

Working Paper No. 39 (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 (vi): Software tools for statistical disclosure control

**BUREAU OF TRANSPORTATION STATISTICS' PROTOTYPE  
DISCLOSURE LIMITATION SOFTWARE FOR COMPLEX TABULAR DATA**

**Contributed paper**

Submitted by the Department of Transportation, United States<sup>1</sup>

---

<sup>1</sup> Prepared by J. Neil Russell (neil.russell@bts.gov).

## **Abstract**

### **Bureau of Transportation Statistics' Prototype Disclosure Limitation Software for Complex Tabular Data**

The United States Department of Transportation's Bureau of Transportation Statistics (BTS) is updating its confidentiality policy and methods to take advantage of the latest advancements in statistical disclosure limitation (SDL) research. BTS is required by law to protect the identity of data providers. And in reviewing current SDL methods, BTS was dissatisfied with complementary cell suppression and how it eliminates significant amounts of data in tables. Hence, BTS initiated a project to develop and implement a new, state-of-the-art SDL method for multi-dimensional (up to five) hierarchical tables. The goal of the project was to increase the amount of information released (compared to cell suppression) while protecting the identity of the data providers.

After reviewing a wide variety of SDL methods for tabular data in the literature, the project team selected the Synthetic Data Substitution (SDS) method proposed by Dandekar and Cox (2002). This SDL method evaluated well for BTS requirements and goals. The SDS method was subsequently enhanced to efficiently process complex large tables, and was designed for use in a static system. (A static system is where the only tables published are those chosen by the agency, and the microdata that created the tables would not be released.) A modified version of the SDS method was then implemented into prototype computer software for testing and demonstration.

Briefly discussed is the research process and development of an efficient algorithm based on the SDS method and Tabu Search that allows for processing of multi-dimensional tables with hundreds of thousands of entries. The software user interface, software functions and options are described, and examples of processed tables using government agency tabular data are provided. Early indications of software testing has shown user satisfaction with the software interface, ease of specifying protection rules and microdata processing, and very short processing times. It was also found that the modified SDS method also preserved the statistical properties of the tabular data when original tables are compared to modified (protected) tables. The paper concludes with suggestions for future software redesigns and research.

Mr. J. Neil Russell  
U.S. Department of Transportation  
Bureau of Transportation Statistics  
400 7th St. SW, Room 4432  
Washington, DC 20590  
USA  
(202) 493-2147  
neil.russell@bts.gov