

Working Paper No. 6 (Summary)

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

Joint ECE/Eurostat work session on statistical data confidentiality
(Luxembourg, 7-9 April 2003)

Topic (ii): New data release techniques

ACCESSING MICRODATA VIA THE INTERNET

Invited paper

Submitted by the Central Bureau of Statistics (Israel)¹

¹ Prepared by Natalie Shlomo (natalies@cbs.gov.il).

Accessing Microdata via the Internet

Natalie Shlomo
Director, Census Methodology Sector
Central Bureau of Statistics
Israel
natalies@cbs.gov.il

Some of the microdata files of the Israel Central Bureau of Statistics are in the process of being placed on the internet to provide remote access to users. This is being implemented in two parallel systems:

1. A multi-facet system is being developed at the Bureau that allows users to build custom-made tables and/or download files, depending on the level of security of the file. With this new application, a decision has to be made as to what type of file should be in the background in order to produce the tables and whether downloading all or parts of the file will be allowed. A public use file (PUF) from which tables can be generated will a priori produce tables that are safe and there is no need for any further restriction or licensing. However, the data generated under such restrictions might have insufficient detail to have real value to researchers. Another possibility is to have the original file in the background of the system from which tables can be generated, but this means placing strict restrictions on the tables themselves, such as a minimum number of cases to a cell, a maximum number of dimensions to the table, and prior registration of the users. The different approaches will be examined by weighing the disclosure risk against the information loss for a particular application.
2. The Israel CBS also has an agreement to place microdata files in a depository managed by the Hebrew University of Jerusalem. Some of the Bureau's PUF microdata files have been put on an internet website of the depository which can be accessed by registered researchers at universities and research institutions of Israel for producing custom-made tables and carrying out data analysis. There are currently no restrictions on the tables and files can be downloaded as well.

With the implementation of these new internet systems and because of the current lack of a systematic way of evaluating the risks of the files, the Bureau recognized the need to develop methodologies to reassess the level of security of the microdata files that it releases. The results of this work will be presented in the paper. Real microdata sets that have been released as PUF files were evaluated by several methods. One method was to estimate the level of risk based on risk measures that are estimated from the sample. In addition, files were also linked to the National Population Register to simulate resources which could be in the possession of potential attackers and the risks of the files were obtained based on the probability of obtaining correct matches for the sample uniques. These results are compared to the estimated measures.

Based on the results of this reassessment, the Bureau will be able to provide better tools for defining the levels of security of microdata according to its intended usage. The ultimate goals of the research will be to develop methodologies for evaluating the risks and categorizing the files according to risk measures; to determine the thresholds and accepted levels of risk; and to develop methods of disclosure control that best suit the growing needs and policies of the Bureau. In addition, there is an immediate need to take disclosure control action during the interim period, in particular for the files that will be placed on the internet.

Keywords : internet access, custom-built tables, risk assessment, disclosure risk measures