

Warning Concerning Copyright Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

This institution reserves the right to refuse to accept a copying order if, in its judgment fulfillment of the order would involve violation of copyright law.

Printing note: If you do not want to print this page, select page 2 to the end on the print dialog screen.

WINDOW ON WASHINGTON

**Katherine K. Wallman,
Editor**



[EDITOR'S NOTE: In December 1989, the Committee on National Statistics of the National Academy of Sciences-National Research Council and the Social Science Research Council convened a panel to study and provide recommendations on one of the most challenging and complex policy issues facing the federal statistical system: the need to better accommodate the increasing tensions between facilitating data access and protecting respondents' confidentiality.

By the end of 1991, the Panel on Confidentiality and Data Access has been charged to develop recommendations to the federal statistical agencies. To

elaborate on the issues that must be addressed and outline some of the key activities that will be undertaken, I am pleased to welcome as guest columnists the panel chair George T. Duncan (director, Decision Systems Research Institute, School of Urban and Public Affairs, Carnegie Mellon University) and the panel study director Virginia A. de Wolf. The views expressed in this article are those of the authors and not necessarily those of the Panel on Confidentiality and Data Access, the Committee on National Statistics, the Social Science Research Council, or the National Academy of Sciences-National Research Council.]

The Federal Statistical System Amidst the Rising Tension Between Privacy and Data Access

**By George T. Duncan and
Virginia A. de Wolf**

Threats and opportunities challenge the federal statistical system—let's call it FedStat—as it mediates the increasing conflicts between privacy and data access. As an agent of both data provider and data user, FedStat must engage both respondents who increasingly demand to be left alone and clients who increasingly demand more detailed data. Prominently joining FedStat on the stage today are the private-sector information organizations—let's call them PrivStat. At first blush this brings to mind a host of established survey organizations (like Westat, NORC at the University of Chicago, ISR at Michigan, and

Roper) many of which have well-served FedStat in the supporting role of contractor. More recently, emerging from the wings is a new group of private information organizations that take data, package it, and make it available for sale. This article looks at the role of FedStat and PrivStat in the mediation of privacy and data access.

Setting the Stage

The clash between the demand for privacy and the demand for information takes place within a larger setting of technology, economics, and values, each with its own icon-like props. The props for technology are a 20 MIPS and 3.5 MFLOPS computer workstation for under \$9,000 and a 5-ounce pocket phone permitting communication between Peoria and Manila via satellite. For economics, the prop is a *New York Times* front page showing President Gorbachev espousing the merits of the free market. For values, the props are a bronze of Kit Carson riding free and symbolizing individual autonomy together with a photograph of Martin Luther King, Jr., inspiring the Civil Rights movement and symbolizing

individual initiative.

Improvements in computers and communication technology are increasing the tension between the needs for privacy and information. On the one hand, researchers encouraged by the prospect of a desktop Cray want remote communication links to large amounts of disaggregated data so they can better model the complexities of society and better inform public policy. On the other hand, respondents are concerned with computers' growing capabilities to store vast amounts of data about them and to communicate this information without their authorization to others via these same high-speed devices.

Turning to economics, we see PrivStat playing a substantial role in the information market. Aided by computer and communications technologies, PrivStat—including such private entities as marketing firms and newspapers—is amassing and disseminating consequential amounts of data. The considerable capabilities of PrivStat derive in part from deliberate government policy. The federal government, for example, has reserved for private firms the tailoring of certain government-produced data, like products from the Census Bureau's comprehensive geographic system—the TIGER files (Topologically Integrated Geographic Encoding and Referencing files). Another part of this growth is a natural free-market response to a vaulting increase in the demand for data, some of which apparently is not being met by FedStat.

Prevailing political values champion a concern for individual autonomy and individual initiative. Affirming autonomy, we value the secret ballot and the privacy of the home, while we deplore universal registries, and fear record systems that may mark us for life. These values are reflected in such legislation as the Privacy Act. Affirming initiative, we value the right to vote and favor empowerment over entitlement, while we deplore restrictions on free speech. These values are reflected in such legislation as the Freedom of Information Act.

Largely, autonomy and initiative are mutually supportive and indeed define an individual in a democratic society as both independent and empowered. The tension between the two values is undeniable however, and no less so from a statistical perspective. Independence requires privacy, protection from the intrusion of others, including both private information gatherers and the federal government. Empowerment requires knowledge, that is, information about the social, political, and economic environment, including about how individuals are faring in the society. Federal statistical agencies, as both data collectors and data disseminators, must mediate this tension, finding ways to be mutually supported by them, if they are to fulfill their function.

Which Actors Form the Company?

However decentralized it may be, FedStat—with its supporting firms—is but one leading actor in a larger company. Performing with FedStat are at least three other leading actors: the respondent or data provider, the researcher or data user, and the private-sector information organizations—PrivStat. The principal challenge to FedStat's traditional and lead role of passing public policy-relevant data from respondent to researcher comes from PrivStat. In thinking about whether it can maintain its role, how does FedStat currently perform with the other two actors—the respondents and the researchers?

In collecting data, FedStat gets data from individual respondents (e.g., the U.S. Fish and Wildlife Service Survey of Fishing, Hunting, and Wildlife Related Recreation elicits data on outdoor enthusiasts), from households (e.g., the Census Bureau's Survey of Income and Program Participation gets data from 20,000 households), from organizations (e.g., the National Agricultural Statistics Service collects information on some 120 crops and 45 livestock items), or governmental subunits (as the National Center for Health Statistics collects data from the 50 states).

In disseminating data, FedStat provides information to its clients. These clients span a range including researchers within statistical agencies, other federal government agencies, state and local government, business firms, Congress, the Administration, the media, and individual members of the public, as well as academic researchers.

The role of PrivStat as both a data collector and a data disseminator is growing rapidly. As a data collector, PrivStat, it might be thought, could not compete effectively with FedStat because it lacks the same politically mandated powers to command response that some federal government statistical agencies, such as the Census Bureau, have. Yet, PrivStat does have massive technological capability to obtain response, for example through random-digit dialing, and the money to buy the latest technology. So its capacity for rapidly acquiring improved technology argues for an increased role as a data collector for PrivStat.

PrivStat does have economic clout. And its actions are grounded in economics, for example, when credit card companies collect information on individuals' transactions and make it available to interested parties for a fee. PrivStat may have cost advantages in the collection of data since often the data are captured through normal business operations.

We can explore how the values of individual autonomy and individual initiative are implemented in legislative guidelines and administrative practice for FedStat and PrivStat. Generally, although some

legislation exists, PrivStat does not operate under the confidentiality strictures that FedStat does. Over time, certain legislative and administrative safeguards have been developed within FedStat to guard against abuse of respondents. Census Bureau policy, for example, does not allow release of public-use files that have administrative records linked to survey data.

Motivated by a lack of safeguards in the private sector, congressional hearings have been held by Representative Robert Wise (Democrat-WV), chair of the Government Information, Justice, and Agriculture Subcommittee of the Committee on Government Operations. A May 16, 1990, hearing focused on data protection, computers, and changing information practices. Chairman Wise in his opening remarks illustrated a wide range of threats to personal privacy, notably:

- A new CD-ROM product available this year will include information on 80 million households and 120 million consumers. The database will include names, addresses, estimated income, and propensity to buy over 100 consumer product categories.
- Travel agents, airlines, car rental companies, and others in the travel industry are fighting over ownership of information about an individual's travel plans maintained in computer reservation systems. Travelers are not aware of the extent to which the industry is trafficking in their private travel plans.
- Some hospitals are using identifiable patient information to compile mailing lists used for selling services through direct mail.

Congress has begun to protect privacy in the private sector information market. The Video Privacy Act (popularly known as the Bork bill), for example, was enacted to protect the privacy of video rental records. Clearly this legislation is of limited scope and has little relevance to FedStat. But it does suggest a general direction of more legislative protection of privacy rights. In its legitimate concern to regulate against certain abuses in the private sector, Congress should not inappropriately burden the public-sector statistical organizations that already have adequate safeguards in place. In the case of either FedStat or PrivStat, a concern is that it seems easier to draft legislation in terms of absolutes, suggesting perhaps zero disclosure risk when this must be conceptually impossible and giving inadequate attention to tradeoffs of privacy with the social benefits of data access.

Whether the organization collecting the data is public or private, to ensure the integrity of research findings, there must be data access by researchers

who are independent. Such researcher access to data allows reanalysis by groups with different agendas; stimulates new inquiries on important social, economic, and scientific questions; improves the quality of data by suggesting improved measurement and data collection methods; and provides information to improve forecasts and resource allocation. A reasonable concern is that the profit motivation for PrivStat will unduly restrict researcher access to those with the ability to pay.

While FedStat has provided one of the richest and most accessible statistical systems in the world, there are instances where agencies have not been adequately forthcoming in providing data that users want. Exemplifying problems in this area, the Bureau of Economic Analysis declined in July 1990 to release the 1989 report on the U.S. Net International Investment Position, a figure that since 1973 had assessed the imbalance between what Americans owned overseas and what foreigners owned in this country. BEA did not release this report because of concerns about data quality problems, particularly in assessing the current value of both foreign and U.S. holdings. It is not unreasonable to forecast that PrivStat will try in some way to fill such a void since there is an apparent demand for the data product.

How Are the Reviews?

Periodically, we need a systematic review of how FedStat is responding to the tensions between confidentiality and data access. Ten years ago, FedStat faced the competing demands of a new Privacy Act and a new Freedom of Information Act. Today, FedStat deals with a decade of legal precedents and administrative experience. Documenting legal precedent, the U.S. Department of Justice's Office of Information and Privacy periodically publishes a compilation of judicial decisions, both published and unpublished, addressing access issues under both Acts. Today, we have as well the up-and-coming actor PrivStat that FedStat must work together with constructively.

To help provide a review for the 1990s, the National Academy of Sciences-National Research Council's Committee on National Statistics and the Social Science Research Council jointly convened a panel in December 1989, to develop recommendations to federal agencies for better accommodating the increased tension between data access and confidentiality. Working over a two-year period, the twelve-member panel will draw on its background in the structure and policy of FedStat, privacy and information access regulation, the research use of federal databases in economics and demography, and the development of disclosure limitation methods. The panel's activities are financially supported

by the National Science Foundation, the Bureau of the Census, the Bureau of Labor Statistics, the Internal Revenue Service's Statistics of Income Division, the National Institute on Aging, the National Center for Education Statistics, and other federal agencies.

Focusing on access to federally collected microdata, the panel is planning an invitational conference on disclosure limitation. The conference will explore basic value issues in the tension between privacy and freedom of information and assess measures of disclosure risk and harm, including economic harm to firms. Also, it will examine statistical disclosure limitation procedures, including researcher guidelines for the analysis of masked data, and the impact computer technology has on the problem. In addition, participants will assess current legislation regarding confidentiality of federal data and current agency administrative approaches to disclosure limitation. Finally, the conference will consider feasible alternatives to the status quo.

Complementing the work of this first conference, the panel is planning a second invitational conference that will deal with respondent impacts and cooperation. Informed consent agreements are being increasingly proposed as a vehicle for providing researcher access to previously inaccessible data sets or new data sets that may arise from surveys to which administrative or other data can be matched. Conference participants will assess what constitutes appropriate informed consent from both individuals and establishments. They also will examine the potential for harm to respondents from research access to original data or from linkage of data with other information. Finally, the consequences of informed consent procedures on response rates and quality of response will be discussed. Among the participants will be researchers on the ethics of social science research.

As part of its assessment, the panel is assembling information about the federal statistical agencies' policies and procedures related to confidentiality and access. Some questions of interest include:

- How do the authorizing statutes and regulations of the federal statistical agencies differ?
- How have certain regulations made data access or data collection by other federal agencies difficult?
- What procedures are used by statistical agencies to limit disclosure risk?
- How do organizational arrangements to establish confidentiality guidelines and to review requests for access differ among the statistical agencies?

While it is obviously premature to suggest what

these two conferences and this data gathering activity may garner, as individuals we would like at this juncture to raise two thoughts for consideration:

As a first thought, we believe that greater coordination of statistical programs within FedStat may save money, provide a higher quality product, and pose little disclosure risk. For example, the U.S. Department of Agriculture and the Census Bureau both conduct farm surveys, yet each develops its own sampling frame for the same population. Can a common frame be developed? What are some hindrances? As another example, consider longitudinal research on individuals. As the cohort ages, there is a potential for cost saving in the transfer of data sets among agencies. What problems and potential do such transfers engender? Is it too far afield to also envision the possibility of mutual gain from greater cooperation in activities between FedStat and PrivStat?

As a second thought, looking to the larger context of FedStat, we sense that each statistical program could benefit from an evaluation component directed at both respondents and users. For certain programs such evaluations are either currently undertaken or are part of long-range plans. Eliciting input from respondents, the Bureau of Labor Statistics has revised its CPS questionnaire using cognitive research on the understanding of questions about family and labor market conditions. Eliciting input from users, the Census Bureau in planning for the 1992 Economic, Agriculture, and Government Censuses will find out the need for new data products, how to improve presentation and dissemination of data, and what data sets are not needed. These kinds of studies could well be a systematic part of all statistical programs.

The Final Act

Today, FedStat faces challenges and opportunities that are unprecedented. The demands of a global society and rapid technological change increasingly press for timely and sophisticated understandings of our social and economic system. At the same time, respondents are increasingly busy and concerned about intrusive data gatherers. Will FedStat be squashed or uplifted by the conflicting demand for information and demand for privacy? Can it balance the rising and conflicting demands of respondents and researchers? Can it act synergistically with PrivStat? To quote Francis Quarles, a 17th-century English poet:

Judge not the play before the play is done;
Her plot hath many changes; every day
Speaks a new scene; the last act crowns the play.