

This paper explores the complex legal and other issues which arise when employing administrative records for statistical purposes. That inquiry involves two principal parts.

The first part undertakes to examine the interests and needs of statisticians which lead them to use information contained in administrative records. In this part of the paper, section 1 points out reasons for the statistical use of administrative records. Statistical use and administrative use are defined, differentiated, and illustrated in section 2, along with terms relevant to legal issues. This leads into the concept of functional separation which is described in section 3 as an analytical tool for data usage, and is formulated in section 4 in general legislative terms as a way to realize the conceptual goal.

The second part of the paper uses a characterization of the existing legal and administrative system as a frame of reference, and suggests a set of organizing principles in which legal and statistical imperatives converge. Section 5 deals with the difficulties associated with the actual application of functional separation concepts to government agency records. Section 6 discusses the application of several confidentiality statutes to particular situations.

Finally in section 7, a brief summary of the paper is provided along with some suggestions for the future.

THE LEGAL AND ADMINISTRATIVE SYSTEM

1. Factors precipitating the shift toward greater statistical use of administrative records

Increased availability of administrative records, and growing limitations on information obtainable directly from individuals on a voluntary basis, have both precipitated a shift toward greater statistical use of program and other administrative records.

Advanced data processing techniques and sophisticated methodologies extract, distill, and illuminate information from massive volumes of data. They also act as a catalyst in the development of social programs which develop complex and fine-tuned adjustments in terms of defined categories of participants, differential eligibility requirements, and other such variables. They result in rich program resources for decision making, whose very existence is a stimulant to the statistician to probe its availability and its adaptability for statistical uses. The selection and molding of administrative records to the needs of the particular, often very narrow, administrative use creates built-in problems of definition and comparability for the statistician who may wish to influence the design of administrative data collections.

There are other factors as well, reflecting the growing resistance of respondents, both persons and firms, to cooperate with voluntary data

collections. One is simply burden of collection activities which have proliferated in number, frequency and detail, without any obvious compensating personal benefit to participants.

Another factor is public distrust of information gatherers, both governmental and private and decline in confidence in the ability of survey organizations to preserve the confidentiality of information entrusted to them.[1], [2]

These factors combine to raise concern about the acceptable level of response burden, counting both voluntary and involuntary collection, which can reasonably be imposed on the reporting public. Where administrative and statistical requirements for information compete, the program requirements generally take precedence.

The extraction of data from information collected by agencies in administering their social and economic programs offers an attractive alternative to mounting new surveys. Compilation of administrative data to produce a microdata file can produce a synthesized equivalent of the response to a personal interview. Even where a "survey" is simulated by linkage of individual data for a sample of persons or firms from records maintained in several different programs or agencies, the cost may still be relatively small compared with the cost of conducting an actual survey. In some instances, cost is a secondary factor, where personal contact would be difficult or even impossible because of inability to interview the necessary sample population, for example, deceased persons.

2. A language framework for legal issues

Generally administrative records mean records which contain information used in making decisions or determinations, or for taking actions affecting individual subjects of the records. Commonly the term refers to records about natural persons, although other entities may be treated by law as legal "persons," about whom decisions and actions are taken, e.g. corporations, partnerships and sole proprietorships. Under IRS rules, similarly, the estate of a deceased person is a "taxpayer" and its records are subject to disclosure rules just as if the taxpayer were a living natural person. In other contexts, legal rules on disclosure might vary depending on whether the particular information refers to an individual in his capacity as a private person, for instance, or as a business proprietor. The juncture of Freedom of Information Act[3] and Privacy Act disclosure rules with respect to a particular set of data may raise just such an issue.

This paper deals with one segment of the large volume of administrative records kept by public and private record keepers. The focus is on records kept by government agencies, mainly Federal, compiled principally in managing their social and economic programs. While agency personnel, law enforcement, regulatory, and other such

such records are also administrative records in the broad sense, they lie outside the scope of this discussion. As a matter of convention, likewise, decennial census records are excluded from consideration as administrative records, although arguments can be made for their inclusion in light of their use for redistricting calculating revenue sharing, providing genealogical data, and so on.

Statistical purposes describe purposes for which information about individual members of a defined study population is aggregated and presented without reference to individual identities. Statistical records may be kept, used, and published in microdata form to maximize flexibility for examining and analyzing the composition, characteristics, behavior, etc., of the group under consideration. Personal identifiers may be kept on microdata records for purposes of record validation and linkage, and the files may be transferred to statistical users with identifiers. Of course, the individual identities of the persons making up the statistical group are not associated with the statistical files once processing is completed, nor are they material to the ultimate statistical results of the process.

Access, use, and disclosure. There are some subtle distinctions in the ideas of access, use and disclosure of records. "Access" to (or availability to see, hear, examine, or otherwise be cognizant of the information contained in the records. "Use" generally refers to the purposes which can be served, or the operations which can be performed with the records by the person who has access. A basic distinction between statistical and non-statistical use is of principle concern. In this connection, the application of statistical methodology is not equated with statistical use. An identifiable person may be singled out for any number of administrative actions--such as promotion, tax audit, and so on--on the basis of a statistical operation, such as ranking by specified characteristics. This would be an administrative use of statistical techniques, and not a statistical use. Quality assurance programs often involve hybrid uses of this sort, and are considered to make administrative rather than statistical use of the data.

Finally, "disclosure" involves providing access or availability to another user, usually by transfer of records, although disclosure can, of course, take place also by word of mouth.

Confidentiality and privacy. Confidentiality refers to limitations which protect records from unauthorized access, use or disclosure. Privacy refers to the protected right of the individual not to be disturbed, or not to have intrusive invasions of his person or property.

3. Concept of functional separation

Statistical use of administrative records has been scrutinized both from the confidentiality side by such agencies and commissions as HEW and the Privacy Protection Study Commission (PPSC)[4], and from the burden side by others such as the Office of Management and Budget (OMB)[5], the

General Accounting Office (GAO), and the Commission on Federal Paperwork.[6] The President's Statistical Reorganization Project also has recently looked at both confidentiality and burden. From these inquiries has emerged a consensual view that the public will benefit from better access by statisticians to administrative sources of information. A caveat is generally added, that better protection of statistical compilation of administrative data to prevent unauthorized use for non-statistical purposes.

These studies have focused considerable attention on an important concept of "functional separation" which was developed in the work of the Privacy Protection Study Commission, and recommended for statutory treatment by the PPSC, the Administrative Privacy Initiative, and the President's Statistical Reorganization Project. Functional separation is based on a promise that there are qualitative differences between program-administrative functions and statistical-research functions which require differential standards for managing the information needed by each.

These standards relate to access, use and disclosure of data. Functional separation means that a separate and distinct approach is necessary for the development of principles, legal rules and practices applicable to data for statistical use. While the principles and standards applicable to statistical use have to take into account the principles and standards which apply to administrative use of information, and in some respects are constrained by administrative rules, the rules for statistical data need not be similar or even parallel to those for administrative use.

Applying the principle of functional separation, to make the rules appropriate for the function that the information serves, data cannot be mixed indiscriminately in statistical and administrative uses. Information designated for statistical use would not be available to administrators for their use except in anonymous or aggregate form, regardless of whether the data were obtained directly through surveys or indirectly from administrative files. With that constraint, records compiled in administering particular programs can be used by statisticians without risk of breaching the right and expectations of program participants about the intended uses of information they give. Stated simply, information with individual identifiers could flow only in one direction, from decision and action records into statistical data sets. Then identification has to be suppressed before the flow can be reversed. This aspect of functional separation has provoked considerable debate with compliance and enforcement officials, and is at the cutting edge of legislative proposals to provide legal protection to statistical files.

4. Options: Legislative approaches to functional separation

There are two principal approaches by which functional separation can lead to protected status for data committed to statistical uses. Both approaches can be found in some recent legislative proposals.

The first approach is to protect designated statistical activities. The method is to name certain organizational units as being qualified users of statistical data, to require safeguards for all statistical data which they manage, and to impose limitations on access and disclosure. This is the design for the "protected statistical center" which is described in the proposed Confidentiality of Federal Statistical Records Act.[7] The model for this approach is the Census statute (Title 13 of the United States Code) which limits examination and use of Census records to employees of the Census Bureau. The difference which would be introduced by this proposed extension of the Census concept is that use would not be limited exclusively to employees of the organization which does the actual collection of the data. Instead, under the proposal, data could be transferred among approved centers with relative ease. Since no data could be disclosed except among protected centers in a way which would permit such data to be associated with identifiable persons or business reporters, the agency which collected the information could even be ordered to transfer its data to other centers which demonstrated their need.

The second approach is to protect specified records or files, regardless of where they are physically located. The method is to designate particular data elements or collections of data elements as "statistical" (or research) records, and to place special conditions on the purposes for which the files can be used. In addition, this approach would restrict disclosures, both as to the form of records disclosed, and as to the type of authorized recipients. This is the approach developed in the proposed Research Records Act.[8]

The Research Records Act would apply to research records as defined. With respect to information about natural persons, this definition is somewhat broader than merely statistical records, except as records are excluded by coverage in such statutes as the proposed Statistical Records Act, Census Act, etc. On the other hand, the research proposal is narrower in scope than the proposed Statistical Records Act, since it would not apply to information about firms or other entities which are not natural persons. The proposed Research Records Act incorporates most of the recommendations of the Privacy Protection Study Commission to provide separate and distinct treatment and disclosure rules--functional separation--for the statistical and research records which it would cover.

This latter approach of protecting data, rather than managing data holders, is also used in the proposed Statistical Records Act referred to above, with respect to files which have been designated by a Chief Statistician as "protected statistical files." These latter conditions would be somewhat less stringent than the conditions attaching to files in protected centers, and would permit the use of protected files as sampling frames for disclosure of names and addresses of entities to contact in order to obtain additional information through surveys or interviews for statistical and research purposes.

Both approaches have been considered in developing the "Standard Statistical Establishment List," and the legislative proposals for widening its availability. At present, the SSEL is maintained by the Census Bureau, and is used in identifiable form only by Census personnel to prepare tabulations which are made available to others in a form not permitting identification of particular firms or establishments. Some proposals for broadening access have recommended the first approach, which would be to name the statistical units which are qualified to use the SSEL information both for preparing tabulations and for drawing samples of enterprises for surveys and questionnaires, and to exclude other statistical users. Other proposals have taken the second approach, available to responsible statistical users, strictly limited to statistical and research applications. In addition, a third type of proposal has offered a "two-tier" compromise. This would create one class of information available to users for statistical and research applications. A second class of information would be available only to Federal statistical agencies for their statistical use.

DYNAMICS OF FUNCTIONAL SEPARATION

5. Dimensions and characteristics of the legal framework

Components operating within the governmental agencies which administer programs use the agency's administrative files as the source of statistical inquiry. The propriety of such use has seldom been questioned, at least within the Federal establishment. Most agency staff, indeed, would not ordinarily consider the availability of program records to in-house statisticians as disclosure at all, although in a legal sense it may be. However, the laws have usually been silent about the conditions of such use.

In the obverse situation, questions have arisen as to the proper extent of access which administrators and compliance officers can or should have to information produced by statisticians from their use of those same administrative records which they sample and use statistically. Auditors claim broad access powers, and recognize few limitations on the uses which they may make of information, regardless of its statistical or non-statistical source.

Disclosure within the agency, a broader view. Authority for use of an agency's records by the agency's own employees for various agency purposes is implicitly assumed on a need-to-know basis, as observed above. Frequently there is no express authorization for such intra-agency disclosures, although the converse, restrictions on use or transfer, even within the agency, may be imposed by law. The Privacy Act, in contrast, provides explicitly for disclosures to the agency's own employees. While the principles of functional separation between statistical and other files are often carried out in administrative practice with respect to intra-agency use, they are less often subject to statutory treatment than are transfers for inter-agency use.

Disclosure to agency contractors. The relationship between the program and the statistician who actually performs the work may become attenuated, and the issues then become more complex. For instance, an agency may wish to use information in its program records to study particular aspects of a client population. It may find that it lacks sufficient or appropriate staff resources to commit to the necessary tasks of preparation and analysis. In such a case, the agency may enter into a contractual arrangement to have the work performed to its specifications by outside organizations. While the work product may be the same as that which would result if the agency relied on its own staff resources, the legal issues and relationships are different when the work is performed by outsiders. The agency must then deal with legal questions related to the disclosure of confidential information to others. A variety of statutory considerations may be involved. Conditions are different for data controlled by the Privacy Act, for example, than for data controlled by the Census statute or the Internal Revenue Code.[9] The Census statute permits no one but Census employees to examine Census returns. On the other hand, the Privacy Act allows disclosure of covered records for a "routine use," and many agencies have determined that disclosure of information needed by contractors to perform their contractual duties would qualify as a routine use of personal information protected by the Privacy Act. In contrast, the Internal Revenue Code (as amended by the Tax Reform Act of 1976) has a provision requiring a particular type of agreement with a contractor to perform data processing functions with tax return information for purposes of tax administration. This provision applies to tax information about business and other taxpaying entities, as well as information about individual taxpayers.[10]

In determining what information can be released to an agency's contractors, and in providing for the disposition of files upon completion of work which agencies contract for, several statutes may impinge on the relationship and complicate the conditions and scope of work. Also, the agency may need to make provision for availability of the working files, with appropriate protections for secondary analysis by the contractor or by others on their own behalf. The nature of the provisions for purging of identifiers, destruction of records, and so on will be influenced by the statutory authority under which the contractual work is done, whether or not the contractor is "maintaining a system of records" as defined by the Privacy Act.

Disclosures among Federal agencies. Alternatively, an agency may serve a population whose members are also covered in whole or in part by a program or activity which is administered by another entirely separate agency. In such a case the two agencies may benefit from creating an enriched file which merges information about a sample of individuals extracted from the separate records of each agency. For instance, a group of Social Security beneficiaries might also be recipients of benefits administered by the Veterans Administration.

Statistical matching techniques[11] may be used, of course, without any individual identification or disclosure. Thus, records of individuals can be selected from each agency's files on the basis of a set of specified characteristics, (e.g. age, sex, race, marital status, etc.), and can be compiled without identifiers. The separate files without identifiers can be merged solely on the basis of similar characteristics, thus synthesizing individual records without any effort to ascertain whether records of the same individuals were in fact merged.

It is more common, however, to create a merged file on the basis of identifiers known to both agencies. When information about a sample of individuals known to the agency is used to create such a merged file, the procedure ordinarily involves some disclosure from one agency to another of both identifiers and of individual data. The process may involve a range of disclosure possibilities, for instance, a one-way flow of identifiable data from a source agency to the agency performing the match, with a return flow of files containing merged records purged of identifiers. There could be a two-way flow of identifiable records between the participating agencies, or there could be a one-way flow from each of the participating agencies to a third agency to perform the match and purge identifiers.

Legal implications of the various possibilities will depend on the legal character of the source information, the cooperative agreements between and among the agencies, and the nature of the resultant files in terms of the potential for matching back against the program or statistical files of the participating agencies.

Use by non-statisticians of statistical files compiled from administrative source records. Statistical analysis selects a small population segment to serve as proxy for a larger target population, focusing on salient characteristics, behavior, relationships, etc. The statistical files and their analysis may, by their design or purpose, provide important information to program administrators, oversight agencies, legislators, auditors, and courts. When these users are satisfied with statistical results based on anonymous or aggregated data, the purposes of the statistician and the non-statistician are compatible, and the statistician can conscientiously make the files available even though the ultimate uses are foreign to his own interests.

Often, however, the administrator, auditor, or regulatory or enforcement officer wishes not only to use statistical results to identify population segments in which he is interested, but wishes also to locate and take action affecting individuals in the group thus identified. (The epidemiological researcher may have a similar design, though for what may be regarded as more benign purposes.) Here the objectives of the statistician are thwarted. Such uses raise doubts about the objectivity of the statistician, the premise of confidentiality on which he bases individual data collection, and the essential fairness of permitting the statistician to have free access to

otherwise confidential information provided by persons for purposes associated with their participation in particular programs.

Moreover, the statistician's sample is usually selected on attributes not associated with the action purposes of the non-statistical user, and the sample data may selectively preserve data about individuals in the sample population which are no longer retained in the underlying program files. The marriage of information from the separate files may also generate new information which was not itself contained in either of the source files, for example a level of income reported by the records in one file which is legally inconsistent with eligibility for benefits whose payment is reported by records in the other file.

Such possibilities raise ethical issues which are beyond the scope of this paper. The statistician takes the general position, however, that the administrator or enforcer ought to have access to aggregate results only, and not to individual data which has been matched for statistical purposes.

6. Some Federal statutes affecting statistical use of administrative records and protection of statistical records from non-statistical use

In general, Federal statutes which have provided confidential treatment of record information have, by providing essentially equivalent treatment to administrative and statistical records, had a dampening effect on productive statistical efforts. For the most part, the laws have discouraged harmless interagency disclosures of identifiable data for statistical purposes at least to the same degree that they have impeded administrative disclosures, and probably more than they impeded enforcement transfers. They have neither assisted the statistician in gaining access to records, nor protected the record subjects from administrative actions based on statistical records.

An exception is the Census statute, Title 13 of the U.S. Code, which gives the Census Bureau broad authority to obtain information, including data contained in agencies' administrative records, at the same time it protects the Census records from being disclosed either voluntarily or by compulsion in a form which makes individual identification possible.

The Federal Reports Act[12] is a record management statute which applies to solicitation of information by Federal agencies from ten or more respondents. Because of its restrictive provisions on interagency transfer, it is not an effective mechanism for authorizing transfers of identifiable data for statistical purposes.

Some recent statutes which have been enacted to protect privacy and confidentiality of information collected by the Federal government have dealt with statistical information in ways that still frustrate legitimate statistical needs. Statistical information can be disclosed only in

a form which does not permit individual identification. Under this provision by itself, no administrative file linkage in identifiable form would be possible for statistical purposes except within the agency which collected all the information in the files to be matched. The Privacy Act basis on which agencies have disclosed data for statistical use is the provision that allows disclosure for a "routine use" which is "compatible" with the purpose for which the agency originally collected the information. Under this provision, some administrative file linkage is performed by agencies which have joint statistical interest in the merged records, and which demonstrate compatibility of agency purposes in order to warrant the necessary disclosure. The Social Security Administration and Treasury, for instance, have created some match files which merge demographic, earnings, and income tax information for a sample of individuals whose records are contained in both agencies' administrative files. Once these files are created, they are purged of explicit identifiers, and are used in anonymous form for analysis by both agencies and by Congressional oversight committee staff. With additional suppression of information (such as geography or extremely high income level) which might lead inferentially to identification of some individuals, a public use microdata version can be produced.[13]

The Tax Reform Act of 1976 has placed stringent restrictions on the disclosure and use of information collected by IRS from and about taxpayers, both individual and business or institutional. Information about earnings and withholding subject to the Social Security Act, including self-employment earnings, is defined by the Tax Reform Act to be within its scope. As such, it is governed by the confidentiality provisions of the Internal Revenue Code, which provides expressly but not generously for statistical applications, and which does not allow discretion as to disclosure to statistical agencies not named in the statute.

The Freedom of Information Act (FOIA) makes a somewhat jagged cut across various disclosure provisions. Information about natural persons which is covered by the Privacy Act, for instance, must be disclosed under FOIA unless its disclosure would be a "clearly unwarranted invasion of personal privacy", or unless it is protected by another statute, such as Census Title 13. FOIA requires that information about business firms must be disclosed unless its disclosure would breach trade secrets or reveal confidential financial information, unless the disclosure is prohibited by another statute, such as the Internal Revenue Code. Other statutes interact with FOIA in similar patterns of inconsistency, insofar as the substantive content of the files is concerned.

In addition to statutes, government agency regulations or guidelines may complicate statistical applications based on administrative records. The Office of Management and Budget recently published guidelines under its Privacy Act authority, applicable to Federal agencies' record matching activities for purposes of fraud detection.[14] These guidelines also apply (somewhat less stringently) to matches for purposes other than antifraud enforcement. Although the guidelines do not prohibit

file linkage, they do require reporting to Congress and OMB in advance of any matching activities. There is an exception for matching of files within an agency, for statistical purposes, but it is by no means clear whether agencies must give prior notice of planned interagency matches derived from administrative files for statistical analysis. Similarly unclear is the status of user files which are provided to agencies to identify sets of individuals for whom record information is to be extracted and matched to augment user information in a file which the agency is asked to create in order to prepare specified statistical tabulations.

7. Summary and directions for the future

The discussion makes clear that the legal issues associated with expanding statistical use of administrative records are fairly complex, often changing, and sometimes inconsistent in their results. Some insights are possible when the legal issues are examined as questions of access, use and disclosure of records. From that starting point the emerging principles can be related to privacy and confidentiality as key concepts underlying those principles, and as embodied in legislative efforts to achieve functional separation.

The current Administration's Privacy Initiative and the President's Statistical Reorganization Project have made recommendations leading to legislative proposals for functional separation which would create quite different mechanisms for the protection of statistical records than for protection of research records, as the terms are defined in the respective proposals. Nevertheless, the line of demarcation between statistical and research records and uses is an arbitrary one, and the two bills would interact to produce a complicated matrix of criteria.

These legislative proposals are complex and need careful thought for the full implications to collectors and users of statistical information in specific applications. In general, however, their dual thrust is to establish conditions permitting freer availability of information among agencies for statistical use, including agency access to Census records, and then to protect files from being used for individual actions and decisions, once the information in them has been compiled and designated by statisticians for statistical use. These broad goals are of great importance to the work of statisticians both inside and outside the government agencies which maintain administrative files.

NOTES AND REFERENCES

- [1] See, for example, Louis Harris and Associates, Inc., and Dr. Alan F. Westin, "The Dimensions of Privacy: A National Opinion Research Survey of Attitudes Toward Privacy," conducted for the Sentry Insurance Co., 1979.
- [2] Privacy Act of 1974; 5 U.S.C. §552a(e).
- [3] Freedom of Information Act, 5 U.S.C. §552.

- [4] Personal Privacy in an Information Society, Report of the Privacy Protection Study Commission, July 1977. Chapter 15, "The Relationship Between Citizen and Government: The Citizen as Participant in Research and Statistical Studies."
- [5] Office of Management and Budget (OMB) Circular A-46.
- [6] Report of the Commission on Federal Paperwork, 1978.
- [7] The Administration's proposed bill "Confidentiality of Federal Statistical Records", is based on the recommendations of the President's Statistical Reorganization Project, and was circulated by OMB for agency review in mid-1979.
- [8] Privacy of Research Records Act of 1979, and administration bill.
- [9] The Tax Reform Act of 1976, P.L. 94-455, 94th Congress, October 4, 1976.
- [10] 26 U.S.C. §6103(n).
- [11] Radner, D., "Report on Exact and Statistical Matching Techniques," to be published in 1980 as an Office of Statistical Policy and Standards (OFSPS) Working Paper.
- [12] Federal Reports Act, 44 U.S.C. §3501-3511.
- [13] Caldwell, S. and Diamond, T., "Income Differentials in Mortality: Some preliminary Results Based on IRS-SSA Linked Data," to appear in the 1980 Proceedings of the American Statistical Association, Section on Survey Research Methods.
- [14] Office of Management and Budget (OMB) "Guidelines for the Conduct of Matching Programs," Federal Register, March 30, 1979.