

Distr.
GENERAL

CES/SEM.47/9 (Summary)
22 January 2002

Original: ENGLISH

**STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

**COMMISSION OF THE
EUROPEAN COMMUNITIES**

CONFERENCE OF EUROPEAN STATISTICIANS

EUROSTAT

**Joint UNECE/Eurostat Seminar on Integrated Statistical
Information Systems and Related Matters (ISIS 2002)**
(17-19 April 2002, Geneva, Switzerland)

Topic II: Secure communications and data confidentiality

THE CASC-PROJECT

Invited paper

Submitted by Statistics Netherlands ¹

Summary

Introduction

1. Statistical disclosure control (SDC) is a field in statistics that has attracted much attention in recent years. Decision-makers demand increasingly detailed statistical information. Researchers have the capacity to perform complex statistical analysis on their powerful PCs and they desire detailed microdata. There is, therefore, growing pressure on the statistical offices to publish increasingly detailed information. But the Statistical Institutes have to preserve the balance between their task as a data provider and their obligation to preserve the privacy of the respondents, who have entrusted their individual information to them. Without respondents there is no statistical information.

2. The CASC-project is an initiative to coordinate research and development in Europe. It is partly subsidized by the 5th Framework programme of the European Union. As a follow-up of the SDC-project it aims at the combination of research and the development of practical tools, the ARGUS-software. We aim both at the SDC-problems for microdata as well as tabular data.

Microdata

3. The existing ?-ARGUS has already useful capacities for the protection of social microdata, based on global recoding and local suppression. However for enterprise microdata, the disclosure protection is much more difficult. Due to the special (skewed) distributions in these datasets the traditional methods are no longer adequate. Research initiatives are:

¹ Prepared by Anco Hundepool (ahnl@krypton.vb.cbs.nl).

- ?? Noise addition;
- ?? Post-Randomization;
- ?? Micro-aggregation.

4. These techniques will be investigated and implemented in new versions of ?-ARGUS. This will offer the opportunity to easily compare the results of the different methods. It will for the first time open the possibility of making safe enterprise microdata. Nevertheless, we should be very careful and also consider the 'safe settings', a controlled analysis centre within the NSIs.

5. In addition, we will study the risk-models that can be used to compute the safety of individual records and a dataset.

Tabular data

6. The current version of ?-ARGUS has solved the disclosure protection of non-structured statistical tables, based on optimization techniques. However experience has shown that many real-life tables have more complex structures. Most code lists have a hierarchical structure, which cause big problems for the protection of these tables. Many sub-marginals as a table could help to undo the cell-suppression.

7. So several new approaches are being investigated to solve this problem. Specialists are working on the extension of the current optimization techniques for hierarchical tables. We will include this solution in new versions of ?-ARGUS, but as we expect that due to the complexity of this problem, the computing time for the protection of very large tables might be very big, we are including alternatives. Solutions based on a clever partitioning of the table are included as well as a solution based on hypercubes. This will lead to a software tool that can offer a range of solutions from a slow, optimal solution to a quick, but not fully optimal solution. As a side effect, the solutions for hierarchical tables can also be applied to solve the problem of the linked tables.

Testing

8. Much attention will be paid to the testing of software and methods. Many nice techniques can be implemented, but in the end the resulting data must be useful. So questions such as the 'analytical power' of a dataset are being investigated. Also, the applicability of the resulting software will be tested in real-life applications.

Conclusion

9. We hope to make a major step forward in both the research as the practical application of Statistical Disclosure Control methods. Further information on the project is available on this web site:
<http://neon.vb.cbs.nl/casc/default.htm>.