

Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality

(Skopje, on 20-22 September 2017)

INFORMATION NOTICE No. 1

This work session will be held from *20 to 22 September 2017* in *Skopje*.

I. PURPOSE AND TARGET AUDIENCE OF THE WORKSHOP

1. The main objective of this meeting is to facilitate the exchange of experience and research findings, and to identify the best practices in dealing with the technical issues related to statistical data confidentiality in national statistical offices. The meeting is primarily intended for experts from national and international statistical offices, as well as invited academics or other researchers dealing with statistical data confidentiality issues.
2. Statistical data confidentiality is an important issue that influences public perception of official statistics. Statistical confidentiality, and protection of respondents' privacy, is included in the Fundamental Principles of Official Statistics, which were endorsed by the United Nations General Assembly in 2014.

II. AGENDA OF THE WORKSHOP

3. The programme of the work session will consist of the following substantive topics:
 - (i) Microdata and output protection
 - (ii) Access to microdata
 - (iii) Confidentiality of big data and special types of data
 - (iv) Utility and disclosure risk in anonymised data
 - (v) Methods and tools for tabular data protection
 - (vi) Census 2021 - Confidentiality issues
 - (vii) Users' perspective on statistical data confidentiality methods and tools
 - (viii) Public use files / open data - Future availability of information
4. Detailed explanatory notes on the nature and expected outcomes of these topics are provided in Section VI of this Information Notice.

III. PARTICIPATION AND ACCREDITATION

5. Representatives of all Member States of the United Nations and of interested intergovernmental organizations are welcome to participate in the workshop. Participants representing non-governmental organizations in a consultative status with the United Nations Economic and Social Council may also attend. All participants must be accredited by the competent authorities of their country or international organization.
6. All participants attending the seminar are requested to have a valid passport and, if required, a visa. Applications for visas should be made as soon as possible, with a reference to the Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality. Please contact the local hosts (contact details provided in Section VIII of this Information Notice) if you require a letter to facilitate obtaining a visa, or have any questions about this issue.

7. Statistical offices and international organizations should express their interest by **Friday 31 March 2017** if they wish to:

1. Participate in this event; and/or
2. Submit abstracts for one or more of the topics mentioned above.

This should be done by filling in the online form on the following website:

<http://www1.unece.org/stat/platform/display/SDC2017/Your+Participation+and+Contribution>

8. Participants should also separately register their security information online at the site <https://www2.unece.org/uncdb/app/ext/meeting-registration?id=RyLcA2> by **Friday 18 August 2017**, in order to be permitted access to the workshop venue. Please note that the registration process must be fully completed before a request is generated. (When this happens, a message will be displayed that says “*You successfully submitted your application. The application is pending approval.*”)

IV. GRANTS FOR PARTICIPATION

9. Whilst the UNECE Secretariat is unable to fund participation in this meeting, the UNESCO Chair in Data Privacy (<http://unescoprivacychair.urv.cat>) sponsors a limited number of travel grants for contributors and delegates from transition countries. For further information please directly contact unescoprivacychair@urv.cat (for attention of Romina Russo).

V. DOCUMENTATION, METHODS OF WORK AND OFFICIAL LANGUAGES

10. The official language of the workshop is English. Any contributions should, therefore, be submitted in English only. Contributions should normally take the form of papers, with an accompanying presentation, though other forms of contribution may be proposed. *Full guidelines for contributions will be given following the acceptance of abstracts.* At this stage, contributors should indicate the preferred format of their contributions if other than a paper and presentation - but please note that due to the nature of the work session, it may not be possible to allocate time to all proposed contributions.

11. The following deadlines and requirements apply:

- A short abstract of the proposed contribution should be submitted as soon as possible and by **Friday 31 March 2017**, by filling in details on the following website:
<http://www1.unece.org/stat/platform/display/SDC2017/Your+Participation+and+Contribution>
Information about the selection of contributions for the workshop, and on guidelines on formatting will be sent to authors using the contact details they enter into the online tool.
- Any written papers must be **uploaded to the online tool** by **Friday 30 June 2017** at the latest. (Instructions on how to upload papers will be supplied upon acceptance of abstracts.)
- Any presentation slides, videos or other electronic materials should also be **uploaded to the online tool** by **Friday 1 September 2017** at the latest.

12. Papers will be made available on the UNECE website at the following location: <http://www.unece.org/stats/documents/2017.09.confidentiality.html>. Presentations will be added to this web site after the workshop. Presentations will not be made available to delegates before the workshop.

13. Participants are encouraged to download the papers from the website and, where feasible, to use electronic devices to read papers in order to minimise paper use. Documents posted on the website before the workshop will **not** be distributed in the conference room.

VI. EXPLANATORY NOTES TO THE AGENDA

Topic (i): Microdata and output protection

Session Organiser: Krish Muralidhar, University of Oklahoma

14. Statistical agencies release many types of data, including microdata (record level data), output from

queries/statistical analysis performed on data, tabular data, and others. Research on microdata protection and output protection has been rich and varied. This stream of research has been, and continues to be, important. With a few exceptions however, research in this area has been performed with a *micro* view of the data (release of a single data set), which was appropriate in the past.

15. The modern data user expects statistical agencies to provide data similar to that provided by large (non-governmental) organizations such as Google. They expect statistical agencies to provide accurate data that is easily accessible and provides analytical flexibility. Statistical agencies have responded with multiple access modes, data releases for different users and other changes. In this new paradigm, it is important that microdata/output protection research takes a *macro* view of the issues relating multiple access modes, data releases, and analyses.

16. We invite research papers relating to both the *micro* and *macro* perspectives:

- (i) The development of new methods or techniques, comparison of methods/techniques, assessment of disclosure risk and/or information loss, user experience, and other issues related to microdata and/or output protection.
- (ii) The impact of different types of data release or access, such as:
 - a. New methods capable of providing access to data defined by the needs of the user;
 - b. Global assessment of disclosure risk/information loss across all types of access;
 - c. Consistency of output across different types of access/release; and
 - d. Other research related to the macro view of the data released by statistical agencies.

Topic (ii): Access to microdata

Session Organisers: Aleksandra Bujnowska, Eurostat & Annu Cabrera, Statistics Finland

17. Several statistical agencies provide access to their microdata (e.g. for scientific purposes), and there are different modes of access, such as release of anonymised microdata files, onsite access (safe centres), remote access systems, remote program execution and remote analysis servers. A good mode of access should maintain statistical confidentiality and privacy of respondents.

18. The aim of this topic is to discuss different approaches to microdata access. We invite papers discussing currently existing best practices, as well as modern modes of access. Contributions to this topic could address both national solutions as well as international trans-border access.

Topic (iii): Confidentiality of big data and special types of data

Session Organisers: Peter-Paul de Wolf, Statistics Netherlands & Josep Domingo-Ferrer, Universitat Rovira i Virgili

19. Whilst statistical institutes have traditionally dealt with survey sample data, they now also have to deal with other types of data, including administrative data, unstructured text, event-based data, and detailed geo-referenced data.

20. In addition to this, there are other types of data, which are being handled by private data controllers, and which may well end up appearing in official statistics, such as social network data, multimedia, streaming data, genomic data, etc. When all of these data types come in large volumes from several sources they are referred to as big data.

21. Making use of big data (and special types of data) poses new questions about privacy and confidentiality. For example, to what extent are the classical methods and theories on statistical disclosure control still applicable? How could we best visualise statistical data confidentiality when publishing big data or special types of data?

22. In this session we welcome contributions on:

- (i) General statistical data confidentiality issues concerning big data, including complex data, such as:
 - a. Administrative data
 - b. Unstructured text
 - c. Event-based data
 - d. (Detailed) geo-referenced data
 - e. Any other “new” type of data (social media, multimedia, genomic, streaming, etc.)

- (ii) Visualising statistical data confidentiality measures taken on (special types of) data (e.g., geo-referenced data).

Topic (iv): Utility and disclosure risk in anonymised data

Session Organisers: Josep Domingo-Ferrer, Universitat Rovira i Virgili

23. The aims of this session are to discuss research on, and practice aspects of, balancing privacy guarantees against utility guarantees. This session should bridge the computer science viewpoint (focused on privacy guarantees) and the official statistics viewpoint (focused on utility guarantees).

24. Contributions to this session should address the following topics in a scientifically rigorous manner, while keeping a practical orientation:

- (i) Formal privacy models;
- (ii) Risk and utility measures;
- (iii) Transparency, verifiability and intervenability of anonymisation; and/or
- (iv) Disclosure scenarios and attacks.

Topic (v): Methods and tools for tabular data protection

Session Organiser: Sarah Giessing, Destatis

25. Tabular data remains the most common type of data released by statistical agencies, despite the replacement of traditional presentations (in form of printed volumes) with more modern ways of providing access to aggregate data. Special challenges arise when disclosure control processes have to deal with large sets of multiple large tables or hypercubes, or when protection is required to be provided “on the fly” for customised tables that users request online.

26. In this session we invite contributions with a methodological focus, as well as papers on new implementations or recent improvements of existing algorithms and software tools. We encourage papers discussing improvement of traditional methods (like cell suppression), as well as papers on more recent approaches and perturbative techniques. Topics of interest include:

- (i) Disclosure limitation methods such as suppression, perturbation or masking of tabulations or original microdata through random noise, controlled adjustment, etc.; and
- (ii) Disclosure risk considerations in tabular data, global and local measures for information loss in the protected data, or issues of consistency.

Topic (vi): Census 2021 - Confidentiality issues

Session Organiser: Eric Schulte Nordholt, Statistics Netherlands

27. All countries are expected to conduct at least one population and housing census every ten years, and most countries will conduct their next census in 2020 or 2021. (European Union countries will use 2021 as the reference year.)

28. The next census round will result in the production of detailed publications and huge sets of detailed tables. Given the increased privacy concerns, it is important that they are properly protected against disclosure. In addition, it is foreseen that microdata for research purposes will be provided by a number of countries.

29. As more and more information is available on internet, it is becoming increasingly difficult to guarantee the safety of the census tables and census microdata for researchers. In many countries and international organisations, projects to study these problems have been started.

30. Contributions are invited addressing issues around tackling the privacy concerns of the output of the coming census round.

Topic (vii): Users' perspective on statistical data confidentiality methods and tools

Session Organiser: Michelle Simard, Statistics Canada

31. Releasing information requires a constant balance between protecting the data and producing useful

information. On the one hand, statistical institutes have to protect the confidentiality of their products, whilst on the other hand, users want more and more data (and some even want them on their laptops). Through various means and efforts, statistical institutes and public institutions have developed programs, systems and modes of access to increase the use of their data, but it is still often considered not enough from the perspective of users and analysts.

32. Contributions are welcome on perspectives of various users (of individuals, and of academics in both the public and private sectors), who access various statistical products (microdata files, tables, online releases or synthetic data files) with regard to the use of disclosure control methods and tools. We will also welcome contributions from statistical institutes that have established and developed client relationship services, either through a regular forum or on-site data centres open to “external clients”.

Topic (viii): Public use files / open data - Future availability of information

Session Organiser: Eric Schulte Nordholt, Statistics Netherlands

33. This panel session will give an opportunity to hear about emerging trends in this area from a number of viewpoints.

VII. VENUE

34. The workshop will be held at:

Ministry of Foreign Affairs
Boulevard “Philip II of Macedon“ Nr. 7
1000 Skopje

A second information notice will be issued and made available on the workshop website nearer to the dates of this workshop, giving details about the location of the workshop, along with additional logistical information.

VIII. FURTHER INFORMATION

35. For further information, please submit queries using the online form described in paragraph 7 of this information note. Alternatively, you may also contact the following organisers:

Local host:

Mr. Kiril Trajanoski
Tel: +389 2 3295 670, email: kiril.trajanoski@stat.gov.mk

Mr. Aleksandar Eftimovski
Tel: +389 2 3295 668, email: aleksandar.eftimovski@stat.gov.mk

UNECE:

Mr. Christopher Jones,
Tel: +41 22 917 2413, email: christopher.jones@unece.org