Journal of Official Statistics
Vol. 9. No. 2, 1993, pp. 271–274

Report of the Panel on Confidentiality and Data Access

George T. Duncan, Virginia A. de Wolf, Thomas B. Jabine and Miron L. Straf[1]

1. Introduction

External factors are rapidly changing the environment in which the U.S. federal statistical agencies operate. Data users, armed with ever more sophisticated analytical techniques and computing power, are chafing at the restrictions imposed, in the name of confidentiality protection, on their efforts to gain access to government data bases and to link different data sets. Faced with a proliferation of requests for information about themselves and concerned about who will have access to their information, individual data providers are becoming more resistant to attempts to persuade them to participate in surveys, whether mandatory or voluntary. And federal statistical agencies, concerned that some breach of confidentiality may adversely affect the trust and cooperation of survey respondents, are exercising more caution in decisions on what data to provide.

American advocacy groups, such as the American Civil Liberties Union and Computer Professionals for Social Responsibility, are paying greater attention to privacy and confidentiality questions. Companies in the private information industry are developing large personal mailing lists and data bases for purposes of marketing, credit evaluation, and other uses that affect all of us. The American Public's awareness of actual or potential abuses in the commercial information sector threaten to have an adverse spill-over effect on major government statistical programs like the decennial census of population.

2. The Panel on Confidentiality and Data Access

To study the problems these factors raise for federal statistical agencies, the U.S. Committee on National Statistics, of the U.S. National Academy of Sciences – National Research Council, joined with the Social Science Research Council to convene in 1989 the Panel on Confidentiality and Data Access. Support for the panel's work was provided by several U.S. federal government agencies.[2] Members of the panel were appointed for their expertise in areas

[1]George T. Duncan, chair of the panel, is Professor of Statistics at Carnegie Mellon University; Virginia A. de Wolf, who served as study director, is now with the Bureau of Labor Statistics; Thomas B. Jabine, is a consultant to the panel; Miron L. Straf is Director of the Committee on National Statistics. The panel was also assisted by Robert Pearson, David L. Szanton, and Wlodek Okrasa of the Social Science Research Council and by Michele L. Conrad of the Committee on National Statistics.

The work of the panel was greatly aided by the generous responses of statistical agencies and others to its requests for background information and its invitations to participate in our meetings and workshops.

[2]Support for this panel study was provided by the National Science Foundation, the Bureau of the Census, the Bureau of Labor Statistics, the Internal Revenue Service Statistics of Income Division, the National Institute on Aging, the National Center for Education Statistics, and, through their contributions to the work of the Committee on National Statistics, many other federal government agencies.

of importance to the panel's deliberations, for example: issues of ethics, privacy, and respondent cooperation; public policy; legislation; the history of the federal statistical system; international perspectives; and statistical methodology. A list of panel members is included in the last section of this article.

The purpose of the study was to develop recommendations that could aid federal statistical agencies in their stewardship of data for policy decisions and research. Three areas were of paramount concern to the panel: protecting the interests of data subjects through procedures that ensure privacy and confidentiality, enhancing public confidence in the integrity of statistical and research data, and facilitating the responsible dissemination of data to users.

The panel's work builds on previous studies, in particular of the Committee on National Statistics and the Social Science Research Council, as well as on four workshops. As case studies, three workshops addressed issues of confidentiality and data access for the Longitudinal Retirement History Survey, the Doctorate Records File and the Survey of Doctorate Recipients, and data collected by the National Center for Education Statistics. The fourth workshop, on disclosure limitation approaches and data access, led to the papers that appear in this special volume.

To make its task manageable, the panel decided to concentrate its attention on major statistical programs of the U.S. government and to look beyond them only to the extent that seemed necessary to address confidentiality and data access questions related to those programs. Given their backgrounds and experience, most members of the panel felt more comfortable dealing with these issues as they apply to personal, rather than organizational data. Nevertheless, major issues relating to data for organizations cannot be ignored, and findings and recommendations pertaining to them are in the report. Confidentiality and data access questions for organizations, however, need more attention by statistical agencies and others.

The dynamics of the trade-off between confidentiality and data access are complicated and, in the U.S. decentralized statistical system, are heavily influenced by individual agency missions and environments. There are no simple solutions that can work for every agency. On the other hand, it was well beyond the panel's abilities to recommend detailed solutions for specific problems faced by each agency with regard to confidentiality legislation, procedures for informed consent, dissemination of public-use data sets, and other matters related to confidentiality and data access. The panel's goal, therefore, was to formulate broad principles and guidelines that would be helpful to each of the statistical agencies in developing its more detailed policies and practices.

Despite the unforgiving nature of the trade-offs, the panel is optimistic about opportunities for improvement. The panel believes that at present there are ways to improve data access without decreasing confidentiality protection and ways to increase confidentiality protection without diminishing data access. In other words, there are some possible courses of action that can properly be described as "win/no lose."

There are many players in this arena. Most directly involved are data providers (the individuals and organizations asked to provide information about themselves in censuses and surveys), data users, and the statistical agencies. The statistical policy office in the U.S. Office of Management and Budget plays a significant role in the establishment of standards for confidential-

ity and data access. The U.S. Congress defines the missions of the statistical agencies and establishes the underlying requirements and constraints for confidentiality and data access. Also important are the custodians of administrative records, agencies like the U.S. Internal Revenue Service and the U.S. Social Security Administration, whose extensive systems of records about individuals have important statistical uses. Finally, there are may advocacy groups that explicitly represent groups of data providers and data users in various ways or that seek to represent the general public, whose views on these subjects are not always well-formed or well-informed, but are nevertheless of fundamental importance.

The majority of the panel's recommendations are addressed to statistical agencies or the statistical policy office and several recommendations would require legislative action. Others are meant primarily for data users outside of the government and for the organizations that represent them. The report tries to make the subject more accessible to all of these groups by including numerous examples of relevant policies and practices of the statistical agencies.

3. Problems Addressed by the Panel's Report

The panel's report, *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics*, will be published by the National Academy Press[3] in late 1993. At that time the panel's recommendations can be released and discussed. The recommendations address five major problems, each of which is an aspect of the trade-off between confidentiality and data access. The panel develops broad principles for its recommended solutions to the following:

3.1. Protection from mandatory disclosure

Many statistical agencies lack adequate legal authority to protect identifiable statistical records from mandatory disclosures for nonstatistical uses. The Energy Information Administration was recently asked to provide survey responses from companies to the Department of Justice's Antitrust Division for use in compliance activities, and it was ruled that the statistical agency was required to do so. In this instance the agency was in the end able to avoid releasing individual company data, but its perceived ability to protect confidentiality has been damaged. Inability of some statistical agencies to protect individual data from such disclosures may lead to serious deterioration of the completeness and quality of the survey data they collect.

3.2. Barriers to data sharing within government

Some of the laws that govern the confidentiality of statistical data prohibit or severely limit interagency sharing of data for statistical purposes. Laws that control access to administrative records, such as reports of earnings covered by Social Security, restrict their use for statistical purposes. These barriers to data sharing for statistical purposes have either prevented important analyses or hampered other statistical activities or else they have led to costly duplication and additional demands on individuals and organizations who are asked to supply information.

3.3. Access to data by nongovernment users

Because of legitimate concerns about the possibility of disclosure of individual information, statistical agencies have limited

[3]For information on how to obtain the report, write or call the National Academy Press, National Academy of Sciences, 2101 Constitution Avenue, N.W., Washington, D.C. 20418, U.S.A.

the amount of detailed data provided to nongovernment users in tabulations and public-use microdata files. This lack of detail limits the ability of users to do analyses that could contribute to the understanding and resolution of significant economic and social problems. Some agencies have developed mechanisms for providing access to more detailed information on a restricted basis, but present arrangements do not meet all legitimate needs.

3.4. Privacy concerns, declining cooperation in surveys

The public feels increasingly, and with some justification, that its privacy is being eroded by organizations that develop and control the use of large data bases, which contain detailed information about them. Linkage of data from different sources is perceived as a particular threat. For these and other reasons, statistical agencies are finding it more difficult to persuade persons and organizations to participate in statistical surveys, whether voluntary or mandatory.

3.5. The changing technical environment

Technological advances in computers and communications offer both opportunities and threats: opportunities to process, access and analyze large data sets more efficiently and threats of unauthorized access to individually identifiable data.

4. Members of the Panel

The panel was chaired by George T. Duncan, H. John Heinz III School of Public Policy and Management, Carnegie Mellon University. Other members of the panel are James T. Bonnen, Department of Agricultural Economics, Michigan State University; Joe S. Cecil, Research Division, Federal Judicial Center; Martin H. David, Department of Economics, University of Wisconsin; Ruth R. Faden, Department of Health Policy and Management, Johns Hopkins University; David H. Flaherty, Social Science Centre, University of Western Ontario; F. Thomas Juster, Survey Research Center, University of Michigan; Gary T. Marx, Department of Urban Studies and Planning, Massachusetts Institute of Technology; William M. Mason, Department of Sociology, University of California, Los Angeles; Donald B. Rubin, Department of Statistics; Harvard University; Eleanor Singer, Center for the Social Sciences, Columbia University; and William H. Williams, Department of Mathematics and Statistics, Hunter College.

Received May 1993