

Working Paper No. 20 (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

## **THE ARGUS-SOFTWARE**

**Invited paper**

Submitted by Statistics Netherlands<sup>1</sup>

---

<sup>1</sup> Prepared by Anco Hundepool (ahnl@rnd.vb.cbs.nl).

# The ARGUS-software

Anco Hundepool  
Statistics Netherlands  
Methods and Informatics Department  
P.O. Box 4000  
2270 JM Voorburg, The Netherlands  
E-mail: [ahnl@rnd.vb.cbs.nl](mailto:ahnl@rnd.vb.cbs.nl)

## Introduction

Statistical Disclosure Control is a field in statistics that has attracted much attention in recent years. Decision-makers demand more and more detailed statistical information. And researchers have the capacity to perform complex statistical analysis on their powerful PCs and they desire detailed microdata. Therefore there is a growing pressure on the statistical offices to publish more and more 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 trusted their individual information to them. Without respondents no statistical information.

The CASC-project is an initiative to coordinate the research and development in Europe. It is partly subsidised by the 5<sup>th</sup> Framework program of the EU. 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

The existing  $\mu$ -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 harder. Due to the special (skewed) distributions in these datasets the traditional methods are no longer adequate. Research initiatives are:

- Noise addition.
- Post-Randomisation
- Micro-aggregation

These techniques will be investigated and also be implemented in new versions of  $\mu$ -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 the 'safe settings' should be considered, a controlled analysis centre within the NSI's.

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

## Tabular data

The current version of  $\tau$ -ARGUS has solved the disclosure protection of non-structured statistical tables, based on optimisation techniques. However the experiences have shown that many real-life tables have more complex structures. Most codelist have a hierarchical structure, which cause big problems for the protection of these tables. Many sub-marginals is a table could help to undo the cell-suppression.

So several new approaches are being investigated to solve this problem. Specialists are working on the extension of the current optimisation techniques for hierarchical tables. We will include this solution in new versions of  $\tau$ -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.

### **Conclusion**

We hope to make a major step forward in both the research as the practical application of Statistical Disclosure Control methods. In this paper we will focus on the software aspects of this project. I.e. the ARGUS software.