

**UNITED NATIONS ECONOMIC COMMISSION FOR
EUROPE**

EUROPEAN COMMISSION

CONFERENCE OF EUROPEAN STATISTICIANS

**STATISTICAL OFFICE OF THE EUROPEAN
UNION (EUROSTAT)**

Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality
(Tarragona, Spain, 26-28 October 2011)

REPORT OF THE OCTOBER 2011 WORK SESSION ON STATISTICAL DATA CONFIDENTIALITY

Prepared by the UNECE secretariat

PARTICIPATION

1. The Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality was held in Tarragona, Spain, from 26 to 28 October 2011. It was attended by participants from: Australia, Austria, Bosnia and Herzegovina, Canada, Croatia, Estonia, Finland, France, Georgia, Germany, Hungary, Italy, Japan, Latvia, Mexico, Netherlands, Norway, Poland, Russian Federation, Serbia, Singapore, Slovenia, Spain, Sweden, Turkey, United Kingdom and United States of America. The European Commission was represented by Eurostat. Representatives of the Organisation for Economic Co-operation and Development (OECD) and European Central Bank (ECB) also attended. Participants from numerous universities and research institutes attended the work session at the invitation of the UNECE secretariat.

ORGANIZATION OF THE MEETING

2. The agenda of the work session consisted of the following substantive topics:
- (i) Disclosure risk assessment;
 - (ii) Software research and development;
 - (iii) Census and other applications;
 - (iv) Balancing data quality and data confidentiality;
 - (v) Privacy for new types of microdata: sequence data and mobility data;
 - (vi) International projects, groups and forum dealing with data access, release and related methodologies;
 - (viii) Trans-border access to microdata;
 - (ix) Statistical disclosure limitation for table and analysis servers: how to make outputs of modern data access infrastructures safe.

For technical reasons, topic (vi) Data integration, was cancelled.

3. Mr. Anco Hundepool (Netherlands) acted as Chairman. He expressed his gratitude to the University Rovira i Virgili in Tarragona for hosting the meeting.

4. Mr. Josep Manel Ricart, Vice-Rector of the Universitat Rovira i Virgili, opened the meeting and welcomed participants. He expressed his gratitude to Prof. Domingo-Ferrer for his work in the field of data confidentiality as UNESCO Chair on Data Privacy. He felt that the meeting will help further development of international networks on a range of topics. He wished the meeting fruitful discussions.

5. Mr. Frederic Udina Abelló, Director of the Statistical Institute of Catalonia IDESCAT, welcomed participants to Catalonia and Tarragona. Although IDESCAT is a small statistical institute, it is increasingly active and welcomed this opportunity to develop further contacts with the international statistical community.

6. The provisional agenda was adopted.

7. The following persons acted as Session Organizers/Discussants: Topic (i) – Ms. Luisa Franconi (Italy) and Ms. Michelle Simard (Canada); Topic (ii) – Mr. Anco Hundepool (Netherlands); Topic (iii) – Mr. Eric Schulte Nordholt (Netherlands); Topic (iv) – Mr. Lawrence H. Cox (United States of America); Topic (v) – Mr. Josep Domingo Ferrer (University Rovira i Virgili, Spain); Topic (vii) – Ms. Luisa Franconi (Italy) and Ms. Michelle Simard (Canada); Topic (viii) – Ms. Aleksandra Bujnowska and Mr. Rainer Muthmann (Eurostat); and Topic (ix) – Ms. Sarah Giessing and Mr. Tim Hochgürtel (Germany).

RECOMMENDATIONS FOR FUTURE WORK

8. The participants reviewed the recommendations for future work on the basis of a proposal put forward by an ad hoc working group composed of Peter-Paul de Wolf (Netherlands), Michelle Simard (Canada) and Johan Heldal (Norway).

9. The participants considered it useful to continue the exchange of experiences in the field of statistical data confidentiality, and recommended that a future work session on statistical data confidentiality be convened in 2013, subject to the approval of the Conference of European Statisticians and its Bureau. Statistics Canada offered to host that work session in Ottawa. The following topics were proposed:

- New methods for protection of tabular data
- New methods for protection of microdata, including multivariate analysis
- Remote access
- Software developments (demonstrations)
- The trade-off between quality, utility and privacy
- Confidentiality of geo-spatial data
- Event-based statistics (events versus units)
- Panel discussions on transparency and/or privacy concepts (such as differential privacy)
- Confidentiality issues related to
 - Large datasets ('big data', streaming data)
 - Data linkage
 - Data visualization
 - Data from social networks
- Case studies on any of the above subjects
- The need for international guidelines on any of the above subjects.

FURTHER INFORMATION

10. The conclusions reached during the discussion of the substantive items of the agenda are contained in the Annex. All background documents, presentations and the final report for the meeting are available on the website of the UNECE Statistical Division (<http://www.unece.org/stats/documents/2011.10.confidentiality.htm>).

11. The participants expressed their great appreciation to the Universitat Rovira i Virgili for hosting this meeting and providing excellent facilities for their work.

12. In view of his approaching retirement, Josep Domingo-Ferrer thanked Mr. Anco Hundepool for having chaired the Statistical Data Confidentiality Work Sessions since the first meeting in Thessaloniki 12 years ago. He said that Anco was like a father to the Work Sessions. He also thanked the Organizing Committee and his team for their efforts in organizing and helping with the meeting.

ADOPTION OF THE REPORT

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

ANNEX

**SUMMARY OF MAIN CONCLUSIONS REACHED AT THE
JOINT UNECE/EUROSTAT WORK SESSION ON STATISTICAL DATA CONFIDENTIALITY**

Tarragona, Spain, 26-28 October 2011

Topic (i): Disclosure risk assessment

Session Organizers/Discussants: Luisa Franconi, Italy and Michelle Simard, Canada

Documentation: Invited papers by Germany and Italy (2); supporting papers by Germany, Canada and IIIA/CSIC

1. With more and more institutions considering the release of synthetic data, methods for assessing risk of disclosure in such data are necessary. This session reviewed the need for proper and formal definitions of disclosure risk as well as sound methodologies for its assessment.
2. The German presentation showed that with remote access there is still a risk of disclosing sensitive information even though the actual data are not directly available. Multivariate procedures such as factor analyses can also lead to a breach of confidentiality if applied in a sophisticated manner to exploit certain features of the data.
3. The representative of Italy discussed the disclosure risk for high dimensional business microdata using the Enterprise System of Accounts survey as an example. The detection of statistical units lacking in consistency with respect to the process generating the majority of observations, represents a relevant first step to assess the disclosure risk of business microdata. As tests are uninformative about the kind of data heterogeneity, any judgment requests further analyses and can imply subjective choices.
4. Italy also presented preliminary findings concerning the simultaneous release of a microdata file for research purposes and a public use file. They provided a brief analysis of the relationship between the two as well as an illustration of different sub-sampling strategies that might be used for disclosure control. The results obtained show that this strategy could be exploited in practical settings for the release of public use files.
5. Other presentations included:
 - Development and challenges of the real time remote access project and the improvements made in its methodological aspects in Canada;
 - The representative of Universitat Autònoma de Barcelona presented a joint project with the Artificial Intelligence Research Institute (IIIA), Spanish Council for Scientific Research (CSIC) on a supervised learning approach for distance-based record linkage which determines the optimum parameters for the linkage process and an evaluation and a comparison between three different alternatives of such a method.
6. During the discussion, the following points were raised:
 - The possible use of the results of the German study for automatic analysis of output of research centres.
 - Evaluation of risk of a given record in the approach outlined in the second Italian presentation.
 - Whether to allow research centre users to integrate their own data with official microdata sets.
 - The possibility of contacting commercial software providers to do some collaborative work to “customize” output modification.

Topic (ii): Software research and development

Session Organizer/Discussant: Anco Hundepool, Netherlands

Documentation: Papers by Austria, Canada, Finland, Netherlands, University of Manchester and Oklahoma State University/University of Kentucky

7. Software always plays a central role in statistical data confidentiality (SDC). This session provided an opportunity for organizations to show recent developments in their software. The following presentations and demonstrations were made:

- Data shuffling to protect confidential data (Oklahoma State University/University of Kentucky). This is a method that combines the strengths of data perturbation and data swapping to ensure maximum protection against identity and value disclosure.
- Key Variable Mapping System II (University of Manchester) where metadata gathered on forms are stored in a metadatabase. The paradigm process is to map two sets of metadata for particular databases. Using a prevalence metric and form classifications allows matching possibilities to be assessed to allow for more sophisticated analysis.
- A computational framework to protect tabular data – R-package sdcTable (Austria). This is a free-of-charge and open source software that is available on the R comprehensive archive network <http://cran.r-project.org>. It provides methods to solve the secondary cell suppression problem for multidimensional and hierarchical tables.
- Negative cell values, singletons and linked tables in Tau-Argus (Netherlands). Recently, some major improvements have been implemented mainly concerning the modular approach: it is now possible to deal with tables that have some negative cell values. A new solution for cell suppression to cope with the presence of singletons has been implemented and an automated way to deal with a set of linked tables is now available.
- New business survey confidentiality software G-Confid, created by Statistics Canada was based on methodology used in the old Confid system for cell suppression in tabular data. It can handle any table size & number of dimensions subject to SAS & hardware limitations.
- A remote access in Statistics Finland was described in a short presentation and poster, including a description of the system and tools, and the guidelines and rules for its use.

8. Software demos were a successful experiment to be repeated.

Topic (iii): Census and other applications

Session Organizer/Discussant: Eric Schulte Nordholt, Netherlands

Documentation: Invited papers by Germany, Poland and United Kingdom

9. Most countries have conducted their census in 2010/2011. The presentations in this topic focused on the practical implementation and solutions for on-going initiatives related to the protection of Census and other detailed data covering the entire field of statistical disclosure control.

10. Poland presented their innovative use of GIS technology in the recent population census. It was used at every stage in which digital maps were required. They were essential tools for census enumerators (for mobility in the field, verification of the sampling frame, etc.), and their managers, who could verify the progress of the census work and the route or location of enumerator on the map. They will also be used to disseminate census results.

11. The United Kingdom provided an overview of the statistical disclosure control process for communal establishments (prisons, hospitals, hotels etc.) for the 2011 population Census. Records for individuals within communal establishments can be swapped where necessary to protect confidentiality, though with the constraint of minimizing data utility loss. The aim of this approach is to provide the necessary protection required by law, without significant damage to the data.

12. In Germany, the 2011 population census was register-based, with an additional sample survey to verify the register and include information that was not available from registers. To secure the confidentiality, additivity and consistency of the published tables, a pre-tabular anonymization method was used. The preferred method among the tested ones was a variant of microaggregation called SAFE. This approach has been tested with real data from the 1987 population census of West Germany.

13. The discussion raised the following issues:

- When moving to a new technology, it is important to address concerns in the population about the confidentiality of data collected during the census.
- Public acceptability in Poland is growing over time but progress must be made in protecting all data.
- How to find suitable candidates for swapping in the UK approach and how to test the results obtained.
- The potential risk associated with swapping attributes from the point of view of respondents whose data are changed.
- Confidentiality issues for geo-spatial outputs need further work and could be considered in a future work session.
- The treatment of confidentiality in the European Union Census Hub project – at present this is dealt with at the national level. All European Union countries make use of one of the agreed methodologies to conduct the 2011 Census.

Topic (iv) Balancing data quality and data confidentiality

Session Organizer/Discussant: Lawrence H. Cox, United States of America

Documentation Session A: Invited papers by University of Kentucky/Oklahoma State University, Germany, Vanderbilt University/IBM-Research-Zurich/Cardiff University; Supporting papers by Japan, United Kingdom/University of Essex and Italy

Documentation Session B: Invited papers by United States, Austria, Universitat Rovira i Virgili; supporting papers by Germany, United States/Ben-Gurion University, Eurostat, University of La Laguna

14. From the papers presented in this topic, it was evident that there is a strong international interest in balancing data confidentiality and data quality. The overall thrust was to move beyond providing protection first and adjusting for quality second to integrated approaches. The session recommended to investigate/use methods whose statistical data limitation (SDL) and quality properties can be examined – and even quantified – analytically. The issues raised covered:

- Variance estimation
- Measuring disclosure risk
- Controlled tabular adjustment
- Hybrid SDL methods
- Differential privacy

15. The paper from the University of Kentucky/Oklahoma State University “Disclosure risk when responding to queries with deterministic guarantees” was about responding to ad hoc queries using Confidentiality via Camouflage, a procedure intended to provide interval responses to queries.

16. The Institute for Employment Research, Germany, addressed upward bias and possibly infeasible (negative) estimates of variance for fully synthetic data and proposed an improved variance estimation methodology in their paper “Improved Variance Estimation for Fully Synthetic Datasets”.
17. Moving beyond traditional focus on survey data, the paper by Vanderbilt University/IBM-Research-Zurich/Cardiff University “On Balancing Disclosure Risk and Data Utility in Transaction Data Sharing Using R-U Confidentiality Map” dealt with data protection and data quality for transaction data and offered some possible solutions.
18. Microdata from Japanese official statistics just started to become available in 2009. The paper on “Quantitative Methods to Assess Data Confidentiality and Data Utility for Microdata in Japan” described early approaches to microdata SDL in Japan.
19. The paper “Sharing risks, sharing benefits: Data as a public good” by the United Kingdom/University of Essex applies economic theory to improve the efficiency of access to confidential data viewed as a public good.
20. The representative of Italy presented a paper describing confidentiality issues associated with ESSnet.
21. “The Case For—Or Against—Hybrid SDL Methods” presented by the representative of the National Institute of Statistical Sciences in the United States questions some pros and cons of hybrid SDL methods through a discussion of five published papers based on an evaluation of data protection, quality preservation, and transparency of hybrid methods.
22. The paper by Austria “Comparison of perturbation methods based on pre-defined quality indicators” examined the extent to which SDL treated data meet quality standards for original data established by Eurostat.
23. In the paper “On differential privacy and data utility in SDC”, the Universitat Rovira i Virgili examined differential privacy from the standpoint of the noise distribution used for perturbation.
24. Germany presented another paper on “Some aspects concerning analytical validity and disclosure risk of CART generated synthetic data” in which it raised and examined technical questions regarding the generation of synthetic microdata from CART models.
25. The paper by the United States/Ben-Gurion University on “Differential Privacy – A Primer for the Perplexed” addressed potential misconceptions regarding the purpose and characteristics of differential privacy.
26. Eurostat reported on a study to quantify the information of released disclosure-treated data vis-à-vis that of original data in their paper “Analysis of information loss in European data due to confidentiality”.
27. The University of La Laguna described a mathematical model for enhancing controlled tabular adjustment, called E(nhanced)CTA.
28. The following issues were raised:
 - The conditions in which controlled tabular adjustment is an appropriate method for disclosure control and how best to compute CTA.
 - An advantage of enhanced CTA is that it can treat large numbers of variables very quickly.
 - In general it is not recommended to combine cell suppression with CTA but there are times when it is suitable.

- The effectiveness of the differential privacy approach in different circumstances.
- Giving access to data by public institutes can be costly, but often has clear benefits. In the United Kingdom, data owners are in general government departments. Researchers need to explain the benefits of their research to encourage more data being made available.
- The appropriate balance between quality and confidentiality depends on the type of data user.
- Conflicts between notions of transparency and confidentiality need to be addressed.
- Rethinking data release formats could help to get around some confidentiality constraints. Recommendations on the risk associated with the different release formats could help.

Topic (v) Privacy for new types of microdata: sequence data and mobility data

Session Organizer/Discussant: Josep Domingo Ferrer, Rovira i Virgili University of Tarragona, Spain

Documentation: Invited papers by Norway, University of Michigan and Universitat Rovira i Virgili; supporting paper by Universitat Rovira i Virgili

29. This session considered confidentiality issues in new contexts. For example, mobility and trajectory data are useful for planners building new transportation systems. Sequence data are increasingly collected from many sources, such as transactions of a customer over time. How to anonymize such data for analysis is a new challenge.

30. Statistics Norway presented a paper on “Anonymized integrated event history datasets for researchers” that described the solutions they have found and some challenges they face in risk assessment and anonymization of such longitudinal data. A preliminary solution is now being tested among researchers. The outcomes of the tests will influence the further development of the anonymization procedure.

31. Research undertaken by the University of Michigan considers a new method for disseminating public-use microdata that contain more geographical details than are currently being released. Observed survey values are replaced with imputed, or synthetic, values generated from a posterior predictive distribution. A hierarchical Bayesian model is used to preserve the small area inferences and simulate the synthetic data. Confidentiality protection is enhanced because no actual values are released. The method was demonstrated using data from a prominent federal statistical survey: the 2005-2009 American Community Survey. The analytic validity of the synthetic data is assessed by comparing the synthetic small area estimates to those obtained from the actual data.

32. The Universitat Rovira i Virgili presented a paper on “Anti-discrimination and privacy protection in released datasets that tackles discrimination discovery and prevention in data mining. The framework for discrimination prevention can be described in terms of two phases: discrimination measurement; and data transformation. The purpose is to transform the original data in such a way to remove direct and/or indirect discriminatory biases, with minimum impact on the data and on legitimate decision rules, so that no unfair decision rule can be mined from the transformed data.

33. Another presentation from the Universitat Rovira i Virgili concerned anonymization of trajectory data that presents two methods for privacy-preserving trajectory publication based on microaggregation. The SwapLocations method works over a cluster of trajectories. The ReachLocations method takes reachability constraints into account.

34. During the discussion, the following issues were raised:

- The problem of loss of outliers in data needs to be addressed
- Data files take about 4-5 hours to create in the University of Michigan approach. The more variables the longer it takes
- How to determine thresholds for “high” correlation in discrimination analysis
- The sequence of events in a person’s life can be compared to trajectories

- In some countries it is difficult to get researchers to accept data distortion
- Ideas presented in trajectory data could be used for longitudinal business data sets

Topic (vii): International projects, groups and forum dealing with data access, release and related methodologies

Session Organizers/Discussants: Luisa Franconi, Italy and Michelle Simard, Canada

Documentation: Invited papers by UNECE, Germany, France/Germany/Netherlands; supporting papers by Italy (2) and Spain

35. This session presented various initiatives and projects being undertaken that deal with confidentiality aspects of data access and release, and ways to improve the services offered to users. The aim was to give participants information on developments in these areas and to improve the exchange of views and experiences in these fields.

36. The UNECE presented the current work programme on statistical confidentiality, approved by the Conference of European Statisticians. This work programme includes an update of case studies for the 2007 publication “Managing Statistical Confidentiality and Microdata Access: Principles and Guidelines of Good Practice” and a review of the implementation of the 2009 “Principles and Guidelines on Confidentiality Aspects of Data Integration Undertaken for Statistical or Related Research Purposes”. Proposals for future outputs are also sought, and could include new topics for principles and guidelines, or an inventory of disclosure control software tools.

37. Germany presented a new “ESSnet” project on Decentralized and Remote Access to Confidential Data in the European Statistical System. This builds on a previous project, and the aim is to implement remote access to confidential data in Eurostat through safe centres in national statistical organizations. The pilot implementation will take place within two years. Legal, administrative, technical and cost requirements have to be defined and pre-tested. If the pilot is successful, it could be extended.

38. The current status of the Data without Boundaries (DwB) project was presented. This project will link to the existing series of workshops on data access. The main goals are to exchange knowledge, define standards and find solutions for best practice. Cutting edge research increasingly requires access to highly detailed and sensitive data from different countries. This issue is now at the top of the European agenda. The main goal of DwB is to have an integrated model on data access where the best solutions are available irrespective of national boundaries and are flexible enough to fit national arrangements.

39. Other presentations included:

- Farm Structure Survey: Considerations on the release of a European Microdata File for research purposes (Italy) which indicates the main issues to be taken into account in the development and implementation of a methodology for statistical disclosure control applied to a specific survey.
- The Basque Statistics Office (Eustat) presented their ongoing work in the elaboration of a public-access microdata facility for researchers.

40. The following points were made during the discussion:

- The possibility for organizations to join the work of the DwB project (but not the formal consortium).
- Governance structures are necessary to ensure that the outputs of research projects can be implemented in practice.

Topic (viii): Trans-border access to microdata

Session Organizer/Discussants: Aleksandra Bujnowska and Rainer Muthmann, Eurostat

Documentation: Invited papers by Canada, United Kingdom and University of Minnesota/Open Data Foundation/Australia; supporting papers by France, Eurostat, OECD and University of Minnesota, Centre d'Estudis Demogràfics

41. When statistical organizations offer access to microdata to researchers, this is usually at the national level. There are now an increasing number of international projects and initiatives aiming at facilitating the trans-border access to microdata. This session presented the various means of trans-border access with an emphasis on legal and organizational aspects.

42. The representative of Canada presented the work of the “Paris Group” in advancing cross-border access to microdata. The biggest challenges for national statistical organizations when it comes to providing access to data beyond their own borders centre around legislation, costs, technical expertise and other risks associated with all microdata access, such as ensuring data confidentiality and maintaining good respondent relations. The Group will continue to seek solutions by working together with projects such as Data without Boundaries (DwB) and the DDI Alliance.

43. The United Kingdom presented proposals within the DwB project for a researcher accreditation model for access to official microdata. It is planned to develop a simple set of criteria for accreditation of researchers who require access to confidential data from multiple countries – a kind of Schengen agreement for data access. In order to reach an accreditation standard, DwB will analyse current practice, identify the essential common features, propose a standard and encourage adoption and implementation, in particular where DwB is joining research infrastructures across boundaries.

44. The University of Minnesota/Open Data Foundation/Australian Bureau of Statistics presented the technical and semantic approaches to dealing with legal and organizational considerations of trans-national microdata access through enforcing clear communication. DDI and SDMX can be used to support the capture of metadata along the statistical process, as defined by the Generic Statistical Business Process Model. DDI is becoming an increasingly popular standard for managing secure microdata in all types of environments, both within the domain of official statistics and for researchers, supplemented by SDMX in some official statistical organizations.

45. Additional presentations included:

- Legal aspects of transnational access of French confidential data (France);
- The future of access to EU confidential data for scientific purposes (Eurostat);
- Integrated cross-border access to microdata in OECD countries (OECD)
- The IPUMS-IECM partnership where a single licence agreement opens access to population census microdata for more than 60 countries to researchers world-wide (University of Minnesota, Centre d'Estudis Demogràfics).

46. During the discussion, the following points were made:

- Development of a “Schengen” agreement for data access was broadly supported.
- All these initiatives to improve data access are very important. However, there is a risk of overlap. It is important to coordinate these projects to avoid this and to achieve “collaboration rather than competition”.
- There is close cooperation between the various initiatives to ensure that their scope does not overlap.
- There is a need to share microdata according to standard approaches and an agreed infrastructure.
- Technology should not be a barrier to microdata access.
- Whilst accreditation should be at the level of the researcher, the institution to which they are affiliated will have to be taken into account and the increasingly collaborative approach to research must be taken into account.

Topic (ix): Statistical Disclosure Limitation for table and analysis servers: how to make outputs of modern data access infrastructures safe

Session Organizers/Discussants: Sarah Giessing and Tim Hochgürtel, Germany

Documentation: Invited papers by Germany (2) and Universitat Politècnica de Catalunya; supporting papers by United States (2) Germany and Australia

47. In recent years, many statistical agencies have started to set up infrastructure to provide secure remote access to confidential microdata for scientists and also to make production of tabular data more convenient by setting up facilities to generate user requested tables. Ideally, these new infrastructures should be accompanied by suitable methodologies to ensure non-disclosiveness of output generated through the system in a convenient automated way. This session considered established as well as new and emerging methods for this purpose, and examined the infrastructures that are being developed.

48. The presentation by Germany on Morpheus described a novel approach to providing remote access to microdata of official statistics. Researchers work on anonymized microdata files with standard statistical software packages and get their results back in real time. A measure of information loss is provided on the level of each individual result. Manual disclosure control of the output is not necessary.

49. Another paper by Germany presented a new method of post-tabular stochastic noise to protect skewed business data. The method modifies every value of a magnitude table. Sensitive cells receive more perturbation than non-sensitive cells. The paper discussed different possibilities to restore the additivity of the perturbed magnitude and proposed a flexible rounding method. The method has been tested on tabular data of German business tax statistics with encouraging results.

50. Controlled tabular adjustment is a perturbation method based on optimization techniques. The paper presented by Universitat Politècnica de Catalunya described current and planned research to speed up the method to make it useable for large applications, and also to make it useable in an online-table generation system.

51. A paper by Evans, Zayatz, Slanta, in 1998 proposed multiplicative noise on the microdata-level to protect enterprise tabular data (EZS method). The United States presentation discussed how to adapt and improve the method for the context of tables released from the US Quarterly Census of Employment and Wages (QCEW).

52. German statistical law considers two different concepts of disclosure risk. An evaluation of primary disclosure risk assessment methods for both concepts was presented. The question is, if application of the relaxed concept which would be legal when making data accessible to approved researchers, can it reduce the amount of manual work to check tabular data outputs.

53. Another paper by the United States provided an overview of various modalities to make sensitive data available to approved researchers. It also gave an overview of National Agricultural Statistics Service (NASS) and its approaches to data accessibility and maintaining confidentiality, while providing maximum data utility to its customers.

54. The Australian Bureau of Statistics (ABS) is developing a remote access service for tabulations of count data, ensuring the confidentiality of individual respondent information. The methodology is robust enough to protect a wide range of queries and analysis against risks from different kinds of attack. A Survey Table Builder supports requests for tabulation output using post-tabular noise for fully automated disclosure control. The Analysis Server supports submission of regression models etc. and includes a query control and output perturbation methods for fully-automated disclosure control.

55. The discussion raised the following points:

- The issue of information loss as a result of perturbation, and how to manage this.

- A standard approach to evaluate disclosure control methods would be a useful output from this group.
- Rounding can help mask sensitive data in some cases.
- Providing data metrics (means, variances etc.) can be a problem for skewed data sets, particularly for business data. The same metrics should be available to all users.
- Data quality varies according to the user and the use. When perturbation is applied, it is difficult to determine the optimum balance between confidentiality and quality. More work is needed on determining statistically meaningful measures to address this issue.
