



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

TOWARD RESOLUTION OF DATA ACCESS CONFLICTS

Proceedings of a symposium at the joint annual meeting of the:

American Agricultural
Economics Association

Canadian Agricultural
Economics and Farm
Management Society

Western Agricultural
Economics Association

Association of
Environmental and
Resource Economists

Vancouver, British Columbia

August 4-8

1990

STP
Quarto
MNUX93
B180

AVENUES FOR RESOLVING DATA SHARING CONFLICTS: A PANEL STUDY BY
THE COMMITTEE ON NATIONAL STATISTICS AND SOCIAL SCIENCE RESEARCH
COUNCIL

George T. Duncan
School of Urban and Public Affairs
Carnegie Mellon University
Pittsburgh, PA 15213

AVENUES FOR RESOLVING DATA SHARING CONFLICTS: A PANEL STUDY BY
THE COMMITTEE ON NATIONAL STATISTICS AND SOCIAL SCIENCE RESEARCH
COUNCIL

by George T. Duncan

I am grateful for this opportunity to share with you some thoughts about the scope and direction of the National Academy of Sciences Panel on Confidentiality and Data Access. This panel, which I chair, is eight months into a two-year effort. It has the able support of Miron Straf of the Committee on National Statistics, Robert Pearson of the Social Science Research Council, and our Study Director, Virginia de Wolf.

The goal of the panel is to provide workable recommendations to federal agencies for better accommodating the increasing tension between data access and confidentiality.

Basic support for this task is being provided by the National Science Foundation, the Bureau of the Census, the Bureau of Labor Statistics, and the Internal Revenue Service Statistics of Income Division, as well as other government agencies.

The work of the panel builds on previous programs of CNSTAT and SSRC, in particular the work that led to the publication in 1985 of the report and edited volume, Sharing Research Data. Two workshops addressing issues of confidentiality and data access have also been

held to provide input to the panel. The first concerned the Longitudinal Retirement History Survey and the second concerned the Doctorate Records File and the Survey of Doctorates.

The workshop on the Longitudinal Retirement History Survey was held in September 1987. Chaired by Jerry Hausman of MIT, the workshop was requested by the National Institute on Aging and the Bureau of the Census. The immediate purpose of the workshop was to evaluate the feasibility of an NIA proposal to reinterview surviving panel respondents and spouses of decedents from the Longitudinal Retirement History Survey. The survey was conducted by the Social Security Administration between 1969 and 1979. Of particular concern to our workshop were the legal, ethical, and policy questions involved in recontacting the respondents and linking their new data with their earlier data and making the resulting microdata files available to researchers. Additionally, mortality information would be obtained through the National Death Index and data on Medicare benefits would be obtained from files of the Health Care Financing Administration.

This workshop recommended obtaining informed consent from respondents or their decedents as the most promising option for making the longitudinal microdata files available to researchers. Following the workshop, a pilot study was undertaken by the Census Bureau, in which respondents to the older female cohort of the National Longitudinal Survey of Labor Market Experience were asked

to explicitly permit the linkage of their records to administrative data, while the respondents are alerted to the possibility that such linkage increases the risk of disclosure.

At the workshop, it was suggested that access to restricted microdata by special sworn employees of the Census Bureau might be allowed at locations other than at Census Bureau headquarters in Suitland, Maryland, and a pilot demonstration of such controlled access was subsequently undertaken. Under a joint statistical agreement with the Census Bureau, researchers at Harvard University have been analyzing data on population coverage in a 1986 test census conducted by the Census Bureau in a section of Los Angeles.

I chaired a second workshop held in November, 1988 to discuss how greater access to the Doctorate Records File and Survey of Doctorate Records could be provided while maintaining the confidentiality of these records. The workshop was supported by the Division of Science Resources Studies of the National Science Foundation, with the cooperation of the Department of Education, the National Endowment for the Humanities, the National Institutes of Health, the Department of Agriculture, and the Office of Scientific and Engineering Personnel of the National Research Council.

Legal constraints did not appear to provide a barrier to access since under the Privacy Act of 1974, disclosure is permitted for a

"routine use", defined as "the use of such a record for a purpose which is compatible with the purpose for which it was collected".

Practical constraints arise from concern about response rates and the prospect of disruption to the data collection programs if there is a failure to protect the data from inappropriate use. A passive waiver experiment in 1981 suggested little problem with response rates. Further, an examination of trends in response rates as confidentiality statements have changed in the Survey of Earned Doctorates (which is used to construct the Doctorate Records File) also suggests little problem with response rates.

Ethical constraints present some of the most difficult problems, and center around the informed consent issues. The workshop suggested approaches appropriate for each of the following areas: (1) expanding and revising the informed consent governing future surveys, (2) the use of existing data under a new consent agreement in which you return to respondents to "renegotiate" this agreement, and (3) the extended use of existing data under current consent agreements.

I would like to turn now to what are current activities.

The composition of the panel is shown in the first figure. This is an able group that has made contributions to the structure and policy of the federal statistical system, the legal bases of

privacy and information access regulation, the research use of large federal data bases for economics and demography, the rights and concerns of respondents, international comparative studies of confidentiality regulation, the conduct of large scale surveys, research on privacy issues, the statistical methodology of sample surveys and observational studies, and the development of statistical disclosure limitation methods.

WHAT IS THE PROBLEM?

The Federal statistical system, which I will call FedStat, must engage both respondents that increasingly demand to be left alone and clients that increasingly demand more detailed data. Whether FedStat will be squashed between these conflicting demands of privacy and data access or will be uplifted as it mediates these demands depends on how it deals with certain issues of technology, economics, and values.

Regardless of the organization collecting the data, whether 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.

If in providing data access to researchers, a disclosure of personal information should happen to take place, it may have legal and other consequences to the respondent, the data-disseminating agency, and the researcher. These consequences can be thought of in a variety of ways, including what persons are affected, how they are affected, whether the agency had a reasoned confidentiality policy in place, and whether the information might already be public and accessible (e.g., vital records, probate records, property rolls, etc.). There may also be systemic damages, including loss of data quality associated with higher refusal rates and increased rates of evasive responses, as well as loss of trust in government. Indeed, respondents seem increasingly wary, in some cases nonresponse rates to surveys are up, and there is a general concern about privacy invasion and social control. Clearly, uncontrolled access flies in the face of privacy concerns.

Government agencies have in many cases not been adequately forthcoming in providing data that users want. Highlighting 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 cited data quality problems, particularly in assessing the current value of both foreign and U.S. holdings, as their reason for stopping the report.

Researchers are also not blame free. Focusing excessively on priorities within their own community, they have too often not aggressively fed the results of their work back to the citizenry. Some have been reluctant to share data with other researchers, thereby raising barriers to reanalysis.

WHAT CAN BE DONE?

Focussing on access to federally-collected microdata, the panel is planning a 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. It will also examine statistical disclosure limitation procedures, including researcher guidelines for the analysis of masked data, and the impact computer technology has on the problem. Participants will also 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 conference that will deal with respondent impacts and cooperation. Informed consent agreements are being increasingly proposed as a vehicle for providing researcher access to heretofore 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 will also 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. Participants will include 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 follow:

- How do the authorizing statutes and administrative arrangements 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?

While it is obviously premature to suggest what these two conferences and this data gathering activity may garner, I would

like as an individual to raise just two thoughts for consideration:

As a first thought, coordination of statistical programs 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 of the 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?

As a second thought, looking to the larger context of FedStat, each statistical program ought to have built into it a systematic 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 Governments Census 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 ought to be a systematic part of all statistical programs.

In conclusion, given the interests and capabilities of the American Agricultural Economics Association--particularly in its concern for quality federal data--I hope that today begins a period of fruitful collaboration with you. The panel would very much like to draw on the resources that you represent. The result, hopefully, is a better mediation of the tension between confidentiality and data access.

Thank you.

Panel on Confidentiality and Data Access
Members as of September 15, 1989

George Duncan (Chair)
School of Urban and Public Affairs
Carnegie Mellon University

James Bonnen
Department of Agricultural Economics
Michigan State University

Joseph Cecil
Federal Judicial Center
Research Division

Martin David
Department of Economics
University of Wisconsin

Ruth Faden
The Johns Hopkins University

David Flaherty
Centre for American Studies
The University of Western Ontario

F. Thomas Juster
Institute for Social Research
University of Michigan

Gary Marx
Department of Urban Studies
and Planning
Massachusetts Institute of Technology

William Mason
Population Studies Center
University of Michigan

Donald Rubin
Department of Statistics
Harvard University

Eleanor Singer
School of International Affairs
Columbia University

William Williams
Strategic COMAPS, Inc.

Staff

Robert Pearson
Social Science Research
Council

Miron Straf
Committee on National
Statistics
National Research Council