STATISTICS AND PRIVACY IN THE NEW MILLENNIUM—EXTERNAL FORCES*

Gerald W. Gates, U.S. Bureau of the Census
U.S. Bureau of the Census, 4700 Silver Hill Road, Washington, DC 20233-3700

Key words: Privacy, Legislation, Regulation

Introduction

In the next millennium, statisticians face considerable pressure from outside forces that attempt to regulate privacy or undertake activities that threaten it. These pressures will come primarily from government, the business community, data users, and the public. For statistical organizations—both public and private—this will require greater attention to the privacy implications of activities that encourage access to and use of personal information. Such activities are primarily related to new uses of technology, including the Internet, and the expanded use of existing public and administrative records.

Traditionally, statisticians see their activities as privacy-invasive only in so far as personal information may be used for non-statistical purposes (that is, making determinations about specific individuals), that confidentiality may be breached, or that the information they ask for may be too sensitive. In the age of networked computers, record linkage, massive databases, and free access to public records, these concerns are heightened. In the future, the mere perception that these concerns are not addressed may create problems for statistical organizations. In the future, failure by other information collecting organizations to respect privacy may have rippling affects on statistics. In the future, data users will demand greater access to information as their abilities expand with increases in technology.

Government is poised to respond to any threats to privacy by creating new laws or regulations. The public is poised to resist the incessant demands for personal information. Data users are poised to push for a partnership in the use of statistical information. Together, these forces have the potential to seriously curtail survey research. My discussion will focus on what is driving the actions of government, industry, data users, and the public and how statisticians can be proactive and avoid major privacy-related disasters and the resulting disruption to their operations.

Government

First, let us look at the government players that may step in to limit privacy threats. Since the Internet has eliminated national boundaries in many aspects, we need to consider the role of both the United States and foreign governments. In the U.S., legislators, regulators, and program administrators have a role to play. In Europe, the European Commission and the Council of Europe are key players.

The U.S. Congress has been active of late in considering legislation to address concerns related to the privacy of medical and financial records. Debate has taken place on the appropriate use of Social Security Numbers to identify individuals and link records. Congress has considered providing for uniform confidentiality legislation and the sharing of data among a selected group of statistical agencies. Also, Congress has recently acted to improve access, under the Freedom of Information Act, to research data collected under Federal grants (the Shelby amendment).

The recent debate over the Shelby amendment to the Omnibus Consolidated and Emergency Supplemental Appropriations Act for fiscal year 1999 has focused attention on protecting the confidentiality of research data collected under Federal grants. Under this amendment, the Office of Management and Budget is required to issue regulations (Circular A-110) ensuring that data collected by scientists/statisticians under grants from Federal agencies are made available under the Freedom of Information Act; just as data collected by Federal agencies on their own account are made available under FOIA. There are many complex issues involved in requiring researchers to provide access to their data. Proponents note that it is difficult to replicate research done under Federal grant due to present restrictions on access. This has been the subject of a National Academy of Sciences report and is clearly a goal of good research. On the opposing side, researchers are concerned about the additional costs and burden, the potential for unfair attacks, threats to exposing proprietary information, and threats to confidentiality. The confidentiality focus is perhaps the most emotional and most misunderstood.

* This paper reports on results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. The report is released to inform interested parties of research and to encourage discussion.

Under the legislation, agencies will get the data from the grantees and must make it available under FOIA. However, agencies currently have several options to withhold confidential information. They can claim a statutory exemption if the data are protected under legal confidentiality requirements (like the Census Bureau's Title 13). They can also claim that release would constitute an invasion of personal privacy. Skeptics believe, however, that it is not quite that simple. Although agencies can and should exercise their privileges under FOIA if they believe that confidentiality is threatened, the nature of disclosure limitation techniques implies that considerable education is needed to inform FOIA Officers of what can be made public use and what is too risky to release.

Regulators have also been active on the privacy front. The Office of Management and Budget issued a Privacy Act directive in May 1998 requiring agencies to analyze and update their Systems of Records Notices that describe identifiable record systems maintained by the agency. The directive also requires agencies to provide clear and comprehensive privacy policies on their Internet sites. These requirements are to be in place by the end of 1999.

Anther key regulator is the Office for Protection from Research Risks in the Department of Health and Human Services, which establishes guidance for agencies in conducting research involving human subjects and oversees the establishment and operation of Institutional Review Boards (IRBs). In 1998, the IRB that oversees the National Health Interview Survey recommended that this survey, which does not have a clinical component, must use a signed consent procedure to ensure that "subjects" fully understand and agree to the conditions of the information collection. Thus, social science research is brought under the same strict consent requirements as clinical testing. The President's National Bioethics Advisory Commission has recently been extended for two more years and continues to review how well the federal policy for human subjects is workiing. The effort may result in more research regulation.

Actions of government program agencies are also raising privacy alarms. With the passage of the Personal Responsibility and Work Opportunity Reconciliation Act of 1996, Congress mandated that the Department of Health and Human Services create and maintain a database of all new hires. This database, sometimes referred to as the "deadbeat dads" database, is to be used by the states and courts to track workers for the purpose of enforcing child support payments. Privacy advocates have been very vocal in their criticism claiming that this database, which will eventually represent every U.S. worker, will be too enticing for government agencies that will use it for unrelated purposes. Government agencies are also under the gun for maintaining the security of their World Wide Web sites. Recently, hackers have successfully penetrated sites of major government agencies including the Justice Department and the White House and changed content. The public's willingness to use new technologies will be affected by the ability of agencies to correct these threats.

State agencies are also taking on activities that threaten privacy. The state of California became involved in a flap earlier this year when it started selling confidential wage records to private companies. Under a newly enacted law, California's Employment Development Department will begin selling salary information to private information companies, car dealers and creditors wanting to check people's annual income. California has traditionally provided this service to government agencies but expanding it to the private sector has raised privacy alarms. Under the new program, designed to reduce fraud, qualified companies can get the information only with the consent of the individuals. However, the Employment Development Department does not require proof that consent was obtained.

In Europe, the European Commission and the Council of Europe are playing key roles in addressing privacy concerns raised by the Internet and other computer-related technologies. The European Union Directive on Transborder Data Flows went into effect in 1998.[1] The U.S. does not meet the Directive's criteria for acceptable levels of privacy protection. As a result, U.S. companies could be excluded from receiving personal information from European affiliates or from their trading partners. This issue is being addressed in the U.S. by the Commerce Department and the new U.S. Privacy Counselor within the Office of Management and Budget who continue a dialog with the European Union on options that meet EU requirements. The current proposal is for U.S. businesses to adopt "safe harbor" principles that meet EU conditions for "notice and choice."

In the area of statistics, the Council of Europe and the European Commission (EC) are also playing key roles. The Council issued its Recommendations for the Protection of Personal Data Collected and Processed for Statistical Purposes in 1998.[2] These guidelines are designed to set standards for statisticians for respecting privacy in the collection and use of personal information. The EC has allocated a large amount of money for research on many social and scientific issues, one of which is statistics. Under the Fifth

Framework for Research, Eurostat has helped identify research topics related to confidentiality and disclosure limitation. Four EU countries have joined in proposing an international research program to study the public's perceptions of confidentiality in light of new ways to collect and disseminate data. The U.S. Census Bureau will conduct similar research in the U.S. should this research proposal be accepted. This will offer the first opportunity to see how perceptions compare across selected European countries and the U.S and to evaluate the role perceptions play on decisions to provide data for statistical research.

The Private Sector

In the United States, activities of the private sector are becoming the focus of great concern about potential threats to privacy. These concerns are magnified by the lack of broad legislation to ensure that information is used only for the purposes for which it was collected. With the trend toward cross-ownership of various business interests by large companies, these concerns become more urgent. For instance, one conglomerate may own an HMO, a retail drug chain, and an insurance company. There is a strong business interest in sharing information across these different functions and the company may not feel a need to seek consent for these uses given that the information in not leaving the corporate entity. Another concern is the trend to electronic commerce. The availability of information on the Internet raises concerns about security and about the proliferation of personal information in electronic databases. The ongoing discussions between the EU and the U.S. are focusing on the lack of enforceability of EU laws in the U.S. Corporate America would prefer that the debate shift to emphasize how competition will bring about a business focus on privacy where privacy-sensitive companies will compete favorably with those that are privacy invasive.

Two recent activities highlight how businesses can make missteps in their move to gather and market personal information. A few weeks ago I got an offer in the mail from a company called DBT Online, Inc. They were trying to sell me their new product Autotrack XP. The ad claimed that one could "easily access more than 4 billion records combining public records and publicly available information including information from major consumer reporting agencies." I could use this service to "quickly identify, locate and profile individuals, assets and businesses." This sounds tempting for a survey organization but it also raises alarms that such a resource is available and that the public is most likely unaware it exists.

Another activity that was reported by *The Washington Post* in June of this year involves a very respected U.S. company--General Electric.[3] In a recent survey of its shareholders--many of them GE employees--the company secretly recorded the names of respondents despite telling them explicitly that this information would not be required. The article's byline noted that "Those sorts of tricks are common in the survey industry." The article went on to say: "In any survey, you should assume your response is not anonymous." Once it was caught, the company said it would drop the practice but the damage to the statistics profession was already done.

Congress is beginning to get serious about business practices that threaten privacy. Specifically, they are debating privacy with regard to medical records and financial records. The Health Insurance Portability and Accountability Act of 1996 included a provision for Congress to enact separate medical records privacy legislation by August 1999 or, failing that, the Department of Health and Human Services was instructed to write regulations addressing this need. That date has come and gone and HHS is busy working on its regulations. On another front, Congress is also considering legislation to allow banks to affiliate with other financial service providers. HR 10 provides for a new regulatory framework for overseeing the conglomerates that would be created under the bill. According to *The Washington Post*, the fight set to go to the floor of the House, "underscores how quickly privacy has leapt from the fringe debates into the mainstream of discussions about the direction of the nation's vast financial services industry. A year ago, privacy was barely addressed when lawmakers debated similar legislation." [4]

Data Users

The third external force consists of data users. In the world of mainframe computers, users of public use microdata were all well funded researchers working under Federal grants or working for large corporations. In the world of the personal computer and the Internet, anyone can become a microdata user—especially with new tools being developed to easily link, tabulate and analyze data. Today, data users are social science researchers, marketers, junk mailers, students, city planners, and even survey respondents. Today's data users are changing the rules for statistical organizations in that they are so diverse, they are becoming more sophisticated, and are better equipped. Today's data users do not have the same motives as social science researchers to protect the "goose that laid the golden egg." Consequently, we need to be concerned that traditional data protection rules are sufficient to deal with users with very different motives.

We are finding that the traditional data users are now only interested in microdata and they want it with information typically removed to protect confidentiality. Survey organizations are responding by creating new access mechanisms and researching new techniques to mask data. Users are pushing the issue further by declaring their rights to the data and their willingness to share the responsibility to protect confidentiality through licensing or bonding.

The Public

The public is the fourth, and the most important external force for survey organizations. We count on the public to willingly provide us their personal information and threats to privacy could easily disrupt this process. In surveys measuring opinions on privacy, the Census Bureau has found that the public doesn't understand new technologies and fears that they may be misused.[5] We also have found that the public is continuing to lose trust in government and does not believe government's promises. When asked about whether agencies keep information confidential, they reply that government agencies freely share information. In fact, they believe that all government computers are connected. When surfing the Internet, people become even more protective of their information and may provide false information to prevent the organization from misusing it. Another apparent fallout of the information age and potential problem for survey takers is the incessant demands from marketers and pollsters for information by phone, mail and the Internet. As the public grows increasing weary and suspicious of these requests, important policy research may suffer as response rates decline.

Observations/Predictions

To put this in some perspective, it will be useful to examine the current statistical climate. As many are aware, the census that will be taken in 2000 has generated considerable debate over the use of sampling to count the population. Little attention has been given to privacy issues. Nevertheless, as Ed Spar, the Executive Director of COPAFS, noted in a recent talk to Census Bureau staff, the census in 2010 is primed to generate considerable attention to privacy. In fact, Ed sees privacy as the big issue in the next decade as the public becomes more and more concerned about the use of personal information. I agree with this observation as recent events will show.

I see several trends that should cause survey organizations to take note. Survey takers are seeing a continuing decline in response rates as people become less willing to spend their time answering questions of marketers and survey takers. When the public writes to complain about having to answer surveys, burden and privacy concerns are most often cited. There is also a growing concern about surveys targeting sensitive populations (e.g. children) and how consent is obtained from these groups. Concerns about recent government abuses raise the real potential that regulation may be forthcoming. Social science surveys with clinical components (e.g. collection of blood or saliva specimens) also raise concerns over consent and the risks to the individual.

The growing use of existing records (public records, private sector tracking databases, and administrative records) raises concerns about the consent provided for these secondary uses and the potential to build a database on every American that could serve to infringe on personal freedoms. Although the statistical uses seem benign to statisticians, a fearful public believing that all government agencies share data may see it differently. Another potential concern involves the trend to provide more data electronically to more users. If we have not sufficiently researched and addressed the added risks, our track record of never breaching confidentiality may end with dire consequences.

These activities are more or less under our control and we can take steps to limit the privacy threats. What are not under our control are the activities of other data collectors. Inappropriate survey procedures like those used by GE or excessive demands for information from other data collecting organizations will sour the public to requests for information that may be needed to address important public policy issues.

Conclusions and Suggestions

There are several things that statisticians can do to address the external environment and head off problems down the road. Recognizing the current issues and concerns, we can work with the government players by being proactive in defending our collection and use of data for statistics. We can make it clear that survey research is not harmful to individuals whereas program uses and clinical studies may be harmful. Consequently, social science research should not necessarily be held to the higher standard. We can spend dollars on security and disclosure research to ensure that technology is not creating an unacceptable risk to confidentiality. We can post meaningful privacy policies to our Internet sites to inform users what information we collect and what we do with it. Finally, we can respond quickly to misinformation about our practices so that it does not lead to inappropriate legislation or regulation.

Inappropriate behavior by fellow data collectors, both public and private, should not be ignored. Statisticians must be vocal critics when necessary. We also need to

respond to misinformation that paints statistics in a bad light. We need to support associations that promote good survey practices and distance ourselves from organizations that do not follow ethical guidelines.

Data users are partners of statisticians and, as such, deserve special consideration. There is no one best way to meet users' needs. We need to provide multiple levels of access to recognize the users' needs and the sensitivity of the data. One option is to provide a secure environment for users to work with confidential data at remote sites.

Perhaps the most important external force—the one driving the activities of all others—is the public. To work with the public we can try to understand what is important to them. We can also reassess, enhance, and explain our security procedures. We can make consistent and re-enforce our assurances of confidentiality. We can actively educate the public on our need for, and use of, personal data. We can ensure that they know about any secondary uses of their information. And, we can work with advocates to understand and address concerns they have about our practices.

---

[1] Directive 95/46/EC of the European Parliament and the Council on the Protection of Personal Data and on the Free Movement of such Data (EU Directive), 1995.
[2] Recommendation No. R (97) 18 of the Committee of Ministers to Member States Concerning the Protection of Personal Data Collected and Processed for Statistical Purposes (Adopted by the Committee of Ministers on 30 September 1997 at the 602nd meeting of the Ministers' Deputies), Council of Europe, Strasbourg, France, 1997.
[3] *The Washington Post*, "Survey says: You're Not Anonymous," by Robert O'Harrow Jr., p. E1, June 9, 1999.
[4] *The Washington Post*, "Bank Information-Sharing Gains on Hill" by Robert O'Harrow Jr. and Michael Grunwald, p E1, July 1, 1999.
[5] Gates, Gerald. and Deborah Bolton (1998), "Privacy Research Involving Expanded Statistical Uses of Administrative Records," 1998 Proceedings of the Government Statistics and Social Statistics Sections of the American Statistical Association, pp. 203-208, Alexandria Va., 1998.