BEYOND THE PUBLIC USE FILE: CONFIDENTIALITY OF ARCHIVAL RECORDS

A CASE STUDY, THE NATIONAL CENTER FOR HEALTH STATISTICS

BY

Cynthia G. Fox
National Archives and Records Service

Governments and society need information to plan, execute and evaluate in a rational manner. Access to timely and accurate information is the cornerstone of liberty. It is no accident that Orwell's hero in 1984 was an information manager or that Henry Ford's motto "History is Bunk," and therefore subject to change was the watchword of Huxley's Brave New World. Free societies have a need and a right to know. In the United States, for example, the Social Security Act exemplifies the notion that information once considered completely private, that is work record and salary, should be collected about individuals to insure rational execution of social programs. Citizens of free societies fund these efforts through taxation and participate in the collection of data by responding to census questionnaires, filing income tax returns, registering births, deaths and marriages, and applying for licenses. In most cases, they do so willingly and often without threat of criminal liability.

This fundamental support for the collection of data, however, is tempered by fear, both real and perceived, of a loss of individual privacy and the creation of the all knowing "Big Brother." At the U.S. Federal level, this fear is translated into legislation and regulation. The Privacy Act of 1974 requires a government to inform the citizens of the existence of systems of information which may contain data about them. It permits citizens to request the destruction of files which the government has created or collected on them or correct misinformation contained in those files. Another piece of legislation, the Freedom of Information Act, permits individuals to obtain access to much of the information the government collects. These two laws attempt to insure that the individual's rights to privacy and access to information are protected. The two laws work in concert. The Freedom of Information Act, which protects the right of access, exempts from disclosure personal information the release of which would clearly constitute an invasion of personal privacy and specifies medical data. The result is that in the United States, information managers at the Federal level are at the center of a triangle composed of 1) the need to collect data of a personal nature in order to assure rational planning; 2) the right of the citizens to access the information collected by the Federal government; and 3) the need to insure the privacy of the individuals about whom the data is collected. If the confidentiality of the individual cannot be protected, then the ability of the government to collect the accurate information it requires will be impaired. Similarly, the right of access to information cannot infringe on the right of the individual to maintain his personal privacy.

The dilemma seems less of a problem when discussing machine-readable records. One collects that data needed to insure rational decision making, drops off the personal identifiers, edits or aggregates the data, and releases that new version, "a public use tape" to researchers. Unfortunately, edited and suppressed records are not always the answer and "disclosure free" data tapes are not always totally

disclosure free. I will attempt to describe the efforts of one Federal agency to prevent the unwarranted invasion of personal privacy and the efforts of the National Archives to continue to protect the confidentiality of this type of record when transferred.

The National Center for Health Statistics, an arm of the U.S. Public Health Service, states that its primary mission is "to develop statistical information on health matters in the United States and to provide that information as quickly and in as useful a form as possible to all who desire it."(1) This mission is conducted within the context of strict controls and guidelines aimed at insuring the confidentiality of individuals and entitites about whom they collect the information.

The National Center for Health Statistics or NCHS functions under two basic propositions which are that the transfer of personally identifiable data from one custodian to another should occur only in accordance with carefully formulated and widely understood written rules and that there is a basic difference between information collected and used only as statistical evidence and personally identifiable data used directly to affect the rights, benefits, privileges, responsibilities, duties, or proscriptions of individuals. They collect data for statistical and reporting purposes only and they do so in a fashion controlled by regulations, written agreements and signed assurances.

There are two laws which permit the Center to provide the confidentiality it requires to carry on its work. Section 308(d) of the Public Health Service Act (42 U.S.C. 242m) provides the basic legal authority for the protection of NCHS files. It states that no information obtained in the course of activities undertaken by NCHS and its sister agency, the National Center for Health Service Research may be used for any purpose other than the purpose for which it was supplied unless authorized by the Secretary of Health and Human Services. The law states that information obtained in the course of health statistical activities may not be published or released in another form if the particular establishment or person supplying the information or described in it is identifiable unless this establishment or person has consented. Whenever NCHS requests information, it specifies to the person or agency supplying the information that the data will be used for a limited purpose or purposes. In most cases, this use is limited to statistical research and reporting.

Under a second law, the Privacy Act of 1974, NCHS has obtained a "K-4" exemption for its statistical systems. This means that NCHS does not have to allow the subjects of its data files to have access to the records about themselves in those files. This exception to Privacy Act requirements is permitted because NCHS does not have in its data files any records that are used in any direct way to affect the persons whose records exist in these files. (2) The files are used strictly for statistical and related purposes.

The Public Health Service Act and the Privacy Act augment the basis for exemption from the Freedom of Information Act (4 U.S.C. 552). Subsection (6) of the act specifically exempts personnel and medical files and similar files "the disclosure of which would constitute a clearly unwarranted invasion of personal privacy" and subsection (3) provides that matter "specifically exempted from disclosure by statute" are also excluded from the disclosure requirements. The Public Health Service Act provides the necessary statutory restriction to prevent the disclosure of individual information.

The Center makes every effort to assure that no breach in confidentiality occurs at any stage of the life of their files. In all cases when NCHS contracts

with any organization outside the Center for the collection or use of information that identifies individuals and/or establishments not advised that all information obtained from them will be made public, the contracts are carefully worded to assure compliance with either the Privacy Act of 1974 or the Public Health Service Act. The contracts contain stipulations to assure confidentiality and physical security of the information and ensure that the contractors' employees abide by the Center's stipulations. (3)

NCHS uses three standard wordings in data collection contracts, depending on the laws governing the particular project from which the information is to be collected. The first alternative is to be used when both the Privacy Act of 1974 and Section 308(d) of the Public Health Service Act apply. The project could involve the collection of information about identified individuals which NCHS is authorized to carry out. The second alternative is used when Section 308(d) of the Public Health Service Act applies to the project but the Privacy Act does not. For example, in the case of a survey of institutions providing health services, the Privacy Act covers only individuals and does not apply. The third alternative would be used when the Privacy Act but not the Public Health Service Act covers the study. Such a situation would be rare but might occur if Congress authorized a study not normally conducted by NCHS and required that NCHS conduct the study. The fourth wording is used in contracts called for when contractors process confidential data.(4)

NCHS contracts with organizations for the collection, processing, and analysis of data because such organizations are specifically equipped and staffed to perform such services effectively, although the contracting organization has no intrinsic rights in the data. These contractors provide such services as extensions of NCHS and, as such, they are subject to all the confidentiality strictures which apply to NCHS itself, and the contractors' employees are subject to the same Privacy Act actions as employees of NCHS. The Center's policy on protection of records also applies. (5)

All contractor employees must sign a "Nondisclosure Statement" as do all NCHS employees. In addition to signing the statement which specifies the force of law and punishment for violation, employees agree to maintain the same physical protections that the records are afforded in the Center. This includes keeping them locked up in fireproof cabinets or locked rooms at all times when they are not actually being used, keeping them out of sight of persons not authorized to work with the records, limiting duplicates, and transferring them in sealed containers. Furthermore, in all statistical programs involving confidential information, records containing identifiers or individuals or establishments should be held to the minimum number required to perform the Center's function. Identifiers are never carried beyond the original survey or report document when the data are processed, unless there is a legitimate and important reason for doing so. Documents containing identifiers are stored in secure areas as soon as possible.

NCHS releases its statistical data in the form of traditional published tables. It also creates public use versions of its data files by deleting personal identifiers or suppressing selected elements to ensure that the identity of the individual is protected. The availability of machine-readable microdata files enhances the value of research conducted by NCHS. It permits other public and private institutions to use the data for purposes other than the rational planning of health related services. The Center has determined that the release of the data in machine-readable form is a desirable objective. (6) However, NCHS must achieve this goal without compromising the confidentiality of its files.

The question of how to achieve both goals raises a wide range of ethical, legal, technical, technological and economic issues. Five different classes of constraint must be considered in deciding if microdata can be released and if so how it should be handled.

I have previously mentioned the Public Health Service Act which provides the legal constraint on release. As I mentioned, the PHS Act makes it clear that data collected by the Center must be processed in such a manner that the identity of no individual is disclosed and that the individual identification will be used only by persons engaged in achieving the purpose for which the information was originally assembled. The identifiers are deleted as soon as possible in the processing sequence. Policy and practices in NCHS and other general purpose statistical agencies have given strict interpretation to this principle, not only protecting individual records against unauthorized use, but also in adopting tabulation and publication procedures that are designed to make it virtually impossible to isolate, identify, or extract facts in such a way that a specific individual or business establishment can be identified by use of released data. (7)

In addition to these legal constraints, there is an ethical concern governing the release of microdata files. On basic human grounds, NCHS has an obligation to protect individual respondents against any invasion of their personal privacy and to be absolutely sure that the confidentiality of privileged communications is not breached. In the United States this ethical responsibility is embodied in the previously discussed exemption of medical files from disclosure under the Freedom of Information Act. However, NCHS is a Federal Government Agency and as such must make every effort to assure that the citizens of the country have maximum access to bodies of information that the government assembles. While NCHS has received a K-4 exemption from Privacy Act disclosure of individual files, the ethical concern embodied in the Privacy and Freedom of Information Acts still exists. A government in a free society is and ought to be responsible for informing its citizens about the scope and content of systems of information that it maintains. Any release of microdata must be done under public scrutiny providing equal access and equal protection for all involved.

The third consideration in the possible release of microdata is a technical constraint. Because NCHS programs are generally designed to produce estimates for the entire United States, some studies require elaborate scientific sampling procedures. Release of the microdata without a complete explanation of the editing, weighting, and ratio adjustments that it has undergone would be close to releasing inaccurate data. The Center must determine if the sampling and manipulating can be explained and duplicated by a researcher.

Technological considerations provide the fourth constraint on the release of data. Hardware dependency is an example of this type of consideration. Another example would be data in other than a statistical form. In some NCHS programs, original data includes such records as xray films, electrocardiograms, paper tapes, tape recordings of speech samples, blood specimens, and photographs. The ability to provide reproductions of these types of information are bound not only by technological difficulties by financial constraints as well.

This is the fifth consideration in the release of data, the economic problems. Financial means have become very important in determining the nature of the Center's statistical output. The Center's resources in funds, personnel, and equipment are limited. The highest priority and prime purpose for which NCHS surveys are conducted are the prompt production of general purpose statistical tabulations for use by a broad spectrum of consumers. The preparation of public use files takes second place to the production of these tabulations. (8)

These five constraints are reflected in the NCHS Policy Statement on release of data for individual elementary units and special tabulations. It reads, "Within prevailing ethical, legal, technical, technological and economic restrictions, it is the policy of the National Center for Health Statistics to augment its programs of collection, analysis, and publication of statistical information with procedures for making available, at cost, transcripts of data for individual elementary units--persons or establishments--in a form that will not in any way compromise the confidentiality guaranteed the respondent."

Micro-data tapes are released after they have been reviewed and approved by the Director of the Center as conforming to guidelines and conditions set forth in the Policy Statement. Operational considerations and attention to the matter of unit costs mean that for most data the effective format for release for the individual elementary units is a Standardized Microdata Tape Transcript. A descriptive catalogue of public use tapes is published. The catalogue is supplemented by a newsletter of updates.

Each public use data set is governed by a procedure designed specifically for that particular study. In general, however, the data set is composed of a standardized transcript, a computer tape image of the edited, weighted, and adjusted data from which all evidence is deleted which might possibly identify the respondent. The image is arranged in a fixed format and is accompanied by documentation which explains the editing and weighting and permits the use of the tape. Any codes which appear on the tape are scrambled or offered in the most general terms. Data which has proved faulty is represented by blanks. The Center will not modify standardized transcripts or produce special tapes when standardized transcripts have been prepared. Finally, each purchaser must sign, as part of the order form, a statement of assurance regarding the use of the data: "The undersigned gives assurance to NCHS that individual elementary unit data on the micro-data tapes being ordered will be used solely for statistical research or reporting purposes."

The Center's reference technical services are handled under contract with the National Technical Information Service (NTIS). NTIS maintains and sells out of print NCHS publications in addition to reference copies of the public use files. However, proprietary responsibility for the data is still in the hands of NCHS. When the agency no longer requires the information to perform its function or when data is replaced or superceded the records become eligible for transfer to the National Archives.

The National Archives and Records Service has been concerned with the protection of confidentiality since its establishment in 1934. Long before "privacy" became a national issue, NARS successfully protected personal, restricted, and classified material from unauthorized disclosure. NARS' general restrictions placed a 75-year restriction on files which contained personal and medical information the release of which could be embarrassing to an individual. This principle has been enforced consistently unless prospective researchers assure NARS that data is to be used for statistical or reporting purposes only.

According to the U.S. Code, records transferred to the National Archives become the responsibility of the Administrator of General Services (the parent agency of the National Archives). Statutory restrictions on their use continue to apply for 30 years from creation or longer if the Archivist and the head of the transferring Federal agency so advise (9), and other restrictions may be negotiated between the transferring agency and the Archivist.

Micro-data files which are deposited into the National Archives are treated in precisely the same fashion. Since NCHS records are subject to the restriction imposed under the PHS Act previously discussed, they are to be used for statistical and reporting purposes only, for a period of not less than 30 years. In addition, NARS General Restrictions, revised February 14, 1983, restrict for 75 years "Records containing information about a living individual which reveal details of a highly personal nature...including but not limited to information about the physical or mental health or medical or psychiatric care or treatment of the individual...not known to have been previously made public." (10) Such records, the restrictions go on, may be disclosed to "researchers for the purpose of statistical or quantitative research when such researchers have provided the National Archives with adequate written assurance that the record will be used solely as a statistical research or reporting record and that no individually identifiable information will be disclosed." (11)

In addition to the guidelines and restrictions which pertain to all archival records, two Office of Management and Budget publications offer guidelines for determining the confidentiality of machine-readable records. OMB's Computer Security Guidelines for Implementing the Privacy Act of 1974 (FIPS #1), prepared by the National Bureau of Standard, requires that provisions exist for stripping "records of individual identifiers so that identities cannot be discerned when statistical research or reporting records are disclosed or transferred I(A) (10) and for ensuring "that an individual's identity cannot be discerned from tabulations or other presentations of statistical data by combining various statistical records or referring to other available information" I (A) (11).

By requiring that the agency transferring records to the National Archives supply a complete and accurate record layout as part of the documentation in hard copy form, the Machine Readable Archives Branch is able to determine with relative ease if a file contains information of a personal nature. These precise definitions of potentially restricted data elements along with detailed information relating to the specific assurances given to the respondents make the production of public use tapes in extract or summary form possible.

OMB's Privacy Act Implementation Guidelines and Responsibilities is the second publication which may be used to determine confidentiality of data elements. According to the Guidelines the elements of personal information may constitute an invasion of personal privacy if retrieved through personal identifiers. This covers a number of areas including: income or census information, private or subjective information about an individual or family, inaccuracies, information about individuals used by a Federal agency for making a policy decision, information gathered under an implied or expressed promise of confidence, or any combination of public elements which combines to form too intimate or too detailed a profile.

The National Archives operates within these guidelines and under its own restrictions and staff guidance and procedures to identify data elements which are subject to protection from disclosure. In addition, the following steps are taken to ensure the physical and intellectual security of restricted files:

a. All files are stored in an environmentally controlled vault;
b. Access to the vault is limited to NARS reference and technical personnel and the Washington National Records Center Security Officer;
c. The file and its documentation are stored separately;
d. No file is maintained in an office area except when in transit;
e. Documentation does not accompany files for computer processing;
f. All programming is controlled;

f. Agency restrictions are strictly enforced;
h. All restrictions are published; and
i. All requests for access to restricted data, even for statistical and reporting purposes, must be approved.

In the case of NCHS data files, the National Archives will accession not only the full master files of unsuppressed microdata but the public use versions as well. The master files meet all of the criteria for restriction for at least 75 years unless an inter-agency agreement reduces that to 72 as has been done for census files. The public use versions created by the agency will be accessioned rather than created by NARS because the Center's record of protecting the privacy and confidentiality of their data has been outstanding. The National Archives and Records Service wants to preserve the records of the Federal Government and provide access to any researcher as completely and conveniently as possible without violating the confidentiality of the information or the privacy of the source of that information. NARS and NCHS share these common goals and are working together to insure that a wealth of health related data is preserved.

NOTES

(1) U.S. Department of Health, Education, and Welfare, Public Health Service, Staff Manual on Confidentiality NCHS, DHEW Publication No. 1 (PHS) 78-1244, (Hyattsville, MD: July, 1978) 10.

(2) Ibid., 3.

(3) Ibid., 13.

(4) Ibid., 26-37.

(5) Ibid., 9.

(6) National Center for Health Statistics, Policy Statement on Release of Data for Individual Elementary Units and Special Tabulations, (Washington, DC: 1978) 3.

(7) Ibid., 5.

(8) Ibid., 6-7.

(9) 44 U.S.C. 2104.

(10) 41 C.F.R. part 105-61.5302.4(a)

(11) Ibid., 3(a).

(12) U.S. Department of Health, Education, and Welfare, Public Health Service, The Model State Health Statistics Act: A Model State Law for the Collection, Sharing, and Confidentiality of Health Statistics, (Hyattsville, MD: March, 1980) 7.