Journal of Official Statistics
Vol. 9. No. 2, 1993, pp. 591–593

Discussion
Statutes and Administration

Nancy J. Kirkendall[1]

1. Introduction

The two papers by Jabine and Cecil provide an informative review of the differences between U.S. federal statistical agencies in terms of their ability to protect individually identifiable data and the innovative methods they are using to provide access to data for research.

2. Restricted Access

Jabine's paper discusses different arrangements which have been made to provide access to sensitive data. He concentrates on restricted access, meaning that special arrangements must be made with the users before access is permitted. The data so accessed range from individually identifiable data (for example, data made available to sworn Census agents); to data which have been stripped of identifiers and subjected to disclosure limitation procedures (for example, the public use data sets available from the National Center for Health Statistics (NCHS) to any one from the public who signs a data use agreement form.)

A focus on restricted access is useful, particularly since among the most recent "success" examples are innovative license agreements granting restricted access to users at their own work sites or worksites other than Washington D.C., and the National Center for Education Statistics (NCES) procedures based on the Hawkins-Stafford Act of 1988 that makes users responsible for guaranteeing the confidentiality of the identifiable data they receive. Perhaps it will be through a review of methods of granting restricted access to data that we will be able to insert some consistency into the federal government's handling of the confidentiality issue. Restricted access is our best hope for resolving problems of research access to data in the near term.

Jabine recognizes two levels of data sharing: (1) for intermediate statistical purposes such as frames development; and (2) for primary statistical purposes such as research. In some sense, the issues surrounding sharing data for intermediate use are more complicated that the issues surrounding sharing of data for research.

3. Providing Data for Research

Government agencies pursue and support research in better methods of disclosure limitation. They work hard to balance protecting sensitive data from disclosure while providing information concerning important relationships in the data. There are no perfect answers yet, but there is interesting and innovative work going on. Additionally, government agencies do produce and distribute public use data sets. Perhaps

[1] Energy Information Administration, EI70, 1000 Independence Avenue, S.W., Washington, D.C. 20585, U.S.A.

they need to be improved, but it is clear that the agencies want to provide public access to data. Finally, as mentioned above, there have been innovative ways of providing restricted access to more detailed data.

Lambert's paper demonstrates that it is virtually impossible to absolutely protect microdata by disclosure limitation techniques. This means that we need some other, more secure way, of providing information for research. The alternative ideas discussed in this set of papers includes (a) restricted access, (b) secure computer data base procedures, and (c) release of "phony" data instead of the real thing. When the feasibility of these ideas have been demonstrated, government agencies will use them.

Reynolds's idea about a Federal Data Base Review Board is interesting. However, the only function such a board could have in our current legal environment is to serve as a central point of contact for researchers desiring access to federal data. It could help them identify the appropriate agency and guide them in preparing an official request for access. It would be a service for researchers, and it would minimize the burden on committees like the Census Microdata Review Board, because the requests they receive would be of better quality.

4. Data Sharing for Intermediate Statistical Purposes

Unfortunately, intermediate statistical purposes involves the sharing of individually identifiable data with other U.S. government agencies. The impediment to this level of data sharing is the law. Even revisions to the law which are currently being considered would allow data sharing only among selected agencies.

Cecil's paper provides an informative review of the Privacy Act of 1974, and of the legal responsibilities of three federal statistical agencies in the protection of individually identifiable data. Cecil's paper describes problems faced by the Census Bureau, NCES and NCHS in balancing the protection of respondent level data with providing statistical information to the research community. The three agencies discussed each have a strong legal responsibility to protect data, and the authority to do so. Many of the agencies which are not included try to protect individually identifiable data, but do not have full legal authority.

As Reynolds points out in his paper, the United States is unusual in keeping list information for companies confidential. The existence of a company is known publicly, and a list of companies would not cause competitive harm. It seems silly that many federal agencies have to buy *Dunn and Bradstreet* and the *Yellow Pages* in order to prepare their sampling frames.

The only complaint I can see companies making about list sharing among government agencies is the perception that they might receive more survey forms. If this were the only harm, what would it take to allow or even encourage more list sharing? Many of Jabine's examples illustrate problems in sharing because of legal barriers. However, the harm question is not addressed. (As Plewes mentioned, the three harms: getting another survey form, competitive harm, and being accused of tax abuses by the Internal Revenue Service are treated as equal problems under the law.)

My agency, the Energy Information Administration (EIA) is one of the agencies which cannot receive individually identifiable data from other agencies. Our law does not allow us to protect the data either from other offices in the Department of

Energy or other Federal Agencies which request the data for official (not necessarily statistical) use.

In fact, EIA does not even share individually identifiable data from our energy consumption surveys with itself. Instead, for the voluntary Residential Energy Consumption and Commercial Building Consumption Surveys, data are collected and processed by contractors who guarantee that they will not provide anyone with the individually identifiable data, even EIA. The mandatory Manufacturing Consumption Survey is conducted for EIA by the U.S. Census Bureau, and is protected under Title 13. Needless to say, EIA does not have access to those data, either.

It is only these consumption surveys that are considered sufficiently sensitive to require that EIA protect the individually identifiable data from itself. EIA does its best to avoid sharing of establishment level data collected for statistical purposes with other agencies who need it for nonstatistical purposes, although the law does not guarantee that EIA can protect it.

Clearly legislative solutions are the only way some of these problems will be resolved and agency approaches to confidentiality can be made more consistent. My initial reaction to the need for legislative change was one of dismay. What can we as statisticians do if problem solving requires changes in the law? I was encouraged to hear from Cecil that there are current legal initiatives underway. Perhaps solutions are ultimately possible.