

# IMPLEMENTING NCES' NEW CONFIDENTIALITY PROTECTIONS

Doug Wright and Susan Ahmed, NCES\*  
Doug Wright, 4804 Jasmine Dr. Rockville, MD 20853

KEY WORDS: Confidentiality, disclosure avoidance

## I. Background

Recently, in July 1988, Public Law 100-297 (Stafford Hawkins Act) was enacted, and this legislation has had a significant impact on how the National Center for Education Statistics (NCES) conducts business. This new thrust comes at a time when the Center has been requested to greatly increase its collection of education data and to provide for greater geographic detail and more linkages across the different education levels. NCES is working to find solutions to provide for the confidentiality of data while serving the data needs of the education community. The following paper will describe what NCES has done and some of the special issues it faces.

## II. Privacy Act of 1974

The principal purpose of the Privacy Act is to protect individually identifiable information maintained by the Federal government that is indexed by a unique identifier assigned to an individual, such as a name or social security number. Such information is considered part of a system of records, and the agency maintaining the information must publish a notice of the system of records in the Federal Register.

The Privacy Act also gives the right of examination and correction to the persons whose records are on file and prescribes the permissible uses (routine uses) of those records.

Until recently, NCES did not have "systems of records" for its individually identifiable information. NCES has now published a system for all its longitudinal surveys including the School and Staffing Survey and the various longitudinal series, National Longitudinal Survey, High School and Beyond, National Educational Longitudinal Survey, and the Postsecondary Longitudinal Survey. Another system is being developed for the National Assessment for Educational Progress, which until recently had been a grant. The individually identifiable information in these systems is covered under PL 100-297 as are the individual data in NCES' other surveys.

## III. Public Law 100-297

Under PL 100-297, disclosure of confidential information under the Privacy Act has been greatly restricted, and the penalties for disclosure have been greatly increased. Also, the new law carries with it the additional protection of the confidentiality of schools from the National Assessment for Educational Progress survey, treating them as individually identifiable information.

The law basically requires that no person may use any individually identifiable information for any purpose other than the statistical purposes for which it is supplied, and that only persons authorized by the Commissioner may examine the individual reports. These data are protected and immune from legal process unless one obtains the consent of the individual concerned. In addition, anyone who is granted access to these data must be sworn to uphold the confidentiality provisions of the statute or, if they are employees of NCES, take the oath of office. There are stiff penalties of up to 5 years in prison and/or up to \$250,000 fine for illegal disclosure.

## IV. Implementation of the New Law

NCES was faced with the problem of how to be responsive to the education community's need for data while adhering to the new confidentiality provisions. NCES had two solutions, one for individual information from "public use tapes" and another for individually identifiable information for which the public use tapes would not suffice.

For "public use tapes," NCES had first to develop a system of records, criteria for conducting disclosure analyses, and a system for reviewing tapes proposed for public release. For methods of sharing confidential data outside of NCES, NCES consulted our Office of General Counsel and outside groups.

### A. Conference on Confidentiality Issues and Sharing Data

In September 1989, NCES convened three

panels to discuss the implications of the new confidentiality legislation for NCES and provide guidance. The public hearing was held to coincide with a meeting of the Advisory Council on Education Statistics (ACES), a group which meets quarterly to advise NCES on the direction of its activities.

Panel I consisted of concerned statisticians; Panel II consisted of Data Users; and Panel III consisted of Data Providers. In order to stimulate discussion, panel members were asked to respond to NCES draft policy papers that described 1) the new law and how it affected NCES, 2) criteria for disclosure avoidance analysis prior to release of public tapes, and 3) possible new initiatives for sharing data.

The views expressed at the conference dealt mainly with the balance between disseminating data and protecting confidentiality. There were few particulars as to how NCES should go about implementing the law. At the close of the session, ACES made a motion to commend the progress NCES had made and to support the basic thrust of NCES' direction. This thrust was that, subject to the necessary confidentiality constraints, NCES' main mission was to maximize the usefulness of its data.

#### B. Affidavits of Nondisclosure

One of the first requirements of PL 100-297 is that persons outside NCES desiring access to NCES individually identifiable information must first swear not to disclose this information. NCES has developed an affidavit of nondisclosure, which can be used to swear in temporary employees of NCES such as contractors and other outside users of NCES confidential data. This information constitutes a system of records and, therefore, has been published in the Federal Register.

#### C. Public Use Tapes

The enactment of PL 100-297 has resulted in delays in the availability of public use data tapes. Education researchers have argued that the tabulations that NCES produces are inadequate for many researchers' needs and that NCES does not have sufficient staff, nor is it NCES' role, to produce all the analyses researchers might want. Researchers require detailed data tapes with individual records. Unfortunately, NCES has never conducted disclosure analyses in the past, and these analyses have required additional time and resources for contractors who were preparing data tapes.

In the last year or two, NCES has initiated a number of new surveys in response to the needs of education policy officials and researchers. Part of PL 100-297 requires the Commissioner of NCES to report regularly to Congress on education issues that many of the new surveys have been designed to address. These surveys are typically more highly linked and provide greater geographic detail than earlier surveys, but this renders much more difficult the task of disclosure analysis and avoidance.

The new School and Staffing Survey (SASS) is a good example. Formerly, the collection of surveys which this survey replaced covered different sectors and levels at different periodicities and provided only national and regional estimates. The new survey is a large linked sample of school districts, schools, principals, and teachers for the public and private sectors selected to produce estimates at the State level.

In addition to the disclosure problems caused by the greater linkage and geographic detail, there is the problem of a multiplicity of national and State files to which sample data can be matched and the uniqueness of many of the records. There are national universes, for example, on institutions of higher education, school districts, and schools and state files on districts, schools, principals, teachers, and, in some cases, students. Of course, the completeness and accuracy of State files varies with the State. There are also files at the district level on schools and principals and even files at the local public library that are sometimes available.

Many records are unique. For example, there is only one school district with an enrollment of 1 million or more, namely, New York City. And when linked, the combination of characteristics can often quickly identify a district/school combination, given the state.

#### D. Disclosure Review Board

The Disclosure Review Board (DRB) was first convened in October 1989. The task of the DRB is to review tapes being prepared for release to the public and the accompanying disclosure analysis documentation in order to determine whether the proposed tape has an acceptably low disclosure risk. The DRB is currently comprised of a representative of the office of the Chief Statistician, the technical planning officers from each division, an outside representative (currently from the Census Bureau), and chaired by the chief of the dissemination division.

The DRB has met at least once a month to consider issues of confidentiality, especially as they relate to the preparation of public release tapes. In its initial deliberations, the DRB discussion focused on learning about the different factors that can affect disclosure and which of them were serious factors at NCES.

NCES faced added complications in preparing public use tapes since PL 100-297 was enacted after several of its large surveys were already in the data collection phase. For example, NCES' use of "unsworn" coordinators to facilitate data collection has made the preparation of public use tapes more difficult.

The DRB takes the following kinds of information into consideration in assessing the risk of disclosure: the existence of "publicly available" files, the age of those files, the correlation of potentially identifying variables between the "publicly available" files and the file proposed for public release, the impact of linkages among files, the level of geography, the use of coordinators (other than NCES employees), public knowledge of the uniqueness of certain records, and the amount of effort and costs involved for someone to "break the code" and thereby disclose individual information from the public use file. While each proposed tape carries with it its own special disclosure problems, the DRB is attempting to refine the techniques used within the Center in order to facilitate disclosure analysis and release. Since the surveys with individually identifiable information are fielded periodically, the length of time to conduct disclosure analyses should diminish over time. The DRB has reviewed or is in the process of reviewing a number of tapes proposed for public release, which are discussed below.

#### E. National Educational Longitudinal Survey

After extensive analysis, NCES has released four linked public use tapes for the National Educational Longitudinal Survey in January 1990. NELS is a linked national sample of schools, students, parents, and the teachers of the students.

Disclosure analysis for NELS consisted of first identifying variables on the proposed public release file common to publicly available universe files. There were two such national universe files available. One was the Quality Education Data (QED) file of schools (private marketing firm). Another was the Common Core of Data file of schools (NCES). The QED had been used as the frame for sample selection. On the basis of "matching" variables, schools on the universe file which

were "similar" to a given school in the sample were identified. When the school on the universe corresponding to a given sample school had a high probability of matching, matching variables were either categorized, collapsed or, on occasion, recoded as missing. Categorization, collapsing, and recoding was also done with a handful of variables on the other (nonschool) files. The overall impact of these changes on the file was minimal.

#### F. School and Staffing Survey

The disclosure analysis for the School and Staffing Survey (SASS) is still in process, probably in its final stages. As mentioned above, SASS was designed to be a state level linked sample of four questionnaires. At the national level, NCES maintains public school and school district universes and QED maintains those universes as well as a universe of private schools. At the state level there exist files of public school principals and teachers as well as school district and school files. SASS is currently considering what level of geography and linkage can maintain confidentiality. The disclosure analysis is looking at a multivariate Euclidean distance measure between each sample combination school district/school record and its corresponding universe "record" and comparing this to the distances of other nearby universe records to see if a combination of matched variables could lead to disclosure.

#### G. National Survey of Postsecondary Faculty

The National Survey of Postsecondary Faculty has been found to include a great deal of data about faculty that is also publicly available in other sets of files and records. In addition, the survey used coordinators at the institution level to facilitate data collection. The combination of these two factors renders it quite simple to disclose the identity of sample individuals.

Since this level of descriptive information is believed necessary for most analyses, the full data tape was not releasable. However, NCES is addressing this public use data problem in two ways. First, NCES has developed a public use encrypted data file and software that will enable users to make crosstabulations of the data while protecting the confidentiality of sample individuals. Second, NCES is preparing numerous tabulations to try to meet the needs of data users and will also respond to special requests for tabulations.

## V. Sharing Data Outside NCES

NCES recognizes that the public use tapes will not satisfy all the needs of the education community. Therefore, a number of other options are being pursued. Some of the options are ones that were initially raised in the NCES draft, which was discussed at the Conference on Confidentiality. Each of these options will require users to swear to an affidavit of nondisclosure and to justify their proposed use of the data and why the public use tape would not be adequate.

### A. User Centers

NCES is planning to establish locations at which sworn users could access data tapes with individually identifiable information in a monitored environment. Some possible organizations that might agree to perform such a function include NCES contractors, Office of Educational Research and Improvement Labs and Centers (especially, the new Data Dissemination Center), and certain licensees.

### B. Licensees

NCES has developed a draft license agreement. This draft is being sent to ED's Office of General Counsel for review. Under this agreement, organizations or agencies that desire to use NCES confidential data in accordance with the statistical or research purposes for which it was collected and wish to do so at their own sites can enter into a license agreement with NCES. Only a limited number of sworn staff of the licensee would have access to the data. The licensee would be responsible for the security of the data tapes and all products. NCES would review each license application to see that the confidentiality of data is reasonably protected.

### C. Dial-In Access

In the future, NCES will explore the feasibility of dial-in access from remote terminals that are not at a licensed site or at a user center. This option would be limited to sworn persons who have a valid statistical purpose for which the public use tapes would not be sufficient.

## VI. Commissioned Papers to Generate Other Data Sharing Approaches

In NCES is striving to be responsive to the education community's need for timely survey

data in yet another way by supporting further research on the topic of confidentiality. NCES has commissioned four papers by researchers and legal scholars to identify other approaches, options and mechanisms to maximize data collection, data estimation, and dissemination while minimizing the likelihood of disclosure of individually identifiable information.

The purpose of these commissioned papers is to identify new approaches and procedures to increase researchers' access to individually identifiable data within the legal, practical, and ethical constraints imposed by law and NCES standards. Furthermore, these papers will be used to generate issues and approaches to be addressed by a proposed "Confidentiality Panel of the National Academy of Science" scheduled for September 1990. It is anticipated that the confidentiality panel will focus mainly on ethical, practical, and legal constraints that limit the distribution of individually identifiable information from NCES surveys, and propose approaches that will maximize access to this information within these constraints.

## VII. Summary

NCES has been exploring ways to maximize the utility of its data subject to the necessary confidentiality constraints. In its implementation, NCES has been guided by the experiences of many other agencies and organizations that have been struggling with the problems of confidentiality for some time.

Public files are undergoing disclosure analysis and modifications prior to release; some files are being encrypted. Modifications include decreasing the geographic detail or linkage to other files as well as the more traditional categorization, random reassignment, and (at times) suppression.

NCES is planning to make files available to sworn researchers at monitored user centers and to organizations through licensing agreements and is exploring the feasibility of dial-in access. NCES is also exploring other possibilities through a set of commissioned papers and a confidentiality panel and hopes to be able to devise ways to meet the needs of researchers and policy officials for timely educational data while protecting the confidentiality of respondents.

\*This paper is intended to promote the exchange of ideas among researchers and policy makers. The views are those of the author, and no official support by the U.S. Department of Education is intended or should be inferred.