General Principles in Establishing Effective Confidentiality Practices

Alvan O. Zarate
Confidentiality Officer
National Center for Health Statistics

Prepared for Session 3 "Statistical Data Stewardship in the 21st Century" of the annual Joint Statistical Meetings, August 11, 2002
New York City

Introduction In the last several years, a variety of methods for assessing the adequacy of privacy practice have appeared. These range from "Privacy Impact" Statements to Privacy Audit or Privacy checklists. They serve a very useful purpose in alerting us to key issues that we may have overlooked or neglected as well as providing newcomers an overall view of acceptable privacy practices.

Today, I'd like to present my own thoughts on Privacy fundamentals that I first developed about four years ago. I had been reviewing testimony before a regulatory agency that presented the multiple precautions taken by a hospital conglomerate to protect the privacy of medical records in its care, and it got me to asking myself how NCHS measured up. When I was through, I had added a number of dimensions from my own experience. In succeeding years, I have added yet more and this is what I propose to offer you today.

At the outset, please keep in mind that these principles were developed in a statistical agency which has the sole function of collecting information for statistical purposes. Data we collect are put to no administrative use. The considerations I'll present today arose out of the NCHS experience as it went from an agency that produced its data exclusively in printed tabular form, to one producing data tapes and floppy disks that were sold under a minimal user agreement, and which today which produces vast amounts of micro data made available at no cost via the Internet or on CD-ROMS.

So I make no claim that all organizations should be in step with NCHS practices. Many of these considerations may be inapplicable to situations of others. I present them only as questions for consideration.

Our need to change has also been stimulated by increasing public and professional concerns as to the safeguards employed by data collectors including, but not confined to electronic security. They encompass issues of responsibility as well as accountability:

- not just for a guarding passwords, but for proper disposal of paper records,
- not just for data we hold, but for the information we share with contractors, collaborators, and others,

- not only for getting new employees to sign a confidentiality pledge, but for continuously reminding them of the high priority we place on keeping our promises of privacy; applying stiff sanctions when this is not done and recognition when it is.

Overview

After mentioning legal and ethical considerations, I'll discuss the scope and special types of responsibility within an organization - including responsibility for imposing penalties and encouraging rewards. Next I'll bring up the issue of "training" - what we mean by it, who gets it and for how long. Often discussed under the rubric of "data flow", many have pointed to the critical question of who actually sees confidential information and where it winds up. The critical question of data access comes next and receives considerable attention with the question of security, electronic and otherwise following. For those who release data or who publish, the subject treated more and more intensely by statisticians - statistical disclosure limitation - must be discussed. Finally, It's always nice to have something in writing, but not easy to do.

Awareness of Legal or Regulatory Authorities and Requirements

All of us in the federal government are subject to the Privacy Act. Some agencies have special legislation, and in the last two years many have felt the effects of regulations arising from the Health Insurance Portability and Accountability Act. In one way or another, our understanding of what we must and must not do in protecting privacy has to begin with knowledge of who to contact for legal expertise.

So, my point is that the first thing that one should do is to find out, preferably with your lawyer, what laws affect ones' organization, what they cover and what is outside their scope, and what they permit or restrict. Equally important is an understanding of how the law governs our response to requests for our data from other parties and the courts. It is always best to have staff understand their rights *before* they are served with a court subpoena or a request under the FOIA.

Some receive identifiable information from other agencies or organizations (e.g. CMS analytical files, NDI) It is not always easy to determine what protections such files have and this is another area of potential confusion. Data in the federal system must be protected, at least, to the extent required by laws under which they were gathered in the first place. So, there is the obligation to be aware of the laws governing data collected by others as well as those covering data we ourselves collect.

Responsibility

I have the good fortune to work in an organization in which there is someone directly responsible for confidentiality and privacy issues. Moreover, the occupant of this position has always received the full and clear support of institutional leaders.

The naming of one person with responsibility for coordination has been essential to the development of a systematic and comprehensive confidentiality program. When such a person or function is provided visibility, the management and staff have a defined place to bring their concerns and problems. In turn, this lays the groundwork for a more comprehensive and informed confidentiality policy. I cannot stress the importance of a confidentiality coordinator/officer enough. How often has information not been transmitted to or obtained from an organization, simply because it was not clear (to insiders as well as outsiders) who is responsible.

So too has been the awareness of responsibility by those in critical positions in the data collection, processing , and dissemination - for example, data handlers such as programmers and analysts. These persons are usually the interface with external researchers and the public and must be aware of the contents of the files under their control - and who is entitled to access them.

This does not mean, however, that only a few people centrally located and with whatever support from "the top" are *responsible* in the sense of actual "stewardship" of confidential information. All persons who come into contact with such information are held legally responsible by NCHS law and regulation. In a statistical agency it cannot be any other way, for we can never be sure when confidential information might be vulnerable and who would see it. In our case, such is the level of cooperation and sense of responsibility that it became clear to me long ago that if were not for this principle we would soon fail in our mission. For the everyday observation of this important obligation accounts for much of the privacy protection we provide. It comes down to this: How many times have I heard from staff member something like "That doesn't look right, let's ask Al" or "Should we be releasing this file to these folks?" or "What is that document doing out?"

In my experience, I have found that is always better to have a clear cut system of both positive and negative sanctions in place *before* it is necessary to use them. In practice, of course, we can rarely anticipate the kind of situation with which we must deal and when they do occur, they are hardly ever clear cut. We can do our best, however, and at least give some thought and, perhaps, discuss with others their experience and maybe even ask our lawyer what courses of action they would support and what standards they would use in judging appropriate sanctions.

Oh, and let's not forget, that we can achieve compliance not only with penalties. It might be wise to look for opportunities to single out those who make special contributions.

Training

This is a difficult topic to summarize, simply because there are so many things that might included. Unfortunately, we do not have a lot of experience when it comes to training for staff in general. In addition some of the items I've included under "Continuing training" are more properly labeled "Awareness". At any rate, it is important to provide training from the outset

of and in a way that impresses on staff that we take these matters extremely seriously. Our practice is to show incoming staff a video with a compact message reinforced by written materials that are as "digestible" as possible. I've recently begun meeting as many new staff as I can, in person to answer any questions. These messages are further reinforced by posters hanging throughout the agency and by notices of proper document handling and disposal at strategic sites.

Among the tools designed to maintain confidentiality awareness are occasional confidentiality "quizzes", special staff presentations, communication of policy decisions and news items of interest,. Others are the encouragement of research in areas such as statistical disclosure techniques and beliefs and understanding concerning consent. This list could go on, but my feeling is that anything that reinforces the "culture of confidentiality" is appropriate.

Data Storage and Handling

Among the issues with which the public has always been concerned, is the maintenance (read "keeping") in a data base of identifiable information - particularly names and other personal identifiers. In the mid 1970s,. the Privacy Act specifically addressed such "systems of records".

When we consider statistical practices, there are usually very good reasons why we keep personal identifiers such as names and address in our files. In one of our surveys (NHANES), names are kept (separately from survey information) in order to contact respondents who, based on clinical test results, might be in need of medical treatment. In other situations, we may wish to re contact respondents to clarify information they have previously provided or to gather information in a new survey.

While it is important that such uses all be somehow covered in the informed consent process, we must bear in mind that the maintenance of this information without a clear reason can put respondents at unnecessary risk. So we have found it necessary to be sure someone is responsible for such information, that they have an acceptable reason for their maintenance and have set a reasonable date by which time they will no longer be needed and destroyed.

We do not collect data for administrative purposes, but it has long been an important practice to separate data intended for research and those related to respondents rights, benefits and privileges. Once a promise has been made to use information only for research we need to sure that promise is kept.

Access to Confidential Data

It would be nice of all information had an easily understandable "label" which would apply in all situations to insure that no one was granted access to information to which they were not entitled.

Until we do develop something effective, we need to rely on two important steps:

1) A clear statement of what makes information confidential (identifiable), so individuals who control data may have a clear idea of what can be released and, at a minimum, what they have to bring to the attention of a supervisor. I still hear staff saying that stripping names and addresses from a file makes it safe.

2) The second important step is the establishment of a policy on granting access that takes into account:

- assurances of confidentiality made to respondents, .
- variations in access qualifications among staff (not all who work in our offices are federal employees or the equivalent as far as confidentiality is concerned. Even if they are, they don't all need access to confidential data)
- the specific purpose for which access is needed (to justify and to be sure we don't give out more than is needed)
- clear identification of who may grant access
- access site
- any required data use agreements and data tracking procedures
- restricted access (Federal Committee on Statistical Methodology, 2002)

Security

Much has been written about electronic security and others are far more qualified to comment on this than I (National Research Council, 1997) . However, appropriate steps in this area can never be under emphasized. The two points to which I want to draw particular attention are:

1) the need to assure ourselves that parties with whom we share confidential data (whether contractors or collaborators, or those under a data use agreement), *themselves* take, at least, the same precautions that we do. Some agencies, particularly those that employ access under licenses use the mechanism of unannounced visits very effectively.

2) There is the continued need for us to look for soft parts in our "armor" and to be sure that simulated attacks (preferably on an ongoing basis) are part of our security program. Moreover, these attacks or tests should involve physical as well as electronic access. It has often been mentioned that intruders often try physical before electronic intrusion.

Review of Data for Public Release

While a number of federal agencies have already established disclosure review boards (Panel on Disclosure Review Boards of Federal Agencies, 2000), many others have no established mechanisms for reviewing files that they wish to make available for unrestricted use. Those considering some kind of review group, might consider the points shown in this overhead transparency. Some of the existing boards have rotating membership which is an important for

the development of institutional capabilities. I think you will find that a literature which provides exact details on how such boards come to a decision is non-existent. That is because while probably all observe certain fundamental precautions (removal of names, addresses) there are great variations in the amount of detail agencies wish to release in geography, race, income and other important data items.

The Census Bureau's desire to publish information for counties inevitably means less detail for certain variables, whereas NCHS can often provide more detail on age, occupation, and income because little detail below the national region is ever released.

Still, it is essential to be aware of what is referred to as the Statistical Disclosure Limitation literature (Federal Committee on Statistical Methodology, 1994, 1999, Doyle, et al., 2001).

A point that I think needs stressing is the need to apply these standards not only to data that our own organization releases, but also to those published by others under data sharing agreements. It is extremely important that there be a mechanism to insure that the same standards be applied to both.

Written Statement of Policy

My organization first published its Confidentiality Manual in 1978. It was first revised in 1984 and is now in its second revision. As one can imagine, this latest revision was a major one given the revolution in electronic technology and data dissemination mechanisms since then.

Although policy in this area will probably always be subject to change, having the major elements in our confidentiality practices available in written (and, now, electronic) form has been a major advantage in explaining our practices to outsiders as well as providing essential guidelines to staff in their daily work.

Other agencies, such as the Census Bureau and the National Center for Education Statistics, have long had similar documents in place.

Conclusion

Our sister agency, the Census Bureau, has the reputation of one that fosters and maintains a "culture" of confidentiality. Perhaps the ideas that I have presented, together, come close to describing the principle elements of such a culture. Of course the essential element of any culture is the extent to which these principles are highly valued and serve to bring about actual behavior.

References

Doyle, P. et al. *Confidentiality, Disclosure, and Data Access*, Amsterdam: Elsevier, 2001.

Federal Committee on Statistical Methodology, Confidentiality and Data Access Committee, Checklist on Disclosure Potential of Proposed Data Releases, July, 1999. http://www.fcsm.gov/docs/checklist_799.doc

Federal Committee on Statistical Methodology, Confidentiality and Data Access Committee, Restricted Access Procedures, 2002.

Federal Committee on Statistical Methodology, *Statistical Policy Working Paper 22 - Report on Statistical Disclosure Limitation Methodology* May, 1994.

National Research Council, *For the Record* Washington: National Academy Press, 1997,

Panel on Disclosure Review Boards of Federal Agencies: Characteristics, Defining Qualities and Generalizability. Presented at the Joint Statistical Meetings August 17, 2000 Indianapolis, Indiana