

**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. 16
English only

Topic (iii) - Administration and policy of statistical data confidentiality

STATISTICAL CONFIDENTIALITY AT THE EUROPEAN LEVEL

Submitted by Eurostat¹

Invited paper

I. Introduction

1. Increasingly, National Statistical Institutes (NSIs) of Member States of the European Union (and of European Free Trade Area countries in many cases) are working closely with Eurostat in the design and co-ordination of statistical systems relating to a wide variety of topics. The concentration on industrial and agricultural statistics in the earlier years of the European Communities has broadened to include statistics on the full range of economic, social and environmental topics. Many official statisticians in Europe will therefore be familiar with some of the functions and procedures of Eurostat. Similarly, many academic researchers will have attended scientific conferences sponsored by Eurostat or will have participated in Eurostat sponsored research.

2. Eurostat aims to work in partnership with the NSIs. It is therefore vital that a good working relationship is maintained, based on trust and mutual co-operation between Eurostat and the NSIs. The NSIs need to be sure that Eurostat will adequately protect their confidential data. Eurostat therefore works closely with the NSIs and other international institutions to further research and foster good working practices in the field of statistical confidentiality. Regular discussions take place with the NSIs to ensure that the production and dissemination of European level statistics is undertaken without endangering confidentiality.

3. Although Eurostat performs many of the functions that would normally be associated with a NSI, there are important differences in role which impact of the treatment of statistical confidentiality. Eurostat is not itself involved in the data collection process, being rather a processor and disseminator of data collected in the main by NSIs and related agencies. Although removed from the data collection process, Eurostat is a trusted receiver of data from NSIs, and as such must uphold the confidentiality promises made by the NSIs. This is required by European legislation which is discussed in the next section. The confidentiality of statistical data is central to the concerns of any statistical institute. An institution which is passing confidential data to another will want to ensure that the same high levels of

¹ Prepared by David Thorogood.

protection are in place in the receiving institution. Both institutions will want to ensure that the transfer of data is allowed by law, and that policies are in place to ensure that these laws are correctly implemented.

II. Legislation

4. In the case of many official statistical studies, individuals or enterprises selected to participate in the study are legally obliged to respond. In other cases where statistical information is obtained from administrative sources, there is no legal right to object to information being additionally used to produce statistics. This legal obligation to supply data is justifiable because of the provision of reliable statistics is in the wider public interest. Such an obligation to respond must however be balanced with a legal protection of the privacy and confidentiality of respondents. It is normal for both aspects to be dealt with together by national law. However, as confidential statistical information is transferred also to Eurostat, similar protection is required under European law. This section briefly outlines two main pieces of European legislation dealing with statistics in general. There are of course other items of European legislation which deal with the provision of statistics on specific topics.

5. Although the issue of respondent burden is less frequently covered by legislation, it must be recognised that responding to statistical enquiries can be time consuming and costly. There is a duty on all statistical institutes to ensure that data are not collected in excess of what is required, and that the best possible use is made of the data which are collected. This has legislative and policy implications as there may be pressures for data to be used in ways which were not envisaged at the time when confidentiality policies were set in place, or pressures to allow an increased range of persons to be granted access to confidential data. It is important that such problems and conflicting pressures are dealt in innovative and imaginative ways, which will try to maximise data use and contain respondent burden while ensuring that the protection of respondent confidentiality continues to be treated as an issue of prime importance.

6. A distinction is frequently drawn at the European level between statistical information which must be supplied by Member States to Eurostat - that which is referred to in topic-specific pieces of legislation - and statistical information which is supplied to Eurostat on a voluntary basis. It will be clear however that the same high standards to confidentiality protection will need to be applied to all statistical information supplied to Eurostat.

- Council Regulation 1588/90

7. This piece of legislation formalised the role of Eurostat as a trusted party in terms of the transmission of confidential data. Previously, certain Member States had been restricted by national legislation from passing confidential data to Eurostat. Under Regulation 1588/90, reasons of statistical confidentiality may not be used to prevent transmission of data to Eurostat. Following the established principle separating the administrative use of data from the statistical use of data, although confidential data are passed to Eurostat, access may not be allowed to other parts of the European Commission.

8. Regulation 1588/90 does not force NSIs to transmit to Eurostat data which could directly identify respondents. Therefore, it is normal for details such as names and addresses to be removed by the NSI before data are sent to Eurostat.

9. National civil servants and other statistical office staff who breach rules protecting confidential data are generally liable to punishment under specific national legislation. As confidential data are placed in the custody of Eurostat civil servants, it follows that there needs to be similar legislation which applies to Eurostat staff. European legislation is generally enforceable via national legislation – national law in the Member States being changed to fit the requirements of European law. Therefore

article 6 of Regulation 1588/90 requires that Member States of the European Union should introduce national legislation to prevent Eurostat officials and contractors from disclosing confidential statistical information.

10. Regulation 1588/90 also established the Committee on Statistical Confidentiality which serves as a forum for discussions between Eurostat and the Member States regarding statistical confidentiality. Proposals regarding the treatment of confidential data are submitted to this Committee for its opinion. The Regulation makes provision for decisions to be reached by a majority vote. However, the aim of Eurostat would be for any such decisions to be reached by general agreement rather than by a majority vote.

- Council Regulation 322/97 - the ' Statistical Law'

11. More recently, the Council Regulation on Community Statistics (Regulation 322/97 - frequently referred to as the 'Statistical Law') increases the role of Eurostat in a number of areas relating to statistical confidentiality. Whereas under previous legislation Eurostat had been allowed to receive confidential data from the Member States, the decision about whether the data were confidential remained with the Member State concerned. Differences in national rules might mean that data collected in one country were seen as safe for release, whereas identical data collected in another country had to remain confidential. The Statistical Law envisages the agreement of common EU-wide definitions of confidentiality based on the identifiability of statistical units. It is no longer seen as being within the competence of the individual Member State to decide what is confidential.

12. The Statistical Law does not provide an absolute definition of what constitutes an identifiable statistical unit. However, article 13 of the Statistical Law states that *'To determine whether a statistical unit is identifiable, account shall be taken of all the means that might reasonably be used by a third party to identify the said statistical unit'*.

13. Thus the concept of reasonableness is applied to the protection of confidentiality. This recognises the fact that, although all reasonable measures must be taken to prevent disclosure, absolute protection of data would prevent the production of virtually all outputs. For example, proof that limited disclosure might be theoretically possible given the expenditure of huge resources and the possession of extensive background knowledge would not necessarily prevent the release of data. Eurostat and the NSIs must work together to reach reasonable and workable definitions of confidentiality.

14. The Statistical Law also allows for persons such as academic researchers to be granted access for scientific research to confidential data held by Eurostat. However, such access can only be granted with the consent of the Member States. Such access, if agreed, would take place in tightly controlled circumstances which would ensure the continued protection of the data. The possible creation of a Data Analysis Centre as a safe setting in which confidential data may be accessed is discussed in section 4. It is hoped that in the future agreement can be reached on ways of allowing this additional use of data.

III. Policies in operation

15. Policies and working procedures have been put in place at Eurostat as part of the implementation of European statistical legislation. These procedures, which cover details of the transmission, storage, processing and dissemination of data, have been agreed with Member States through the Committee on Statistical Confidentiality. Where appropriate, more detailed discussions have also taken place at topic-specific working groups and task forces.

16. Regulation 1588/90 stipulated that all necessary regulatory, technical, administrative and organisational measures must be put in place to ensure the confidentiality of data transmitted to Eurostat. As part of this, a manual on the protection of confidential data at Eurostat was prepared. This sets out a series of principles and procedures which are to be followed in dealing with confidential data. The principles referred to in the manual include:

- The protection of confidentiality is of central importance, over-riding considerations of ease of access and processing;
- Adequate levels of protection must be provided, in line with agreed standards, for all stages of data handling;
- Access to confidential data must be limited to those staff who have a clear need to work with the data;
- For each dataset there is a named official who has primary responsibility for the protection of confidentiality.

17. Final responsibility for the protection of confidentiality rests with the Director General of Eurostat. Others have specific administration and advisory functions on related issues such as legal matters, information systems security, and statistical methodology.

18. The manual makes specific recommendations regarding the physical and logical security of confidential data. A secure area has been created on Eurostat premises where confidential data can be processed and stored.

19. As part of the implementation of the Statistical Law, task forces of experts from the Member States and Eurostat have addressed a number of issues relating to the treatment of confidential data. The work of these task forces is continuing, and will help to ensure that, where necessary, procedures first developed under Regulation 1588/90 can be updated in line with the Statistical Law.

IV. Looking ahead - current work and future developments

20. Technical or policy work on statistical confidentiality takes place against a background of developments in information technology and statistical methodology. Official statistics are produced in an environment where there are changing political and societal demands and perceptions which must also be considered. It is possibly then a mistake to consider that one is working towards a permanent solution to all problems of confidentiality. Instead it is necessary for work to be ongoing and to attempt to keep pace with changes in the wider world.

- Improving communication

21. The importance of Eurostat working closely with the NSIs has been mentioned already. Efforts are being made to improve communication and consultation on confidentiality issues between Eurostat and the NSIs. Although the Committee on Statistical Confidentiality must remain as the formal venue for discussions on confidentiality, there is a place for more regular communication using e-mail and the Internet. To this end, a web site is being established to which members of the Confidentiality Committee will have access via a password. It is intended that papers and items of interest to members can be placed on this restricted access site, with comments and opinions returned by e-mail. Rather than taking the place of formal meetings, it is hoped that this development will make the Committee

meetings more productive. Policy documents presented at meetings can take into account the views of Committee members on earlier draft versions.

- Defining what is confidential

22. Various national rules are in place defining what data are to be considered confidential. Although these are commonly based on established tests of confidentiality such as thresholds or dominance rules, there are differences in the critical figures used. As mentioned in section 2, as part of the implementation of the Statistical Law, there is a need to develop common confidentiality rules. For several datasets - such as PRODCOM and Structural Business Statistics data - discussions with Member States have led to specific charters which define when data should be considered to be confidential. In the case of PRODCOM, a charter was put in place to cover data collected before the Statistical Law came into force. This has recently been revised to cover data collected after the Statistical Law. Given the great variety of statistical outputs produced by Eurostat, it will be difficult or impossible to compile detailed general rules applicable to every dataset and type of output. A more likely outcome is a set of general rules and principles, which may be applied to all outputs for which a specific charter has yet to be agreed. A task force on the methodology of confidentiality has begun work on this.

- Producing microdata outputs

23. There is an increasing demand from the users of statistics for microdata outputs. These allow a far wider range of analyses to be undertaken. Many NSIs have experience of releasing microdata, and it is clear that the particular confidentiality risks associated with this form of output can be successfully overcome.

24. Within Eurostat, there has been work towards producing microdata files based on the European Community Household Panel (ECHP) survey. This survey, which addresses a range of socio-economic issues, involves annual interviews with a representative sample of households and individuals in each country - an annual sample of approximately 60 thousand households and 130 thousand adult individuals.

25. The longitudinal nature of data from a panel survey such as the ECHP can lead to additional disclosure risks. Changes in the structure or circumstances of a household which would be recorded in the ECHP data - such as the birth of a child, a migration, or a change in employment - are readily visible and at risk of identification. It was necessary therefore to create a user database to which disclosure control measures had been applied.

26. The fact that the data come from a small sample of the population already provides a good degree of protection to the data. In producing the user database, changes are made to reduce the amount of identifying information which is contained on each record. These modifications, which are of the type frequently made by statistical institutes when releasing microdata, include

- limiting the geographical information ;
- reducing the classification detail or limiting the upper codes for certain variables;
- removing sampling information which might be used to define geographical details.

27. It is still felt necessary to retain some degree of control over what is done with the released microdata. Data recipients may not distribute the data to third parties, and they must not attempt to identify (or claim to have done so) any respondent. Copies of any papers or documents produced using the microdata have to be supplied to Eurostat. The release of microdata from the ECHP, and the decisions on the modifications which were made to the user database for disclosure avoidance were

agreed following detailed discussions with the NSIs. As national data from the panel survey may be released by the relevant national agencies, it is important that Eurostat continues to work closely with these national agencies to ensure that the release conditions are complied with.

- Increasing the use made of Eurostat data

28. The vast amount of data held by Eurostat represents a unique information resource on the economic and social situation of all Member States of the European Union. These data have been collected at public expense, often with a considerable burden being placed on respondents. It is clearly in the interests of governments, citizens and enterprises that these data are fully exploited. The extent to which the data held by Eurostat can be analysed is limited by the finite number of expert analytical and research staff available within the organisation. It is desirable therefore to increase the use which can be made of Eurostat data by external researchers. Ways are therefore needed of simplifying and accelerating the procedures to allow this greater use, without of course reducing the degree of protection for confidential data.

29. It has been proposed that a Data Analysis Centre (DAC) be established on Eurostat premises where external researchers can be permitted access to confidential data under contractual agreements which cover the maintenance of confidentiality, and which place strict controls on the uses to which the data can be put. The DAC can be seen as a 'safe setting' in which confidential data can be analysed

30. The DAC itself would consist of a secure hermetic working and data storage environment in which the confidentiality of the data for research can be ensured. Both the physical and the IT aspects of security would be considered here. The DAC would also include administrative and support facilities to co-ordinate responses to requests for data access, and to ensure that the agreed conditions for access to the data were complied with. Researchers from NSIs, international bodies such as OECD or the European Monetary Institute, universities and similar institutions would be invited to submit applications to undertake analyses of data held by Eurostat.

31. Although there was general support for the aims of the DAC, certain Member States expressed reservations about granting standing permission for the operation of the DAC. It has been agreed to look at ways of proceeding with this option on a 'case by case' basis with permission being sought from Member States for particular pieces of research, to allow a better understanding of opportunities and problems of such a facility.

32. Various sanctions and pressures may be exploited to avoid breaches of confidentiality by persons granted access to confidential data. For the bona fide researchers who are envisaged as the users of the DAC, the threat of professional disgrace and the lack of data for further research would be extremely powerful incentives to comply with the agreed conditions of access.

- Eurostat involvement in research

33. Eurostat remains closely involved in research into statistical confidentiality. For example, funding of the Statistical Disclosure Control project under the European Commission's Fourth Framework Programme lead to great improvements to the ARGUS statistical disclosure control software, and to advances in disclosure control theory and methodology. Smaller scale projects under other programmes have been aimed at improving the understanding of particular techniques such as microaggregation. There is also a need for a better understanding of respondents' attitudes to confidentiality and of their understanding of the confidentiality promises which are given by statistical agencies. Such research should also address the extent to which the confidentiality of data reduces with the data age. It is thought that this may be particularly the case for enterprise data where information may cease quite quickly to be commercially secret. It is hoped that information from this research

could be used to minimise respondent concerns about confidentiality, and to guide policy on statistical confidentiality. It is intended that work on these issues will continue to be sponsored through the Commission's Fifth Framework Programme.

V. Conclusions

34. Developments in statistical confidentiality policies and procedures at the European level require a working partnership between Eurostat and the NSIs. Changes in legislation can be seen as moving towards EU-wide definitions of confidentiality. There is an increased role for Eurostat in co-ordinating and leading work on these issues. However the active involvement of the NSIs is of vital importance in ensuring that appropriate policies are agreed and implemented. It is important that information on best practice is shared among statisticians and researchers working on confidential data. It is also important that users and providers of official statistics gain a better understanding of each other's needs and concerns. This will help in the creation and implementation of appropriate policies.

35. Innovative approaches and ongoing research are required to ensure the continued protection of confidential data in an environment of technological change and increasing user demands.