shall not, without the consent of the person furnishing such information, be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceedings. (Encoded as 42 U.S.C.A. §3771.)

The implementing regulations for the statute (41 Federal Register 54816, December 15, 1976) acknowledge that one of the purposes of providing such protection for criminal justice research is to increase the credibility and reliability of federally supported research and statistical findings by minimizing subject concern over subsequent uses of identifiable information.

The scope of protection under this statute is quite broad, extending to research and statistical information obtained either directly or indirectly by the Law Enforcement Assistance Administration, or under any agreement, grant, or contract awarded under the Omnibus Crime Control and Safe Streets Act. Such a relationship can be demonstrated for almost all federally sponsored criminal justice research. Protected information may be used only for statistical or research purposes and may not be used as evidence in judicial or administrative proceedings. However, the statute leaves research information vulnerable to a legislative subpoena, an exception which the Law Enforcement Assistance Administration considers to be a shortcoming (Madden, 1977). Also, the statutory protection does not extend to information concerning future criminal conduct.

A number of other federal statutes offer protection to record systems which may occasionally be used in social scientific inquiries. Some statutes offer protection for special categories of data. Special statutes protect census data (13 U.S.C.A. §§8, 9) and social security data (42 U.S.C.A. §1306), but these protections extend only to federal researchers and to outside researchers hired temporarily as agency personnel.

Another form of federal statutory protection extends to specific topics of research. Federal surveys concerning venereal disease have received statutory protection. Information derived from venereal disease treatment programs can be released for statistical or research purposes only if the identity of the individual provided care under a treatment program is not disclosed (42 U.S.C.A. §247c(3)(5)).

Statutory protection has also been provided for specific research projects commissioned by Congress. For example, when authorizing a comprehensive statistical survey of runaway youth, Congress indicated that identifiable records gathered in the course of the survey could not under any circumstances be disclosed (42 U.S.C.A. §5732).

2. Technical and Organizational Measures

The Privacy Act dedicates no special attention to technical and organization measures for handling social science data. The handling practices which it alludes to are general, e.g., rules on disclosure are general rather than specific. The Privacy Protection Study Commission recognized specific topics in this context, however. Their report recognized and defined concepts such as randomized response, microaggregation, random error injection and the like, but provided no references. In the section on recommendations about information collected solely for statistical research purposes, the Commission Report proposes: notifying respondents that records may be disclosed for secondary research purposes (Recommendation #10); using institutional review boards to oversee collection of information from those incapable of providing voluntary consent (#11).

Neither the Act nor the Privacy Commission recommendations appear to affect the use of conventional methods for assuring confidentiality and data handling. Those methods, summarized in Boruch and Cecil (1978), include:

Procedural Approaches. For longitudinal data collected periodically within the same framework, the simple device of using alias identifiers is obvious but underutilized. The alias may be created by the respondent and used consistently in response to permit intrasystem linkage. It may be created by social scientists, provided to the respondent, then purged from the social scientists' files to achieve the same ends. To decentralize the process, some neutral brokerage agency (a census bureau, a nongovernmental agency) may similarly create an alias for the respondent and destroy its own records of any linkage between clear identification and alias.

The strategy has been field tested with some success in U.S. drug studies, political attitude surveys and the like. Aside from longitudinal problems, its major shortcomings are the limitations imposed on linking the data elicited under alias with any other existing data on individuals.

To accommodate some logistical problems as well as the limitation on intersystem linkage, procedures such as the link file system have been developed. In this technique, a dictionary of double aliases is created by the social scientist and given over for safekeeping to an independent agency. The decentralization of the process enhances physical security, and if the agency is legally entitled to resist governmental appropriation of files, the procedure is legally secure. The dictionary is used as a basis for linking information which is periodically obtained from individuals. The main benefit of the strategy is that it reduces the social scientists' need to maintain longitudinal records on identified individuals, in general, and it reduces the time during which the social scientist has access to any given wave of data containing identifiers to an arbitrarily short period (see Astin & Boruch, 1970).

For those cases in which records from different archives must be linked, a variety of methods have been developed to permit linkage without violating the customs or law governing linkage. Among the better known systems for doing so is the "mutually insulated file approach," used in the Schwartz-Orleans (1967) referred to earlier. Basically, the system involves two files of records operated under different auspices; all records

are identified and there is some overlap between the samples of individuals on which the records are maintained. To accomplish the linkage, the first archive (assume it is the social scientist) cryptographically encodes the information portion of each record, producing a new file without meaning to any outsider, which is then transmitted to the record archive. The archive then matches the encoded records with its own records, based on the clear identifiers appearing in each record. Upon completion of the match, identifiers are deleted and the linked records are returned to the social scientist who then decodes relevant portions of the linked records and conducts his statistical analysis of the anonymous records. Variations on the strategy are described in Boruch (1972) and Campbell et al. (1975).

Statistical Approaches. The devices just described are most often relevant to more impersonal forms of observation--questionnaires and the like--rather than to direct interview research. And in some instances, the logistical difficulties attached to their use are considerable. Partly for these reasons, it may be more appropriate to capitalize on one of the statistical strategies which have been developed to reduce depreciation in privacy. A variety of these approaches exists and these may be used alone or in conjunction with the procedural devices.

The best known class of approaches is the randomized response tactic currently under test and development by Greenberg and others (1974) in the United States, Dalenius (1975), Lanke, Swensson, Svensson, and Eriksson in Sweden; Warner in Canada, Moors in Holland, and others. In the simplest variation of the approach, the social scientist simultaneously presents a sensitive inquiry to an individual, e.g., "Did you cheat on your income taxes this year?" and an insensitive one, e.g., "Do you prefer potatoes over noodles?" The individual is then instructed to roll a die and respond to the first question if a one or two shows up, and to the second question if a three, four, five, or six shows. He is also told to refrain from giving the interviewer any indication of which question was answered. When the process is carried out on two large samples of individuals and the instructions are followed by the respondent, it is possible to estimate the proportion of individuals in the sample who have cheated on their income tax forms and the proportion who prefer noodles. In particular, given some simple laws of probability, the odds on answering one or the other question, and the observed proportion of Yes-responses, the estimation is a matter of simple algebra.

The technique permits us to establish the statistical character of sensitive properties of groups of individuals. And moreover, it does so without disclosing to the social scientist any information about a particular individual. It has been field tested in drug studies, in fertility control studies and other areas, and those tests continue in the U.S., Canada, Sweden, and elsewhere. The basic method is being refined to make it more efficient in a statistical sense, more acceptable to the respondent in a social psychological sense, and less vulnerable to corruption in a legal sense.

Another class of statistical techniques which has received some attention is aggregation of the sample. The technique requires that one obtain data not on single identified individuals but rather on very small and carefully constructed clusters of individuals. If the cluster's composition remains the same over time, each cluster can, under certain conditions, be regarded as a synthetic person, a composite of all the properties of the small set of individuals it comprises. Some informative data analyses can be conducted on those aggregates and, insofar as aggregation helps to assure anonymity of individual response, there is no depreciation of individual privacy.

The applications of sample microaggregation have so far been limited to economic research on commercial units. Banks, for example, may be reluctant to release information about their operations to any outside economist. They are willing, however, to have the social scientist analyze aggregates of banks in the interest of reconciling bank privacy with future research. And indeed, a major system of data maintenance and dissemination has been built up on this theme by the University of Wisconsin (see Bauman, David, & Miller, 1970).

G. Data Protection Within Social Science: Codes of Ethics

The majority of the social science professional groups in the United States have developed codes of ethics for their membership. Major exceptions include the American Economic Association and the American Statistical Association. Each of the codes are described in a monograph by Robert Bowers (1978).

It is unlikely that researchers' codes of ethics played any major role in formal or informal discussions leading up to the Privacy Act: research itself was not given notable attention. The Privacy Protection Study Commission did elicit testimony from major professional organizations, but codes of ethics were not discussed in any detail. Testimony from the American Sociological Association did emphasize the researcher's ethical responsibility to assure that researcher assurances of confidentiality were in fact maintained by the researchers. The effect of this emphasis, or of sporadic reiteration of ethical codes to potential interrogators of research records and to legislators is still unclear.

H. Implementation of Data Protection Laws

1. Specialized Personnel

There are no special provisions in the Privacy Act bearing on agency "privacy officers," i.e., the individual within agency who takes primary responsibility for meeting demands of the Act. Nor does the Act itself require officers. Similarly, no other privacy legislation deals directly with special personnel. Officers within agency have been designated to oversee implementation of the Act, to meet the Office of Management and Budget guidelines on implementation.

In fact, we know very little about the selection of individuals with responsibility for assuring adherence to the Privacy Act or to any other privacy legislation. We know less about their performance, though it's clear that in some instances, the officer's lack of skill and knowledge about statistical social research has led to a conservative posture and even truculence toward the research. It's clear that some officers feel uncomfortable enough about the ambiguity of their role of their agency's policies to meet periodically to exchange views. Nothing is being done to permit officers to learn from one another in a formalized way, through workshops for example, or to consolidate their experiences, through a regularly issued newsletter, for example, if complicated law is to be implemented well (Boruch; 1977).

2. Dominant Bureaucratic Behavior

There are several stereotypical behaviors.

Generally, agency staff are more cautious about attending to privacy matters. Regardless of the setting or outcome, the caution takes time and negotiation; consequently some new research takes more time to initiate. At best, a few agency staff are quite aggressive in understanding access regulations, and trying to assure that the researchers' needs are met in the confines of the act.

At the other end of the spectrum, a few privacy officers are sufficiently intimidated by the legislation's emphasis that saying "No" to researcher requests is almost automatic. Under these conditions, the appeals process available under the Freedom of Information Act can be helpful. For instance, we recently requested that the Drug Enforcement Administration supply us with the names of researchers who had been granted testimonial privilege under DEA auspices. A low level officer in DEA refused initially to supply the names, claiming that disclosure would violate the individuals' privacy. We appealed that decision under FOI and, after 10 months, received the list.

3. The Effect of General Rules on Reducing Discretion

The privacy laws and the regulations do reduce bureaucratic discretion in restricting access to identifiable data. They may also help incidentally to impede access to statistical data though that is not the purpose of the law. In this sense, the laws increase discretion by providing an excuse for refusing to disclose data.

For example, a large school district recently refused to permit a researcher's sampling of anonymous records from school files. The researchers were interested in assessing the quality of record contents, such as accuracy in diagnosis, especially on minority group students enrolled in special (Federally supported) programs. The school's refusal to provide a sample of anonymous records was based on the Privacy Act and on other privacy law, but was countermanded by the courts. In this and similar cases, the Privacy Act becomes a convenient vehicle for insulating the institution from legitimate research and investigation, rather than protecting the individual.

Discretion for government managers may also be reduced indirectly through increased demands on their time. Jabine (1975), for example, and others observe that "several more steps have been added to already complex clearance procedures for initiating and conducting surveys," that "access and accounting requirements will cause delays in normal processing activities" as a consequence of the Privacy Act of 1974 in the Office of Research and Statistics of the Social Security Administration. That reduction in discretion may reduce the quality of data or operations, at least temporarily, since privacy legislation includes no additional financial support or staff for meeting its requirements.

Footnotes

1. Prepared for the International Conference on Emerging Data Protection and the Social Sciences Need for Access to Data, sponsored by the Committee of European Social Science Data Archives (CESSDA) and the International Federation of Data Organizations (IFDO), Cologne, Germany, August 9-11, 1978. This paper was developed with support of a grant from the National Science Foundation (NSF-APR 77-00349). Our background research on privacy in social research has been supported by the National Institute of Education (NIE-C-74-0115).
2. The most thorough analysis, to date, of the research implications of the Privacy Act is given by Cecil (1978) from which some of the current material is adopted. Papers on the topic have been developed by executives of federal archive agencies such as the U.S. Census and the Social Security Administration; see Duncan (1975) and Jabine (1975).
3. One model in which the Census Bureau serves as broker for the Bureau of Labor Statistics is clearly still legitimate under the Privacy Act. Census obtains employment information, on behalf of BLS, from identifiable individuals in its Current Population Surveys. The Bureau strips resulting records of identifiers and turns the microdata over to BLS for analysis (Yates, 1974). The linkage of samples of various types of records, from the Internal Revenue Service, Social Security, and Census, for example, also remains legitimate under the Act; the Census Bureau also serves as broker in this research on consistency of information from different sources (Jabine, 1975).
4. The Act's posture toward disclosure of anonymous records for research purposes is consistent with the policy of various U.S. agencies. Public use tapes are available from, among others, the U.S. Census Bureau, the Social Security Administration, and the National Center for Educational Statistics. The spirit of the Act is also consistent with the Bellagio principles (Flaherty, 1978) which encourage the use of government microdata, in anonymous form, for research purposes. Both the Act and the Bellagio principles are compatible with more recent efforts by some federal agencies to assure that data stemming from federal program evaluations are available, with identifiers removed, for secondary analysis. Those agencies include the National Institute of Education which has made data from the Educational Vouchers Experiment and others available. And it includes the Law Enforcement Assistance Administration which has an explicit policy on disclosure of statistical data stemming from any LEAA supported project.

References

- Astin, A. W. & Boruch, R. F. A link file system for assuring confidentiality of research data in longitudinal studies. American Educational Research Journal, 1970, 7(4), 615-624.
- Bauman, R. A., David, M. H., & Miller, R. F. Working with complex data files. In R. L. Bauman (Ed.) Data bases, computers, and the social sciences. New York: Wiley-Interscience, 1970.
- Boruch, R. F. Maintaining confidentiality in educational research: A systemic analysis. American Psychologist, 1971, 26, 413-430.
- Boruch, R. F. Statishe und methodudishe prozeduren zur sicherung der vertraulichkeit bei forshung. In A. Eser and K. F. Schumann (Eds.) Forschung im konflikt mit recht und ethik. Stuttgart: Ferdinand Enke Verlag, 1976, 396-436.
- Boruch, R. F. Testimony on the Privacy Act. In: Hearings of the Privacy Protection Study Commission on Social Research and Statistics. Washington, D.C.: Privacy Protection Study Commission, January 5, 6, 1977.
- Boruch, R. F. & Cecil, J. S. Report to Science, Technology and Policy Office, National Science Foundation, on the impact of the Privacy Act of 1974 on social research. Psychology Department, Northwestern University, Evanston, Illinois, 1977.
- Boruch, R. F. & Cecil, J. S. Assuring confidentiality in social research. Philadelphia, Pa.: University of Pennsylvania Press, 1978 (in press).
- Bowers, R. and others. Ethics in social research. New York: Praeger, 1978 (in press).
-
- Campbell, D. T., Boruch, R. F., Schwartz, R. D., and Steinberg, J. Confidentiality preserving modes of access to files and to interfile exchange for useful statistical analysis. Evaluation Quarterly, 1977, 1(2), 269-299.
- Cecil, J. S. Regulation of research record systems by the Privacy Act of 1974. Research Report NIE-01C, Department of Psychology, Northwestern University, Evanston, Illinois, 1978.
- Dalenius, T. The invasion of privacy problem and statistics production: An overview. Sartryck ur Statistisk Tidskrift, 1974, 3, 213-225.
- Dalenius, T., & Klevmarck, A. (Eds.) Privacy protection and the need for data in the social sciences. Stockholm: Swedish Council for Social Science Research, 1976.
- Duncan, J. W. The impact of privacy legislation on the federal statistical system. Public Data Use, 1975, 3(1), 51-53.

Federal Committee on Statistical Methodology, Subcommittee on disclosure-avoidance techniques. Report on statistical disclosure and disclosure-avoidance techniques. Washington, D.C.: U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, (May) 1978.

Flaherty, D. H. Final report of the Bellagio Conference on Privacy, Confidentiality, and the Use of Government Microdata for Research and Statistical Purposes. Federal Statistical Reporter, 1978, 78-8 (May), 274-279.

Hedrick, T. E., Boruch, R. F., & Ross, J. On ensuring availability of evaluative data for secondary analysis. Policy Sciences, 1978, in press.

Hedrick, T., Nelson, R. L., & Cecil, J. S. Grants of confidentiality: A survey analysis of a statutory researcher-subject privilege. In R. F. Boruch (Ed.) Solutions to ethical and legal problems in social research. Report, submitted for publication. Psychology Department, Northwestern University, Evanston, Illinois, USA, 1978.

Horvitz, D. G., Greenberg, B. G., & Abernathy, J. R. Recent developments in randomized response designs. In J. N. Srivastava (Ed.) A survey of statistical design and linear models. Amsterdam: North-Holland, 1975, 271-285.

Jabine, T. B. The impact of new legislation on statistical and research uses of Social Security Administration Data. Proceedings of the American Statistical Association: Social Statistics Section. Washington, D.C.: ASA, 1975, Pp. 221-230.

Kruskal, W. B. Taking data seriously. Chapter 6 of Elkana, Y., Lederberg, J., Merton, R. K., Thackray, A., & Zuckerman, H. (Eds.) Toward a metric of science? The advent of science indicators. New York: Wiley, 1978, pp. 139-170.

Moore, G. B., Kuhns, J. L., Trefftz, J. L. and Montgomery, C. A. Accessing individual records from personal data files using non-unique identifiers. Washington, D.C.: U.S. Government Printing Office, 1977.

Muller, P. J. Datenschutz und sicherung der individualdaten der empirischen sozialforschung. In Eser and K. F. Schumann (Eds.) Forshung im konflikt mit recht und ethik. Stuttgart: Ferdinand Enke Verlag, 1976, 437-454.

Muller, P. J. Memo to all country rapporteurs...2. The form and substance of country reports. University of Cologne, Cologne, Germany, February 27, 1978.

Privacy Protection Study Commission. Personal privacy in an information society. Washington, D.C.: U.S. Government Printing Office, 1977. (Stock No. 052-003-00395-3).

Singer, E. Informed consent: Consequences for response rate and response quality in social surveys. American Sociological Review, 1978, 43, 144-162.

Yates, P. B. Implications of the Freedom of Information and Privacy Acts for Voluntary Labor Statistics Programs. Proceedings of the American Statistical Association: Social Statistics Section, Washington, D.C.: ASA, 1975, pp. 217-220.