UNITED
NATIONS

E

Economic and Social Council

Distr.
General

ECE/CES/2006/4/Add.1
8 May 2006

ENGLISH ONLY

ECONOMIC COMMISSION FOR EUROPE **STATISTICAL COMMISSION**

CONFERENCE OF EUROPEAN STATISTICIANS

Fifty-fourth plenary session
Paris, 13-15 June 2006
Item 2 of the provisional agenda

REPORT OF THE NOVEMBER 2005 WORK SESSION ON STATISTICAL DATA CONFIDENTIALITY

Note by the secretariat

1. The Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality was held in Geneva, Switzerland, from 9 to 11 November 2005. It was attended by participants from: Australia, Austria, Bulgaria, Canada, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Italy, Japan, Latvia, Lithuania, Mongolia, Netherlands, New Zealand, Norway, Portugal, Romania, Russian Federation, Slovakia, Slovenia, Sweden, Switzerland, Turkey, United Kingdom and United States of America. The European Commission was represented by Eurostat. Representatives of the European Central Bank (ECB) and the Organization for Economic Cooperation and Development (OECD) also attended. Participants from numerous universities and research institutes attended the work session at the invitation of Eurostat.

2. The agenda of the work session consisted of the following substantive topics:

(i) Web/on-line remote access (techniques, confidentiality protection and organizational issues);
(ii) Disclosure risk, information loss and usability of data;
(iii) Confidentiality aspects of statistical information taking into account register-based data;
(iv) Access to business microdata for analysis;
(v) Confidentiality aspects of tabular data, frequency tables, etc.;
(vi) Software for statistical disclosure control;

(vii) General statistical confidentiality issues (legal framework, political and conceptual aspects, terminology).

3. Mr. Anco Hundepool (Netherlands) acted as Chairman.

4. The representatives of Eurostat and UNECE addressed the meeting. They stressed the importance of protecting data confidentiality, the impact of technological developments, and recalled the tradition of international cooperation in this field. The importance of the Work Session was emphasized by the participation of 90 participants and 51 discussion papers submitted.

5. Mr. Pedro Diaz (Eurostat) expressed his gratitude to the UNECE for hosting the meeting. He attached a great deal of importance to the topics being discussed at this Work Session and hopes that progress in the field of data confidentiality can be speeded up. He noted that throughout the European Union, the harmonization of risk disclosure criteria has not been done. The frontier of what is confidential and what is not, is difficult to define. There is a lack of common methods and policy. He also mentioned that progress in remedying these problems has been very slow. A task force will be created at a higher level in the ESS to work on policy issues for confidentiality.

6. Mr. Heinrich Brüngger (ECE) identified the main problem in this area of how to reconcile the strict rules of statistics with the needs of researchers to have access to data. He noted that this Work Session brings these two groups together. He drew attention to the Fundamental Principles of Official Statistics and in particular to the 6th principle, which states that data are to be strictly confidential and used exclusively for statistical purposes. In surveys we promise respondents that their identity will be protected and we must honour this pledge. He also mentioned that in the manipulation of microdata files by adding noise or perturbation, this may reduce re-identification but may increase the risk of the wrong information being attributed to a unit. This could affect the credibility of the data provider. He noted that countries in transition still need assistance in developing their own models of good statistics.

7. The following persons acted as Session Organizers: Topic (i) – Mr. Anco Hundepool (Netherlands); Topic (ii) – Ms. Luisa Franconi (Italy); Topic (iii) – Mr. Eric Schulte Nordholt (Netherlands); Topic (iv) Mr. Josep Domingo Ferrer (University Rovira I Virgili, Spain); Topic (v) – Mr. Lawrence H. Cox (United States of America); Topic (vi) Ms. Sarah Giessing (Germany); and Topic (vii) – Messrs. Pedro Diaz and Jean-Marc Museux (Eurostat).

PUBLICATION OF PAPERS

8. Eurostat will publish a special volume in the series Monographs of Official Statistics devoted to this work session, containing a wide selection of the papers presented. In view of the large number of papers, it was agreed that invited papers should not exceed 12 pages and 8 pages for supporting papers. This requirement must be strictly observed and authors were invited to submit the final edited versions of their papers in PDF and Word or LaTex formats to:
maria-joao.santos@cec.eu.int no later than **2 December 2005**. The publication will be available in March 2006.

RECOMMENDATIONS FOR FUTURE WORK

9. The participants reviewed the recommendations for future work on the basis of a proposal put forward by an ad hoc working group composed of Ms. Sarah Giessing (Germany), Ms. Luisa Franconi (Italy), Mr. Anco Hundepool (Netherlands) and Ms. Maria Joao-Santos (Eurostat). When preparing the proposal, the working group took into account suggestions made by other participants during the meeting.

10. The participants considered it useful for national and international statistical offices to continue the exchange of experiences in the field of statistical data confidentiality. The Work Session, therefore, recommended that a future meeting on statistical data confidentiality be convened in 2007, subject to the approval of the Conference of European Statisticians and its Bureau, with a study programme as suggested below.

11. The following substantive topics were recommended for the study programme of the proposed 2007 work session:

- Web/online remote access questions (techniques and organizational issues);
- Glossary of terms for statistical data confidentiality;
- Disclosure risk, information loss and usability of data;
- Confidentiality aspects of tabular data, frequency tables, Geographical Information Systems, etc.;
- Software for statistical disclosure control;
- Synthetic data files for research/analysis training;
- User case studies on practical applications (more countries);
- User issues, SDC more visible to users.

Participants recommended that an organizing committee of the future work session try to group these topics into a smaller number of themes. They also recommended that the future work session should provide more time for discussion, and look at alternative ways of presenting numerous supporting papers (poster sessions, etc.). Organizing panel discussions was also recommended as a possible method of work. It was suggested that the work session be organized in such a way so as to make it more visible to the outside world (e.g. renaming the sessions to address more practical issues).

12. The participants felt that the following topics may also be of interest, and possibly discussed at one of the future meetings on statistical data confidentiality or a related subject (in order of preference):

- Linking external data;
- Access to business microdata for analysis;
- Intruder aspects;
- Case study on legal aspects in different countries;
- Statistical data confidentiality aspects of population and housing censuses taking into account classical, register-based and continuous censuses;
- Building a standard set of test data sets;
- Statistical disclosure control methodology assessment;
- Safety aspects of analysis results, disclosure assurance of analytical outputs;

- Links coordination with the Comité du Secret;
- Reducing the burden/costs of statistical disclosure control;
- Editing/imputation/SDC.

FUTURE WORK WITHIN THE FRAMEWORK OF THE CASC PROJECT

13. The participants took note of information on ongoing and planned activities within the Computational Aspects of Statistical Confidentiality (CASC) project. In particular they were invited to a conference that will take place in October-November 2006 in Rome. More information can be found at: **http://neon.vb.cbs.nl/casc/** .

FURTHER INFORMATION

12. The conclusions reached during the discussion of the substantive items of the agenda are contained in the Annex. All background documents and presentations for the meeting are available on the website of the UNECE Statistical Division (**http://www.unece.org/stats/documents/2005.11.confidentiality.htm**).

ADOPTION OF THE REPORT

13. The participants adopted the present report before the Work Session adjourned.

ANNEX

SUMMARY OF THE MAIN CONCLUSIONS REACHED AT THE WORK SESSION ON STATISTICAL DATA CONFIDENTIALITY

I. Web/on-line remote access (techniques, confidentiality protection and organizational issues
Session Organizer: Anco Hundepool (Netherlands)
Documentation: Invited papers by Denmark, Sweden, United States and University of Minnesota/Autonomous University of Barcelona; Supporting papers by Netherlands and United States

1. Papers presented under this topic focused on two types of access: (i) remote execution, which is less flexible but provides better disclosure control and where all outputs are checked; and (ii) remote access, which is more flexible but disclosure control is more difficult and final output is checked. When implementing remote access, statistical offices have to consider issues related to information technology, organization of remote access and selection of data/preparation of data sets to be accessed by researchers. Researchers are showing a growing interest in remote access to anonymised microdata.

2. The participants discussed examples of the technological setup. Dedicated computers are assigned separate from the production environment of the statistical office. Several offices use solutions relying on the server side – the most frequently used software packages are available for statistical analysis on the server in several countries. Other offices provide specific desktop utilities for researchers. For security reasons, researchers are usually not able to print or download the data or outputs to their workstations. Sophisticated authentication approaches are used (e.g. secure ID, smart cards and biometric identification). Various firewall configurations provide physical protection for data. In some applications, results were sent by e-mail. Another example moved outputs behind the firewall after checking them. Participants also confirmed the need to keep the set-up user-friendly and secure at the same time.

3. There was a discussion on the risk that a session is recorded (video-taped, photographed, etc.). This can hardly be avoided. The participants felt that this risk is not so important as the recorded sessions would not be too useful. It was more important that the data could not be downloaded and there was no copy-paste possibility between desktops. Technical solutions were implemented to avoid such risks.

4. On the organizational side, the usual practice is that access is not granted to everyone. Only research institutions (public or private) with a stable research environment (having a management and a number of researchers) are allowed to access the anonymised microdata. Another approach is to limit the access to concrete people, projects, variables and a specified time period. There were various approaches to checking the outputs. In some cases only the final output was checked and made available to the researchers; in other cases all outputs, including intermediary ones, were checked. One example is when the statistical office’s staff

performed the verification, while another example used robust algorithms for verification.

5. There was a question by one delegation about whether the anonymised microdata should be provided only to researchers. Some participants considered that there should be a distinction between the public use files that may be available to everyone, and the detailed microdata files that had to remain restricted.

6. There were various approaches to the selection of data. In some cases the approved data sets were available to all users, while in other cases only variables related to the research project. In some cases very detailed universe formation rules are defined in order to assess the risk and limit the creation of tables accordingly.

7. Access to microdata may be interesting also from the historical viewpoint. A project aiming at preserving and making available historical data sets from population censuses (IPUMS) was presented. The issues included ensuring that no important historical facts were lost and that all data sets were harmonized. The legal, administrative and technical setting of the project ensures confidentiality protection. The protective measures include suppression of geographical details, converting dates to ages and blurring/aggregating the sensitive codes. Legal measures imposed signing the licenses agreements binding the researchers as well as the researchers' institutions, and prohibited any redistribution of data.

8. Software tools for two projects on remote access introduced within this topic were presented at a special session: OnSite@Home (Statistics Netherlands) and IPUMS (University of Minnesota).

II. Disclosure risk, information loss and usability of data

Session Organizer: Luisa Franconi (Italy)
Documentation: Invited papers by University of Naples/University of Plymouth, United Kingdom/University of Southampton/Hebrew University, Hebrew University/University of Southampton, Carnegie Mellon University and University of Alicante and University of Southampton; Supporting papers by Italy, New Zealand, United Kingdom, United States, University of Kentucky/Oklahoma State University and University of Constance

9. The discussion on this topic focused on the release of microdata files that may lead to risk of disclosure. Comparison of anonymised microdata files with accessible registers and archives may permit re-identification of records.

10. Statistical literature provides a definition of risk measures, but there is no formal definition when the file is considered to be safe. Generally, there is no zero risk of disclosure. Some computer science literature defends the opinion that any data must be released with noise (the noise must be small enough so that information on large data sets remains useful). However, the question was raised whether releasing data with noise is acceptable for statistical agencies.

11. The participants discussed several methods for assessing disclosure risk, and agreed that

risk assessment is crucial to the disclosure control. The discussion stressed the importance of statistical models. The approaches included a set of models using contingency tables. Superpopulation models were constructed using the population and sample frequencies. Several models were developed for estimating the record level measures of the disclosure risk using statistical models based on Bayesian methods. Log-linear models applied to contingency tables aim at estimating the sample disclosure risk. An approach based on canonical correlation analysis for assessing both re-identification and value disclosure risk was discussed.

12. The discussion stressed the issue of complexity of calculations. The experience, on the part of some participants, showed that it was possible to run complex algorithms on present computers. Some approximations, aiming at decreasing the computational complexity, were also suggested. The computational complexity of different models is very different and is not related to the complexity of the model itself.

13. There was also a presentation dealing with techniques of microdata release. A random orthogonal matrix masking methods (ROMM) was presented. The measures for risk when using this method and when using a concurrent noise introduction method were compared. The participants discussed the applicability of this method to categorical and continuous data. The evaluation of the method has not been finished yet.

14. The participants also discussed the use of Confidentialized Unit record Files (CURF) for dissemination of categorical microdata. The method uses small samples that have almost exact census proportion for a few census variables and minimized sampling error for others. CURFs represent a new advantages and limitations. The method is currently being evaluated by a pilot group of researchers.

15. The Post Randomisation Method (PRAM) also serves dissemination of categorical microdata. It perturbs the original data – categorical variables are changed according to prescribed probability. The tests revealed that the method shows good results in minimizing the disclosure risk as well as the damage to the original data set.

16. The disclosure risk of public use files created from confidential files was also considered in the discussion. An automatic record linkage experiment was suggested and is currently being experimented on in one country to assess such risk.

III. Confidentiality aspects of statistical information taking into account register-based data
Session Organizer: Eric Schulte Nordholt (Netherlands)
Documentation: Invited papers by United Kingdom/University of Southampton/Hebrew University, Eurostat and University of Alicante; Supporting papers by Italy, Norway and Norway/Eurostat

17. Participants recognized that presently there are no European best practices available. There are differences in the perception of risk resulting in significant variations of disclosure risk. An international harmonization would ease international comparison. Eurostat is attempting to pursue harmonization efforts and is currently working on common attacker/intruder

scenarios, risk measures and thresholds and common rules for anonymisation. Residual flexibility would allow adapting to national situations following national assessment according to common standards.

18. The comparison between pre-tabular record swapping and post-tabular small cell rounding was reviewed. The strength of record swapping was in keeping the table totals consistent; preserving marginal distributions at higher aggregated levels, and its usability for geographic variables. Targeted swapping lowers the disclosure risk, but increases distortion in the distributions of the table. The advantages of small cells rounding comprise the full protection of high-risk cells, low information loss, transparency for the users and possibility of taking the method into account in statistical analysis.

19. The longitudinal data, originating from two or more sampling frames with measurements for multiple time periods were discussed The possibility of linking different time periods may be a cause of additional disclosure risk. The classical disclosure control methods may not be efficient if used alone. It was suggested to combine methods such as data masking, synthetic data, licensing and research data centres.

IV. Access to business microdata for analysis

Session Organizer: Josep Domingo Ferrer, Rovira i Virgili University
Documentation: Invited papers by Germany, University of Tübingen, National Institute of Statistical Sciences/University of Cincinnati/Duke University/Bristol-Myers Squibb, Rovira i Virgili University and IIIA-CSIC; Supporting paper by United Kingdom

20. The agencies acting within the system of official statistics are very careful when sharing their possibly confidential or proprietary data with others who own related databases. This may represent an obstacle for conducting statistical analyses. Several methods were discussed at the Work Session for secure computation that may allow sharing data without compromising data confidentiality. These methods included secure summation protocols, secure matrix product protocols, and synthetic data approaches.

21. The Work Session examined another study on estimating re-identification risk from partially synthetic microdata. Guidelines on how to tune synthetic data generation were provided as a conclusion of this contribution.

22. The Work Session examined a study on estimating disclosure risk using cross-database matching for anonymised business data in combination with external data. The re-identified units are further analysed to determine if they contribute benefit to potential data intruders. A stronger anonymisation method is subsequently applied on these units. This approach was used for the creation of Scientific-Use-Files. The opinion was expressed that intruders follow the economic rationale, and they would refrain from re-identification efforts if the risks of failure were too high.

23. The simultaneous consideration of data quality aspects together with disclosure control deserves the attention of statistical offices. A possible way to achieve this is to accompany the

anonymisation by testing the information loss. Several methods used for synthetic data were discussed. The question was raised how the data provider knows about the information needs of potential users.

24. The participants also discussed the secure computation approaches. These include secure servers where users can conduct their analysis without disclosing the confidential data. The case of health insurance companies was quoted as an example: the full data set may need to be analysed, but under no circumstances insurance companies would accept the risk of having their data disclosed.

V. Confidentiality aspects of tabular data, frequency tables, etc.

Session Organizer: Lawrence H. Cox (United States of America)
Documentation: Invited papers by United Kingdom, United States, United Kingdom/University of Southampton/Hebrew University and University of La Laguna; Supporting papers by Australia, United Kingdom/University of La Laguna, United Kingdom and United States

25. There are various forms of tabular data. One of them is frequency (count) data organized in contingency tables. Tabular data may also take the form of magnitude tables (e.g. income, sales, number of employees, etc.). Magnitude tables are published in a large number and they can disclose contributions more easily than other tables. Therefore, statistical disclosure control for these tables has received great attention in literature and a computer package mainly devoted to the protection of this type of tables, τ-Argus, has been developed.

26. Recent research has gone beyond cell suppression, which thwarts and distorts statistical analysis, to perturbative methods like controlled tabular adjustment (CTA) and further to balance data quality and confidentiality, quality-preserving CTA. A pressing issue is how to efficiently compute such problems.

27. Rounding is one method used to protect tabular data. Effects of rounding on data quality and utility were considered at the Work Session. Various methods for rounding and various choices for the rounding base are available. These may be compared in two ways: (i) bias and variance and (ii) effects on the underlying distribution. Methods discussed at the meeting included conventional rounding, modified conventional rounding, zero-restricted 50/50 rounding and unbiased rounding.

28. An alternative to cell suppression and use of fixed intervals was suggested. This method consists of fixed intervals that contain the sensitive value. The purpose of this method is to prevent outside intruders from gaining identifiable information on individual contributors to the cell. The participants also considered using inference techniques for minimizing the disclosure risk.

29. The discussion also considered the effect on information loss of disclosure control techniques on tabular data. The studies showed that stochastic disclosure control processes have varying effects on data quality. Some suggestions were made that could lead to the development

of more detailed guidelines: a table that has only one or two columns of small values and the rest large values should not be suppressed since inevitably the secondary suppressions will involve some of the larger cells. A table that is uniform has less information loss regardless of the perturbation method. It is clear that the information loss measures perform differently depending on the characteristics of the table and the perturbation methods need to be tailored to the specific type of table. It was emphasized that each type of tabular data needs a specific disclosure control method to prevent information loss.

30. The discussion focused on the specific issue of perturbation. Some of the methods alter the small aggregates – e.g. in cases when it is possible to re-identify the concrete business. While this prevents the disclosure of true information, there is a new risk of identifying wrong information about a particular business. Some participants suggested that a solution could be to indicate what cells were perturbed in the final output. Another suggestion was to pursue a policy debate on this issue, as it appeared more policy related than technical.

31. The participants attempted to compare disclosure control methods for microdata with methods used to protect tabular data. Those for microdata are mostly statistical methods. Complex mathematical algorithms are used for protecting tabular data. In this connection, the question of computational complexity was also considered. Experiments showed that an ordinary computer could handle a table in the order of 1'000’000 cells within 40 minutes. On the other hand much smaller but more complex tables could not be solved to optimality after several hours of computation. In these cases, near optimal solutions are provided. It means that the complexity of the table, contribute to the length of time needed for computation.

32. Similarly, as mentioned under previous topics, national statistical offices also feel the need for a consolidated approach to tabular data protection. The Code of Practice and Protocol on Data Access and Confidentiality by the UK ONS were quoted as an example.

VI. Software for statistical disclosure control

Session Organizer: Sarah Giessing (Germany)
Documentation: Invited papers by Germany, Germany/Technical University of Catalunya, United States and University of Manchester; Supporting papers by Netherlands and United States

33. The presentations in this topic covered practically the entire field of statistical disclosure control.

34. One presentation focused on a method producing safe output for complex statistical analysis in a remote access environment. The “jackknife” method overcomes typical disadvantages of anonymised microdata that are the only protection against specific disclosure scenarios, non-protection of certain individual values, low quality results, biased results of complex analyses and impossibility of some analyses. Some prototypes of the “jackknife” method are available for evaluation, namely classical and robust univariate descriptives and tests (as in SAS Proc Mean / Proc Univariate and some more); two-way frequency tables with table statistics and tests (as in SAS Proc Freq); and (non-)linear least squares regression (as in SAS

Proc NLin). Presently the replacement values are drawn from the distribution defined by the administration. During the discussion, suggestions were made to draw them from a distribution calculated from the data, for example the imputation model.

35. An algorithm for Controlled Tabular Adjustment (CTA), using minimum distances, was presented at the Work Session. The aim is to change values of sensitive cells sufficiently and adjust values of non-sensitive cells minimally to maintain the additive table structure. The quality criteria for the CTA method were also considered. They include the possibility of combining the CTA and cell suppression and that the CTA and cell suppression should result in a comparable amount of accurate data being released taking into account the data's relevance and significance. The future development will include processing of multivariate data, determining adjustment senses for sensitive cells in restricted CTA and researching practicability of using CTA to handle disclosure limitation issues of table servers.

36. Complementary cell suppression methodology was discussed. It was proved that using interior point solvers it is possible to derive close estimates of confidential cells. To avoid this, it was suggested to use the small cells suppression with caution, and not to be afraid of over-protection. The future work may involve looking into synthetic tabular data. Another example of cell suppression, which was applied to the census of agriculture, was also presented at the meeting.

37. A concept of a program (SUDA) for classifying cells according to their disclosure risk was presented. The concept used for this program is called special uniqueness. The algorithm permitted efficient searching for special uniques. Work on refinement of the program is under way, both in terms of the computer science background and the statistical disclosure algorithms on which it is based. A project is under way that aims at grid enabling SUDA, which is expected to increase efficiency and overcome some disadvantages. This program can run on Windows and Linux computers. However, its distribution is subject to a licensing procedure because the software is believed to be useful also for intruders, and such use has to be avoided. The discussion stressed the there is little attention paid to group disclosure. SUDA may have a chance to solve this, and the authors promised to look into it. There was also a discussion on whether to select only key variables when searching for special uniques or use a broader model.

38. The use of τ-Argus software for cell suppression was also discussed. The presentation focused on the clever usage of microdata. Confidentiality patterns in τ-Argus can be influenced. A practical presentation of μ-Argus and τ-Argus software, for microdata and tabular data protection, was organized for interested participants at a special session. More details can be found at http://neon.vb.cbs.nl/casc/.

39. There are already many complex algorithms available in various software tools. The development is resource intensive, and so duplication is not desirable. It was therefore suggested to consider creating a library for statistical disclosure control software.

VII. General statistical confidentiality issues (legal framework, political and conceptual aspects, terminology)

Session Organizer: Pedro Diaz and Jean-Marc Museux (Eurostat)
Documentation: Invited papers by Canada/Germany/Netherlands/University of Manchester, United Kingdom, Eurostat and UNECE; Supporting papers by Canada, Portugal and UNECE

40. The participants were informed about present work on the Glossary on Statistical Disclosure Control. The authors tried to look at different websites and find common definitions. The ambition of the project was not to invent new definitions, but to make those that existed easily available on the Internet. The draft is available on the CASC website (http://neon.vb.cbs.nl/casc/). Comments are welcome, and can be sent to Eric Schulte Nordholt (esle@cbs.nl). The authors offered assistance to those who would volunteer to translate the glossary into other languages. It is currently available in English, with a tentative plan to translate it into French and German.

41. A balance needs to be found between the need to provide users with access to microdata and the need to protect the confidentiality of respondents. The legal aspects often comprise a complex variety of general laws on access to information and laws specific to statistics. In addition to the legal regulatory measures, official statistics need to develop policies for access to microdata. This is undertaken at various levels within departments at the national level. There are also activities aimed at identifying agreed good practices on confidentiality protection and access to metadata. The practical application involves risk assessment and methods to control disclosure risk. It also covers various access options that restrict access to authorized use and authorized users and prevent eventual intrusion.

42. When discussing the dissemination of potentially confidential information by international organizations, the lack of harmonization was pointed out. This leads to accepting the most restrictive rule from countries concerned. There are EU regulations on access to confidential data for scientific purposes (831/2002), and these were discussed in depth. ESS work on statistical disclosure control will also be pursued through the Centres of Excellence (CENEX). A consolidated legal confidentiality framework for the EU should be sought in the future. Plans to streamline and better implement EU Regulation 831/2002 will be presented to the Committee on Statistical Confidentiality in December 2005. Eurostat is interested in promoting remote access for researchers, and a current reflection on the production of public files will is on going.

43. A representative of the UNECE presented the work currently undertaken by the CES Task Force on Confidentiality and Microdata. Mr. Denis Trewin, Chief Statistician of Australia, chairs the Task Force. The membership consists of representatives from Autralia, Canada, Denmark, Georgia, Italy, Poland and the UNECE. The goal is to prepare the "Core Principles of Access to Microdata". The Task Force plans to complete the work with the adoption of the Principles at the June 2006 plenary session of the Conference of European Statisticians. More information can be found on the UNECE website at **http://www.unece.org/stats/documents/tfcm.htm**. Comments may be sent to dennis.trewin@abs.gov.au and tiina.luige@unece.org.

44. Statistical offices also shared experiences in their approaches to confidentiality. The first example was the use of Argus software for tabular data protection and preparation of the Public and Scientific Use Microdata Files (PUMF, SUMF and microdata files for use in the office premises). The other presentation dealt with managerial issues of statistical confidentiality, with particular attention to preserving trust in the statistical agency and protecting a fair competitive business environment. Plans to take into consideration international aspects such as international access and data provided to international organizations were also discussed.

45. There seems to be a need to differentiate between the sensitivity of statistical data based on the subject matter area. For example, the data on turnover may be more sensitive than those on the number of employees, etc.

46. The participants considered it useful to develop a framework for measuring the utility. Some countries have regular contacts with the user community, and they try to take on board their needs when taking decisions about data confidentiality.

47. Participants stressed that the environment convention sometimes obliges disclosing data that would normally be considered confidential. Any public authority must provide information on emissions within twenty days. This paradox would deserve further discussion by statisticians. The suggestion was made to increase awareness, mainly of the management in the statistical offices, in the legal issues related to data protection.

- - - - -