

**STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

**COMMISSION OF THE
EUROPEAN COMMUNITIES**

CONFERENCE OF EUROPEAN STATISTICIANS

EUROSTAT

**Joint ECE/Eurostat Work Session on
Statistical Data Confidentiality**
(Thessaloniki, Greece, 8-10 March 1999)

Working Paper No. 6 (Summary)
English only

Topic (i): new applications of disclosure control methods

MAKING DATA MORE ACCESSIBLE IN A CLIMATE WHERE PERCEPTION MATTERS

Submitted by the U.S. Bureau of the Census¹

Contributed paper

Summary

1. The U.S. Census Bureau is planning linkages of data on individuals with data on businesses to study the relationship of workers and employers within communities. At the same time, the Bureau is developing new ways to provide access to data from these linkages and from the population census to be taken in the year 2000. Access is planned through different mechanisms based on the sensitivity of the data and the needs of users. For the linkage project, data will be available to researchers primarily through Research Data Centers where the Bureau can better assure control over access and use of the confidential data. Access to the Census 2000 data will be provided through the electronic Data Access and Dissemination System (DADS) which will provide Internet access to pre-defined and custom tabulations. There are real disclosure issues arising from these activities that are being addressed by security and disclosure limitation methods. Nevertheless, issues remain about unknown threats from intruders who have new tools to break our protections and concerns that the public may perceive that we have not done enough. This paper discusses access plans for DADS and the new Linked Household-Employer Dynamics Protect (LHEDP) and identifies some obstacles and possible solutions.

2. For DADS, broad access to Census 2000 data is planned over the Internet. Users will have the potential to access data in any one of three levels: press releases/reports, predefined tables, and user-defined tables. In planning for predefined tables, the Census Bureau has chosen data swapping as a primary method to protect data, as it did in 1990. This method permits showing values for very small cells (ones or twos) when displaying data for fine levels of geography such as blocks or block groups. While the method proved highly desirable in 1990, new issues have arisen about the degree of swapping required and the potential for greater public concern that personal information is being disclosed. Options are being considered to add more swapping or to include controlled or random rounding to limit the perception of disclosure. For user defined tables, the Census Bureau is prepared to implement filters and firewalls to screen requests for tabulations. Tabulations that would reveal, directly or indirectly, characteristics that have not been swapped would be denied. Final plans for DADS disclosure methods are to be completed by the end of 1998. Among the disclosure issues to be considered are the broad availability of data to unsophisticated users, the inexpensive powerful new tools to link and mine data files, and the public's perception that data may not be adequately protected.

3. For the LHEDP, access will be more restrictive than for Census 2000 and there are no plans to provide these data through DADS. This project raises its own sets of issues, both from the disclosure viewpoint and on the basis of

¹ Prepared by Gerald Gates.

privacy concerns. Initial plans call for linking data from the Census Bureau's Standard Statistical Establishment List (SSEL), the Social Security Administration's earnings history file, both the Current Population Survey and the Survey of Income and Program Participation, and eventually, the American Community Survey. Data will be linked using the employer's Employer Identification Number (EIN) and the employee's Social Security Number (SSN). Since EIN and SSN are not available on all files, some will be obtained indirectly using name and address from the source files. The privacy issues raised by these linkages are of real concern and the Census Bureau is evaluating the public's expectations regarding controls over access and use. The research-valuable data sets that are created will be used by Census Bureau researchers and specially-sworn outside researchers strictly at the Census Bureau's secure Research Data Centers. Unique security arrangements will be put in place to ensure that the data are protected and that uses are restricted.