

## Information Notice No.1

### I. PURPOSE AND TARGET AUDIENCE OF THE MEETING

1. The 2021 joint UNECE/Eurostat Expert meeting on Statistical Data Confidentiality will be hosted by Statistics Poland in coordination with Poznań University of Economics and Business. It will take place in Poznań, Poland on 1 to 3 December 2021, starting at 09:30 am on Wednesday 1 December, and ending by 12:00 noon on Friday 3 December.

---

The planning of this meeting envisages an improved public health situation to allow this meeting to take place in-person in Poland. However, uncertainties remain about this due to the COVID-19 pandemic. Therefore, by early September, a decision will be made taking into account the overall situation in the UNECE region. If an in-person meeting is not possible, it will be changed into an online (only) meeting during the same week.

---

2. This meeting will be organized as part of the Conference of European Statisticians' work programme for 2021, within the context of the High-Level Group for the Modernisation of Official Statistics.

3. The objective of this expert meeting is to identify innovative approaches and best practices in statistical data confidentiality, and to provide a platform for practitioners to exchange experiences and foster collaboration in this area. In addition to the more traditional presentations, the agenda will include target-driven small group discussions and activities to identify best practices and new opportunities. Delegates will be asked to contribute to the development of internationally-coordinated work in the field of statistical data confidentiality. The meeting is primarily intended for experts from national and international statistical offices as well as invited academics dealing with statistical disclosure limitation.

### II. AGENDA OF THE MEETING

4. The programme of this meeting will consist of the following substantive topics. (Further details about these topics can be found in the annex at the end of this Information Notice.):

- Access to microdata;
- Microdata protection;
- Tabular data;
- Risk assessment: Privacy, confidentiality, and disclosure;
- Other emerging issues;
- Software tools for statistical data confidentiality; and
- Communication of statistical disclosure control methods.

### III. PARTICIPATION AND ACCREDITATION

5. Representatives of all Member States of the United Nations and of interested intergovernmental organizations are welcome at this meeting. 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 meeting are requested to have a valid passport and, if required, a visa. Applications for visas should be made as soon as possible to the Embassy of Poland in the country in which the participant resides, with a reference to the 2021 joint UNECE/Eurostat Expert meeting on Statistical Data Confidentiality. A letter to facilitate obtaining a visa can be requested from the local contacts in Poznań (contact details below).

7. Participants should register online by 1 October 2021 by following the link:

<https://forms.office.com/r/sLEvz1PaDS>

8. 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. CALL FOR PAPERS, METHODS OF WORK AND OFFICIAL LANGUAGES

9. Participants are strongly encouraged to consider submitting an abstract, summarising the content of their proposed contribution. These should cover one or more of the topics of the meeting programme. The official language of the meeting will be in English, and therefore all contributions should be submitted in English only. No translation or interpretation during the meeting will be provided.

10. Regarding abstracts received, the steering committee will send notification in due time about 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. Information about the selection of contributions for the meeting, guidelines on formatting, and means of submission will be sent to authors by email. Please note that due to the nature of the meeting, it may not be possible to allocate time to all proposed contributions.

11. The following **deadlines** and requirements apply:

(i) A short abstract of the proposed contribution should be submitted via the following webpage as soon as possible and by **Friday 18 June 2021** at the latest:

<https://forms.office.com/r/sLEvz1PaDS>

(ii) Any written papers must be supplied by **Friday 1 October 2021** at the latest. A link will be sent to the authors where documents can be uploaded.

(iii) Any presentation slides, videos or other electronic materials should be supplied by **Friday 5 November** 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.

12. Papers will be made available online for this meeting before the meeting via the following website: <https://unece.org/statistics/events/SDC2021>. Presentations will also be added to that site after the meeting. Presentations will not be made available to delegates before the meeting, unless requested by the presenters.

13. Participants are encouraged to download the papers via the webpage and, where feasible, to use electronic devices to read papers in order to minimise paper use. These documents will not be distributed in the conference room.

## V. VENUE

14. Provided that the public health circumstances allow, the meeting will take place in Poznań, Poland at:

Poznań University of Economics and Business  
al. Niepodległości 10  
61-875 Poznań,  
Poland  
<https://ue.poznan.pl/en/>

15. A second information notice with practical information will be shared with registered participants and on the meeting website in due time before the meeting.

## VI. FURTHER INFORMATION

16. For further information you may contact the following organisers:

*Secretariat of the United Nations Economic Commission for Europe:*

Mr. Christopher Jones, email: [JonesC@un.org](mailto:JonesC@un.org)

Mr. Taeke Gjaltema, email: [Taeke.Gjaltema@un.org](mailto:Taeke.Gjaltema@un.org)

*Local Contacts in Poznań:*

Mr. Tomasz Klimanek, email: [T.Klimanek@stat.gov.pl](mailto:T.Klimanek@stat.gov.pl) and [Tomasz.Klimanek@ue.poznan.pl](mailto:Tomasz.Klimanek@ue.poznan.pl)

Ms. Grażyna Dehnel, email: [G.Dehnel@stat.gov.pl](mailto:G.Dehnel@stat.gov.pl) and [Grazyna.Dehnel@ue.poznan.pl](mailto:Grazyna.Dehnel@ue.poznan.pl)

### DEADLINES

18 June 2021	<b>Abstract</b> or proposal for intended contribution
1 October 2021	<b>Registration</b>
1 October 2021	<b>Paper</b> or detailed abstract if no paper is provided
5 November 2021	<b>Presentation</b> or other material to be presented
1 to 3 December 2021	Meeting

---

## VII. ANNEX: EXPLANATORY NOTES TO THE AGENDA

### Access to microdata

Topic Leads: Aleksandra Bujnowska, Eurostat and Eric Schulte Nordholt, Statistics Netherlands

*Description:*

17. The aim of this topic is to discuss different approaches to microdata access. Several statistical agencies provide access to their microdata (e.g. for scientific purposes). Different modes of microdata access exist, such as release of (partially or fully) anonymised microdata files, onsite access (safe centres), remote access systems, remote program execution and remote analysis servers. An appropriate access mode should balance statistical confidentiality with the usefulness of information offered to the user community.

18. 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.

19. Contributions to this topic could include:

- Case studies of remote access to microdata in virtual labs;
- Real-time confidentiality (solutions allowing users to query the data with confidentiality applied instantly, including querying systems and table builders);
- Optimisation and automatisation of output checking (how to ensure in an efficient means of making the results of microdata analysis safe to be published);
- Data user perspectives;
- Efficient management of disclosure risk;
- Planned innovative microdata access for the 2020 census; and/or
- Public use files for research.

### Microdata protection

Topic Leads: Josep Domingo-Ferrer, Universitat Rovira i Virgili and Krish Muralidhar, University of Oklahoma

*Description:*

20. This topic will cover different approaches to generate anonymized microdata, including masking and synthetic data generation. Papers on anonymizing emerging types of microdata, such as mobility data, geo-referenced data or even unstructured individual data are also welcome. Given the current decentralization trend in computing, decentralized approaches to microdata protection are also of particular interest.

21. This could include:

- Microdata protection;
- Mobility data;
- Geo-referencing/geospatial data;
- Synthetic data;
- Decentralized/local microdata protection; and/or
- Protection of census microdata.

## **Tabular data**

Topic Leads: Steven Thomas, Statistics Canada and Sarah Giessing, Destatis

### Description:

22. Pressure is growing in statistical institutes for modernisation of tabular data production systems, making them capable of producing not only carefully planned volumes of tabular data, but allowing for much more flexibility, even to the extent of tables generated on user demand. However, such changes in output production processes tend to hamper the efficiency of traditional Statistical Disclosure Control (SDC) practices based on cell suppression techniques, and call for changes in the disclosure control processes and techniques. Developing and implementing alternative techniques which do not rely on cell suppression imposes numerous challenges.

23. In the field of business statistics especially, cell suppression and related techniques are expected to remain the most relevant methods for disclosure control. Here the main challenges are to provide improved algorithms, able to handle larger tables, or complex constraints, and in the context of frequency tables to find solutions to overcome a lack of tools and methods to properly handle attribute disclosure risk.

24. Topics of interest include:

- SDC methods for online tabulation query systems;
- Disclosure control without cell suppression in practice;
- Utility and risk measurement;
- Improving state of the art algorithms for cell suppression, controlled tabular adjustment and rounding;
- Methods to handle attribute (group) disclosure risks in practice; and/or
- Protection of census tables.

## **Risk assessment: Privacy, confidentiality, and disclosure**

Topic Leads: Josep Domingo-Ferrer, Universitat Rovira i Virgili and Krish Muralidhar, University of Oklahoma

### Description:

25. Any data release, whether it consists of summaries, tables or microdata, inherently involves the risk that information about the respondents in the data set may be leaked. Risk assessment attempts to quantify the risk of leakage 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, including privacy measures, confidentiality measures, and disclosure measures (identity and value), for example. Yet, there remains considerable confusion about what these measures mean and/or how they relate to one another. On the other hand, performing SDC implies the loss of information (due to suppression or perturbation of data). There are a number of methods for measuring it, e.g. assessing differences in distributions, variations or correlation between data before and after SDC. Some of them can have slightly universal form. Ultimately, a balance between minimization of the risk of disclosure and minimization of the information loss due to SDC should be established.

26. For this topic, we welcome papers that address:

- Relationships between the different measures of risk, information loss and/or an exposition of these measures;
- Development of new measures that go beyond existing measures;
- Trade-off between risk and utility (opposite of the information loss) in data releases;
- Record linkage for disclosure risk assessment;

- Privacy models in official statistics: k-anonymity and extensions, differential privacy, etc.;
- Risk assessment procedures for the 2020 census; and/or
- Any other topic related to risk and utility measures.

### **Other emerging issues**

Topic Leads: Peter-Paul de Wolf, Statistics Netherlands and Janika Tarkoma, Statistics Finland

#### Description:

27. 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, network data, and detailed geo-referenced data. In addition to the heterogeneity of data types, some of these data can be privately held and there exists the need of designing access techniques able to make privacy-preserving the process of treating the data for statistical purposes. Alongside "traditional" methods and techniques for data protection there is also increasing interest in new(er) techniques, such as perturbative methods and privacy preserving techniques, for example.

28. The aim of this topic is to bring together new techniques and methodologies, applied to (old and) new types of data, and to discuss issues concerning the design, the implementation and the communication of these new techniques.

29. Contributions to this topic could include:

- Perturbative methods (use cases, opportunities and threats, communication);
- Privacy preserving techniques;
  - (i) Secure data sharing;
  - (ii) Secure multi party computation;
  - (iii) Privacy preserving record linkage;
- Privacy and machine learning;
- The relationship between input and output privacy;
- Anonymity in network data;
- SDC and visualisations;
  - (i) Visualising confidentiality measures;
  - (ii) Confidentialising visualisations; and/or
- Privacy and geo-referenced data.

### **Software tools for statistical data confidentiality**

Topic Lead: Peter-Paul de Wolf, Statistics Netherlands and Mr. Andrzej Młodak, Statistics Poland

#### Description:

30. The final stage in the process concerning statistical data confidentiality is the implementation of 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 such methods. Contributions can be either as a paper or in a "walk around" fashion (where a live demonstration can be given), allowing participants to discuss the features they would most desire to be available in SDC tools software.

31. Possible areas of contribution include:

- General purpose SDC software;
- New implementations (or even prototypes); and/or
- Automated output checking.

## **Communication of statistical disclosure control methods**

Topic Lead: Aleksandra Bujnowska, Eurostat and Janika Tarkoma, Statistics Finland

### Description:

32. Users receive a lot of information on how the data are compiled. There are metadata standards, quality reports and other forms of information to users. But statistical offices do not share widely the information on SDC methods applied on the data. This is considered risky. Modern SDC techniques like perturbative methods should be properly communicated to the users, as they have impact on the whole data, their accuracy and additivity. Having clear information on SDC technics allows users to properly understand and interpret the data.

33. For more advanced users, statistical offices develop special SDC trainings, which make participants aware about how to produce safe results when using various statistical tools and methods.

34. The objective of this topic is to discuss efficient SDC communication strategies for data providers (including respondents), data analysers (including researchers) and general users.

35. Contributions to this topic could include:

- How to communicate perturbative methods in census data protection;
  - Communicating confidentiality to providers (incl. respondents) and users of data (policy/public); and/or
  - Training on SDC.
-