I. PURPOSE AND AGENDA

1. This work session will be organized as part of the Conference of European Statisticians’ work programme for 2019, within the context of the High Level Group for the Modernisation of Official Statistics. It will take place at the Statistics Netherlands premises in The Hague, Netherlands from 29-31 October 2019.

2. The main objectives of the meeting are to facilitate the exchange of experience and identify the best practices in dealing with technical issues related to statistical data confidentiality in statistical offices. The meeting is primarily intended for experts from national and international statistical offices as well as invited academics dealing with statistical disclosure limitation.

3. Statistical disclosure limitation is an important issue influencing public perception of official statistics. Statistical confidentiality and protection of respondents’ privacy is included in the Fundamental Principles of Official Statistics.

4. The programme of the work session will consist of the following substantive topics:
   - Access to microdata
   - Tabular data
   - Risk assessment: Privacy, confidentiality, disclosure
   - Emerging issues
   - The framework for confidentiality
   - Software tools for statistical data confidentiality
   - Confidentiality issues of the Census 2020/2021 Round

II. CALL FOR CONTRIBUTIONS

5. Participants are strongly encouraged to consider submitting an abstract that summarises the content of their proposed contribution. These should be covering one or more of the topics of the programme of the work session. For further details on the coverage of the topics, refer to the annex at the end of this note. Any contributions should be submitted in English only.

6. A short abstract of the proposed contribution should be submitted by email to the UNECE Secretariat, tetyana.kolomiyets@un.org and taeke.gjaltema@un.org, as soon as possible and by Tuesday 30 April 2019 at the latest.

7. The steering committee will notify in due time whether the submission is accepted or not and might request changes. Contributions should normally consist of a paper, plus an accompanying presentation. Other forms of contribution may be proposed. Please note that it may not be possible to allocate time to all proposed contributions.
8. Any written papers must be provided by **Wednesday 4 September 2019** at the latest. A link will be sent to the authors where documents can be uploaded.

9. Any presentation slides, videos or other electronic materials should be provided by **Tuesday 15 October 2019** at the latest. Any equipment required for practical demonstrations must be provided by the participants. A link will be sent to the authors where presentations can be uploaded.

### III. REGISTRATION

10. Participants should register their security information online by **Wednesday 4 September 2019** at the site [https://statswiki.unece.org/x/3gNqDg](https://statswiki.unece.org/x/3gNqDg) in order to be permitted access to the work session 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.”).

11. Representatives of all Member States of the United Nations and of interested intergovernmental organizations are welcome to participate in the work session. 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.

12. All participants attending the seminar are requested to have a valid passport and, if required, a visa. Applications for visas to the Netherlands should be made as soon as possible to the Embassy of the Netherlands in the country in which the participant resides, with a reference to the Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality. A letter to facilitate obtaining a visa can be requested from Martin van Sebille (mt.vansebille@cbs.nl) from Statistics Netherlands.

13. Participants and/or their offices are requested to make their own travel arrangements and hotel reservations. The UNECE Secretariat regrets not being able to offer any financial assistance regarding travel and accommodation arrangements.

### IV. GRANTS FOR PARTICIPATION

14. Whilst the UNECE Secretariat is unable to fund participation in this meeting, the UNESCO Chair in Data Privacy ([http://unescoprivacychair.urv.cat](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 LANGUAGE

15. The official language of the work session is English. No interpretation or translation will be provided.

16. The work session will consist of presentations, small group discussions to exchange experiences and to identify lessons learned, and interactive discussions to identify future work in the area of Data Confidentiality. Participants are expected to actively contribute to the discussions.

17. Papers and presentations will be made available in pdf format on the work sessions website (http://www.unece.org/index.php?id=50848).

18. 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 work session will not be distributed in the conference room.

VI. VENUE

19. The work session will be held at:

Central Bureau of Statistics/
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

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

VII. FURTHER INFORMATION

21. For further information you may contact the following organizers:

UNECE:
Ms. Tetyana Kolomiyets, tel: +41 22 917 41 50, email: tetyana.kolomivets@un.org
Mr. Taeke Gjaltema, tel: +41 22 917 1272, email: taeke.gjaltema@un.org

Eurostat:
Ms. Aleksandra Bujnowska, email: aleksandra.bujnowska@ec.europa.eu
VIII. ANNEX: EXPLANATORY NOTES TO THE AGENDA

Access to microdata

Organizers: Aleksandra Bujnowska (Eurostat), Janika Tarkoma (Statistics Finland), Steven Thomas (Statistics Canada)

22. Several statistical agencies provide access to their microdata (e.g. for scientific purposes). There are different modes of access, such as release of (partially or fully) anonymised microdata files, onsite access (safe centres), remote access systems, remote program execution and remote analysis servers. A good mode of access should balance statistical confidentiality with offering useful information to the user community.

23. The aim of this topic is to discuss different approaches to microdata access. We invite papers discussing current practices, as well as innovative modes of access. Contributions to this topic could address both national solutions as well as international trans-border access. Papers can be both from the perspective of the organization that provides access or from the perspective of the users of microdata.

24. Contributions to this session could include following topics:

(i) Case studies of remote access to microdata - virtual data labs
(ii) Real-time confidentiality (solutions allowing users to query the data with confidentiality applied instantly)
(iii) Optimisation of output checking (how to ensure in an efficient way that the results of microdata analysis are safe to be published)
(iv) Synthetic data creation
(v) Data user perspectives
(vi) Efficient management of disclosure risk

DEADLINES

30 April 2019  Abstract or proposal for intended contribution
4 September 2019  Registration
4 September 2019  Paper
15 October 2019  Presentation or other material to be presented
29-31 October 2019  Work session
Tabular data

Organizer: Sarah Giessing (Destatis)

25. While the traditional formats of presenting tabular data in printed volumes get replaced more and more by various electronic formats, the challenges connected to statistical disclosure limitation of statistical data aggregates that used to be presented in the traditional printed volumes are not ceasing. Special challenges arise when protection is required for tables generated on user demand, or when protection even has to be provided “on the fly” for products requested through on-demand tabulation or more general tools for secure statistical computing. Moreover, as more and more different methods for tabular data protection arise in the literature, evaluation criteria are needed to guide decision processes.

26. 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.
   (ii) Disclosure risk considerations in tabular data, global and local measures for information loss in the protected data, or issues of consistency
   (iii) The user’s perspective
   (iv) Increasing risk appetite and its implications
   (v) Evaluation criteria for tabular disclosure limitation methods
   (vi) Case studies of on-demand tabulation, issues arising when setting parameters (like for random noise

---

Risk assessment: Privacy, confidentiality, disclosure

Organizers: Josep Domingo (Universitat Rovira i Virgili), Krish Muralidhar (University of Oklahoma)

27. Any data release, whether it is summaries, tables, or microdata, inherently involves the risk that information about the respondents in the data set may be leaked. How this leak is quantified represents the risk assessment when releasing data/tables/summaries. Data release risk assessment has received considerable attention in the literature for over half a century. A number of measures have been proposed: privacy measures, confidentiality measures, disclosure measures (identity and value), etc. Yet, there remains considerable confusion about what these measures mean and/or how they relate to one another.

28. In this session, we welcome papers that address:
   (i) The relationship between the different measures of risk and/or an exposition of these measures
   (ii) Development of new measures that go beyond existing measures
   (iii) The trade-off between risk and utility in data release
   (iv) Record linkage for disclosure risk assessment
   (v) Any other topic related to risk and utility measures
Emerging issues

Organizer: Josep Domingo (Universitat Rovira i Virgili)

29. While 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.

30. 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.

31. 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?

32. In this topic we welcome contributions on:
   (i) Statistical disclosure control of new types of data, including complex data, arising in big data environments:
       • Administrative data
       • Unstructured text
       • Event-based data
       • (Detailed) geo-referenced data
       • Mobility data
       • Streaming data
       • Any other “new” type of data (social media, multimedia, genomic, etc.)
   (ii) Merging several anonymised sources and privacy-preserving record linkage
   (iii) Visualising statistical data confidentiality measures taken on the above new types of data

The framework for confidentiality

Organizers: Aleksandra Bujnowska (Eurostat), Janika Tarkoma (Statistics Finland)

33. The General Data Protection Regulation (2016/679) (GDPR) applies in the EU since 25 May 2018. It has an effect on organizations handling all kind of personal information. Statistics has its own national and international legislation and the GDPR has also to be taken into account. The GDPR states that organizations should be able to demonstrate compliance of 'lawfulness, fairness and transparency' when processing the personal data. It would be of great interest to learn from the experiences of different organizations: what kinds of changes are needed to comply with the new requirements.

34. The GDPR has also raised the general awareness on the privacy issues. Both media and individuals are more interested in the use of personal data. Organizations should have plans on how to proceed in the unfortunate event of breach of confidentiality. Since the GDPR applies, organizations have to inform their data providers about the breach in case there is a risk for rights and freedoms of the data provider. Usually this means that the breach is also discussed in public. We are interested to learn about the different approaches to crisis management in case of breach of confidentiality.

35. Contributions to this session could include following topics:
   (i) Good practices on complying with new legal requirements:
• Awareness raising and training
• Re-organization of administrative processes
• Technical issues including security
• Documentation of the data processing workflow

(ii) Crisis management:
• Detection of confidentiality breach
• Procedures and guidelines
• Communication on the event
• Training

36. We also invite contributions not linked to GDPR but covering general confidentiality frameworks, i.e. technical (security), organizational and administrative aspects.

Software tools for statistical data confidentiality
Organizer: Peter-Paul de Wolf (Statistics Netherlands)

37. The final stage in the process concerning statistical data confidentiality is the implementation of statistical disclosure control (SDC) methods. Either because of the complexity of the methods or the volume of the data to be protected, it is preferred to have dedicated software to apply SDC methods. Moreover, since it concerns software to protect privacy aspects of collected data it is almost essential that this software is Open Source.

38. In this section we will focus on Free and Open Source software and welcome contributions on:
   (i) Using existing “standard” SDC tools:
   • Tests of functionalities
   • How to embed the tools in production processes
   • Best practices and available workarounds
   (ii) New SDC tools, either prototype or production ready
   (iii) Organizational aspects of using SDC tools

39. We would like to organize part of this session as a "walk around" session where presenters can give a live demonstration of their (prototype) software and where participants can discuss the features they miss but have always wanted to be available in SDC tools software.

Confidentiality issues of the Census 2020/2021 Round
Organizer: Eric Schulte Nordholt (Statistics Netherlands)

40. 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.

41. The next census round will result in the production of 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.

42. 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 organizations, projects to study these problems have been conducted.

43. Not only well-known techniques (e.g. global recodes and local suppressions), but also newer techniques
(e.g. targeted record swapping and the cell key method) will be used to protect the huge sets of census tables while minimising information loss.

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