

**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. 8  
English only

Topic (ii): software and computing developments

**BETTER SOLUTIONS TO DISCLOSURE CONTROL PROBLEMS: RESEARCH  
UNDER THE FIFTH FRAMEWORK PROGRAMME**

Submitted by Eurostat<sup>1</sup>

**Invited paper**

**I. The Fifth Framework Programme**

1. As part of the European Commission's Fifth Framework Programme, there will be significant amounts of funding available to sponsor research into a number of areas. A key aim of the programme is work towards a 'user friendly information society'. Statistics are seen as being a key part of the information society, and research in statistics therefore forms a significant topic in the planned research.

2. Although this paper will concentrate on research in the field of statistical confidentiality and disclosure control, the statistical research programme attached to the Fifth Framework Programme will address all aspects of the production of official statistics.

3. There are various challenges faced in the development of the European Statistical System – these challenges are in the main common to both Eurostat and the National Statistical Institutes (NSIs), as well as to many other individuals and organisations working within the information society. There is pressure for more information to be made available, with increasing (and often conflicting) requirements regarding timeliness, quality and cost. There are also concerns about response burden and maximising the use of available data.

4. Developments in information technology have lead to an increased demand for more flexible 'non-standard' statistical outputs which are more amenable to further processing and analysis. These may include microdata, detailed user defined tables and geographic information systems related products. There are also on-going changes in the methods by which statistics are delivered to the user. There is less emphasis on the production of volumes of printed tables. Instead, statistics are now likely to be supplied in electronic form via e-mail and the Internet. Such changes look set to continue.

---

<sup>1</sup> Prepared by David Thorogood.

5. It will be clear that the changes and conflicting demands outlined above imply greater problems for statisticians who must be concerned with maintaining the confidentiality of statistical data. Additionally, the greater availability of powerful computing hardware and record matching software means that even data which it was safe to release previously may now need to be considered unsafe.

6. This paper aims to highlight a few of the areas which might be considered for research funding under the Fifth Framework Programme. This should not be considered to be a definitive list, and any applications to participate in Fifth Framework Programme work should be based on the official Programme documentation. The European Commission does not commit itself to fund research on any of the specific areas mentioned here.

## II. Experience from the Fourth Framework Programme

7. Research in a number of areas relating to statistics and information systems was funded under the previous Fourth Framework Programme. In relation to statistical confidentiality, the most significant piece of work has been the Statistical Disclosure Control (SDC) project. This project, which was led and co-ordinated by Leon Willenborg at Statistics Netherlands, has encompassed work on a number of issues relating to the theory and methodology of disclosure control, as well as making significant developments to the ARGUS disclosure control software.

8. The project included the work of NSIs, universities and research institutes in the Netherlands, Italy and the UK. Each organisation has contributed its own skills and particular expertise to this wide-ranging and productive project.

9. The software development has involved the production of two separate versions of ARGUS have been produced:  $\mu$ -ARGUS which works with microdata files and  $\tau$ -ARGUS which addresses disclosure control problems of tabular data. Both pieces of ARGUS software now include an easy to use Windows-type interface. The software includes tests for disclosure risks in data, and offers readily useable methods - such as recoding or record suppression - which reduce that level of risk. Users are able to look at the effects - in terms of degree of protection and information loss - of implementing particular disclosure control solutions.

10. It is important to remember however that, although such software is a powerful tool to protect data, there is scope for considerable theoretical and methodological research to guide the correct use of this tool. Although ARGUS is perhaps the most visible outcome of the Statistical Disclosure Control project, there has been much useful work on, amongst other things;

- efficient algorithms for disclosure control;
- disclosure risks of producing outputs for multiple geographies;
- opportunities and motives for attempting to disclose information.

11. There are numerous references which give a more detailed description of ARGUS and of the SDC project more generally (see for example Willenborg and Hundepool, 1998; Elliot *et al.*, 1998, Benedetti *et al.*, 1998).

### **III. Future research under the Fifth Framework Programme**

#### *Likely research areas*

12. There are a number of statistical confidentiality issues which appear suitable for further research under the Fifth Framework Programme. Despite the significant developments of ARGUS, and of other similar software, there is scope for further software work. Areas of particular interest include improving the ability of software to deal with complex microdata files, tables which have a complex and/or hierarchical structure, and linked tables.

13. There is also a need to work on the evaluation of new and existing disclosure control methods and software tools. It is necessary to look at the degree of protection conferred on data and the amount of information which is lost. There may be scope to develop standard testing procedures and datasets that will help ensure the comparability of test results. It would also be useful to develop the use of disclosure risk models to help identify cells which pose the greatest disclosure risk and to guide the use of disclosure control measures.

14. Other possible areas for software and methodological research include:

- further methodological research into improved and refined disclosure control techniques;
- examination of ways of allowing safe access and processing of confidential data held on distributed databases;
- ways of measuring and addressing the disclosure risks which result from multiple requests for similar data. This could include research into the risks of differencing which can result from data being supplied for flexibly defined geographical areas or classifications.

15. For any disclosure control methods and software to be correctly implemented, it is important to have better information about respondent perceptions about confidentiality. There is also a need for cognitive research to address:

- the extent to which respondents understand and believe the confidentiality promises made by statistical institutes;
- which variables are seen as sensitive and which therefore require the greatest protection
- the extent to which data ageing is felt to make certain types of data less confidential.

### **IV. Bidding to participate in the Fifth Framework Programme research**

16. It is intended that there will be several calls for proposals to undertake work under the Fifth Framework Programme. The first of these calls is expected to be in March 1999. Proposals will be assessed according to published criteria relating to scientific excellence and relevance to the research programme. At the time of writing, rules for the formation of research consortia and for funding options have yet to be finalised. However, funding is likely to be on a shared-cost basis, and there is likely to be a requirement that consortia consist of two or more non-affiliated organisations from two or more EU Member States.

17. Organisations which are considering participating in this programme are invited to consult the programme Internet site at <http://europa.eu.int/en/comm/eurostat/research/index.htm>

(Specific questions may be addressed to Jean-Louis Mercy at Eurostat (jean-louis.mercy@eurostat.cec.be). Although it is necessary to bid as part of a consortium, Eurostat will attempt to put potential individual participants in touch with each other.

18. Ideally consortia should also be multi-disciplinary in nature. It is envisaged that participants may include individuals and organisations such as:

- statisticians in government and academic institutions;
- information systems specialists;
- experts from other relevant fields – economists, psychologists etc.;
- software development companies; and
- data providers and users more generally.

## **VI. Conclusions**

19. Although certain details remain to be finalised at the time of writing, it is clear that the Fifth Framework Programme will include significant amounts of funding for research in statistics. This will cover many topics relating to statistical confidentiality, as well as to statistics more generally.

20. Potential participants in this work are invited to contact Eurostat to obtain further information about the research opportunities and bidding procedures.

## **References**

Benedetti, R., Franconi, L. and Piersimoni, F. , Per-record risk of disclosure dependent data. *Proceedings of the Statistical Data Protection'98, Lisbon, 25-27 March 1998.*

Elliot M. *et al.* Special Uniques, Random Uniques and Sticky Populations: Some counter intuitive effects of geographical detail on disclosure risk. *Proceedings of the Statistical Data Protection'98, Lisbon, 25-27 March 1998.*

Willenborg, L.C. R. J. and Hundepool, A. J., ARGUS for Statistical Disclosure Control. *Proceedings of the Statistical Data Protection'98, Lisbon, 25-27 March 1998.*