



**Economic and Social  
Council**

Distr.  
GENERAL

CES/2003/5/Add.1  
21 May 2003

ENGLISH ONLY

---

**STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE**

**CONFERENCE OF EUROPEAN STATISTICIANS**

Fifty-first plenary session  
(Geneva, 10-12 June 2003)

**STATISTICAL DATA CONFIDENTIALITY**

Paper submitted by the ECE Secretariat

**I. INTRODUCTION**

1. In preparation of the Seminar on “Statistical confidentiality and microdata”, the UNECE secretariat, as a member of the organising committee, consulted Member Countries about any concerns they might have regarding confidentiality issues. Furthermore, in order to identify any specific concerns of the transition economies, an ad-hoc survey was carried out in these countries with the help of a short questionnaire. The results of the ad-hoc survey are available in document CES/2003/5 “Data Confidentiality – results of the ad-hoc survey carried out in the transition economies”.
2. All Member States were invited to provide UNECE with short reports about any concerns they might have regarding confidentiality issues; they were asked to share their problems and concerns regarding confidentiality rather than current practices and legislation.
3. The following eighteen countries provided ECE with short reports about their concerns with data confidentiality: Armenia, Austria, Bulgaria, Canada, Czech Republic, Denmark, Germany, Italy, Kazakhstan, Kyrgyzstan, the Netherlands, Norway, Slovak Republic, Sweden, Switzerland, Turkey, United Kingdom and the United States.
4. This note summarises the main confidentiality concerns reported by the countries. The information provided was used in the preparation of the CES Seminar on Statistical Confidentiality and

Microdata. Some issues, however, could not be covered by the current Seminar and it might be worthwhile to bear them in mind for future consideration.

## **II. PROBLEMS AND CONCERNS OF DATA CONFIDENTIALITY IN THE ECE COUNTRIES**

5. The following key issues can be derived from the information provided by the countries which reported to the ECE secretariat:

### **Access to microdata without jeopardizing confidentiality**

6. More than half of the contributing countries refer in their reports to the challenges of various issues related to providing access to microdata.

7. Increasing demand from users in many countries to access microdata and the assumed “societal value added” of an efficient and creative way of using statistical data demands an improved access to microdata. In order to balance access to microdata and the protection of confidentiality and privacy, feasible and innovative solutions need to be found. Additional challenges arise from the availability of more and more information from various sources (internet, administrative registers) and increased computing power and new data mining and matching methods.

8. The following issues have been identified by the countries as being of some concern related to access to microdata:

- a) How to provide access to microdata? Where is the access to be provided? Access can be provided through centres at the agency, on-line/remote access, or through grids. How much control and supervision can/should the agency have over the researcher who is using the confidential data? Contractual issues – contents of the contract between the researcher and the NSO; this can also include the question of who covers the costs involved.
- b) Increased use of microdata results in a greater use of complex analytical models, which presents challenges in the application of disclosure avoidance methodologies, i.e. adaptation of disclosure avoidance to analytical outputs.
- c) Survey contents become more and more complex and detailed. The challenge is in providing needed design information in such a way as not to compromise the data confidentiality (for example by revealing the location of sample units or allowing the linking of related sample units).
- d) Use of files containing administrative data can lead to an increased ability to link back to the administrative data source, thereby resulting in the disclosure of respondent identities.
- e) A decentralised statistical system might create problems regarding both access to microdata and confidentiality issues (different policies, standards, combination of data from different sources/registers).

- f) Confidentiality protection of business microdata is generally considered to be more difficult than the protection of microdata of private persons.
- g) Statistical information is published by many different users, with often contradictory results. This is confusing for users.
- h) Users' interest for data vs. confidentiality concerns of respondents. (See also: Awareness of respondents and users of confidential protection, below).
- i) A common research project of Germany, the Netherlands, Italy, Spain and UK deals with computational aspects of statistical disclosure control. The importance of the implementation of best practices anonymisation methods is stressed. Well-established and commonly accepted standards for anonymisation methods are seen as a fundamental basis for EU-level release of scientific use microdata files.
- j) Is automatic confidentiality protection possible at all?
- k) Funding costs for research and development of new and improved methods.
- l) To what extent will researchers accept microdata with "noise" that statistical agencies insert to protect confidentiality, and how can agencies educate the research community about this issue?
- m) The linking of data might be legally forbidden by law even for scientific analysis. Research institutes and other national bodies show little understanding in this regard.

#### **Disclosure control for tabular data**

9. Several countries emphasise that there are new challenges related to the confidentiality protection of tabular data. These challenges arise on the one hand from an increasing demand from users (including government requirements) for more detailed information, and on the other from new dissemination methods (internet, large volume of multidimensional, hierarchical tables, cross-publication disclosure). Together with an increased computing power and new data mining and matching methods, this provides greater opportunity for attacks on confidentiality and thus leads to a need for more secure protection of tabular data.

10. Correspondingly, several countries ask for new or improved methods to ensure confidentiality in tabular data. Perturbative methods such as rounding and their impact are mentioned as an area for further research and for developing standards.

#### **When are data confidential and what to do – the need for concrete guidelines**

11. Several countries implicitly or explicitly express the need for concrete guidelines both for identifying confidential data and for protecting them (both for tabular data and microdata). This includes

the desire to learn from other countries about existing practices and concrete implementation of legislation. At the same time, it raises the question of how to improve methods of statistical disclosure limitation.

12. One issue concerning the need for concrete guidelines is software – which one to choose, learn more about best practices, and also introduce these best practices in all agencies that produce official statistics. In this context, one country also argues that, especially for microdata, automatic tools might not be completely sufficient (distortion of interference, ways to circumvent such tools).

13. The issue of actual implementation of data confidentiality principles is of great concern to ECE countries. Some sent a list of questions asking for guidance on how to deal with confidentiality issues in practice. To name just a few, the questions include whether public sources are confidential; rules for access to confidential data for all producers of official statistics; administrative, technical and organisational precautions against legal access of data; possible penal sanctions; rules and terminology for confidentiality; categories of data confidentiality and their delimitation (such as “top confidential”, “confidential within certain limits”, “no limits on dissemination”); access to confidentiality protection software such as ARGUS.

#### **Data confidentiality – legal and organisational aspects of its implementation**

14. Legal and organisational aspects of data confidentiality are mainly of concern to transition economies (see also CES/2003/5 “Data Confidentiality – results of the ad-hoc survey carried out in the transition economies”).

15. Some CIS countries still need to provide particular enterprise information to other government agencies “for internal use only”. In addition, one country reports on cases where other government agencies ask enterprises directly for their statistical reports for control and supervision purposes. At the level of regional statistical bodies, violations of confidentiality are observed. The need for further improvement of the implementation of confidentiality protection of legal units is also stated.

16. Several transition countries emphasise the need for a better “mainstreaming” of confidentiality protection in the statistical system (i.e. compliance with confidentiality protection of all producers of official statistics). Some transition countries work towards a better coordination and interaction in solving methodological and IT problems through setting up a Council on statistical data confidentiality. Another option mentioned to ensure compliance with data confidentiality is aiming at centralising the processing of statistical data. In addition, one country plans to organise seminars and other events to promote the principles of statistical confidentiality to all producers of official statistics.

17. The reports also show that some pre-accession countries have concerns regarding the implementation of the EU regulations with regard to confidentiality. Several very specific questions on how to implement the EU law applying to confidentiality have been raised by various countries.

## **Awareness of respondents and users of confidential protection**

18. Several countries make reference to the importance of both respondents' and users' awareness of confidential protection. The need to promote the capacity of producers of official statistics to handle confidentiality protection properly is stressed.

19. Another concern expressed by countries is that data users and their strong pressure to obtain access "to everything" is much more prominent in the public debate than the interests of the respondents with regard to confidentiality. As a result, increased access to data is provided and this might deteriorate the respondents' perception of confidentiality protection. Statistical agencies, however, very much depend on the confidence of the public and their willingness to respond to data requests.

20. With regard to studying respondents' perception, some countries conduct specific surveys on private persons' perception. The purpose is to increase the confidence of users/respondents, and thus the quality of data.

21. With regard to access to administrative registers, one country reports that procedures to access administrative registers in other government agencies are not yet in place. Therefore, many users request the data from the NSO, and the NSO's staff spend a lot of time explaining why they cannot submit the requested individual information. A common approach towards the work with administrative data is therefore needed.

## **Other issues mentioned**

22. Several countries refer to confidentiality issues in an international context. International organisations ask countries for increasingly detailed information. At the same time the requested data could be confidential at national (small country) or regional level. The question therefore arises as to whether these confidential data are to be released to international organisations and how to ensure confidentiality protection for the data then published by these organisations. Another aspect of the international context is the limitations that arise at international level if national/regional data are only provided in a confidentiality protected way (inconsistencies, loss of information). Therefore, one recommendation of the joint research project of Germany, the Netherlands, Italy, Spain and UK is to harmonise disclosure control methodology and practical tools at the international level.

23. Another issue of concern is the confidentiality protection of spatial data within a Geographic Information System where traditional confidentiality protection methodology is not designed to work.

24. One country argues that statistical agencies should move from the releasing of predefined sets of tables to allowing users to query tables as freely as possible, maintaining high standards of confidentiality protection of respondents. Therefore, statistical disclosure control methods must be applied to Websites according to their structure and flexibility as well as to the type, quality and level of detail of the information released. In many cases, statistical disclosure control must also be combined with electronic

access control. This makes clear how several problems have to be solved both from the statistical methodology point of view and the IT point of view.