
REPORT OF THE WORKSHOP

1. The Joint UNECE/Eurostat Workshop on Statistical Data Confidentiality was held in The Hague, from 29-31 October 2019. It was attended by participants from: Australia, Austria, Belgium, Bosnia and Herzegovina, Brazil, Bulgaria, Canada, Croatia, Finland, France, Germany, Hungary, Ireland, Israel, Italy, Japan, Kazakhstan, Mexico, Netherlands, North Macedonia, Norway, Portugal, Republic of Korea, Romania, Russian Federation, Serbia, Slovenia, Spain, Sweden, United Kingdom of Great Britain and Northern Ireland, the United States of America as well as by representatives Eurostat, FAO, OECD, The World Bank Group and UNECE. Participants from numerous universities and institutes attended the workshop at the invitation of the UNECE secretariat.
2. The workshop was hosted by Statistics Netherlands. It was organised under the responsibility of the High-Level Group for the Modernisation of Official Statistics. The Steering Committee consisted of Steven Thomos (Statistics Canada), Janika Tarkoma (Statistics Finland), Sarah Giessing (Destatis, Germany), Eric Schulte Nordholt and Peter-Paul de Wolf (Statistics Netherlands), Aleksandra Bujnowska (Eurostat), Krish Muralidhar (University of Oklahoma), and Josep Domingo (Universitat Rovira i Virgili). Peter-Paul de Wolf was the overall chair of the workshop.
3. The agenda included the following substantive topics:
 - (i) Access to microdata
 - a. Microdata access facilities
 - b. Microdata protection
 - (ii) Tabular data
 - (iii) Risk assessment: Privacy, confidentiality, disclosure
 - (iv) Emerging issues
 - (v) The framework for confidentiality
 - (vi) Software tools for statistical data confidentiality
 - (vii) Confidentiality issues of the Census 2020/2021 Round
4. Forty contributions were submitted. These were allocated and presented in seven substantive sessions. Through a consultative process, topics for future work were solicited. The workshop was concluded with an interactive plenary discussion on these and additional topics could be brought forward for inclusion. ay forward in Statistical Data Confidentiality. This led to the following list:

SDC techniques

- Geo-referencing/geospatial data
- Differential privacy applications
- Production of safe tables through non-standard software
- Zero-suppression in practice
- Improving SDC algorithms
- Practicality of techniques
- Utility/risk measurement
- Integration of SDC into NSI operating practices

- Group disclosure
- Integrated databases (administrative/big data)

Efficiency in micro data

- Sustainability/Managing demand
- Practical solutions (in each of the five safes)
- Synthetic microdata
- Benefits of provision of access to microdata
- Impact evaluation
- Quantification (5 safes)

Services and Tools

- Shared guidelines and handbooks facilitating data protection
- Safe rooms
- Output checking
- Certification
- Online tabulation, query systems
- Reproducibility
- SDC tools for data protection

Concepts

- Legal vs ethical issues
- Confidentiality and privacy
- Trust (with other organisations, and with public)
- Impact of GDPR – different country experiences (incl. response in non-EU countries)
- Utility measurement
- standardisation (methods/techniques/tools used)

SDC in specific fields

- Agriculture, Economic statistics, social statistics
- Big data, Census
- Sensitivity of specific concepts/variables

Follow up work

- Updating glossaries
- Inventory of handbooks/guidelines
- Projects teams working on specific issues
- Roadmap for microdata access maturity model

5. Further details on the outcome of the interactive session and group discussions can be found in Annex 2. An overview of requests for assistance and support offers, can be found in Annex 3. For reference, the timetable is included as Annex 4. All abstracts, papers, presentations, and other output from the workshop are available at the workshops wiki pages (<https://statswiki.unece.org/x/0oy9Dg>). The main documents are also available from the UNECE website (<http://www.unece.org/statistics>).

Annex 1 Agenda of the Workshop

A. Topic (i): Access to microdata

6. This topic was organized by Aleksandra Bujnowska (Eurostat), Janika Tarkoma (Statistics Finland) and Steven Thomas (Statistics Canada). The topic was divided into 2 parts, Microdata access facilities and Microdata protection. It included the following presentations:

Microdata access facilities:

- Romania NIS – Microdata for scientific purposes;
- Hungary – Joint Safe Centre of the Hungarian Central Statistical Office and the Hungarian Academy of Sciences;
- Access to microdata in the State Statistical Office of the Republic of North Macedonia;
- Canada – Virtual data labs - A more flexible approach to access Statistics Canada microdata;
- Mexico – Harnessing the potentiality of microdata access risk management model;
- Israel – Data Confidentiality in ICBS Research Rooms;
- The United Kingdom – Accessing Data in the ONS Secure Research Service: A Certification Regime for Remote Connectivity.

Microdata protection:

- Norway – Synthetic data generation for anonymization purposes. Application on the Norwegian Survey on living conditions/EHIS;
- The Health Foundation and Universities of Essex and Manchester – Promoting Statistical Disclosure Control for novices: A Handbook;
- World Bank – A practice guide for microdata anonymization;
- University of the West of England – Training research output checkers;
- Institute of Economic Research, Hitotsubashi University – Creation of synthetic microdata using dummy random variables of high dimension statistics based on big data.

B. Topic (ii): Tabular data

7. This topic was organized by Sarah Giessing (Destatis). It included the following presentations:

- Germany – Concepts for generalising tools implementing the cell key method to the case of continuous variables;
- France – Prodcum disclosure control with non-nested national and European classification;
- Universitat Politècnica de Catalunya – Using a stabilized Benders algorithm for cell suppression;
- Norway – Releasable inner cell frequencies by post-processing protected tabular data;
- Brazil – Primary analysis of disclosure risk in tabular data from a Brazilian economic survey;
- The Institute of Statistical Mathematics, Tokyo – Algorithmic Matching Attacks on Optimally Suppresses Tabular Data;
- Australia – ABS perturbation methodology through the lens of Differential Privacy.

C. Topic (iii): Risk assessment

8. This topic was organized by Josep Domingo-Ferrer (Universitat Rovira i Virgili) and Krish Muralidhar, (University of Oklahoma). It included the following presentations:

- University of Manchester – Trade-off between Information Utility and Disclosure Risk in GA Synthetic Data Generator;
- University of Manchester – The Synthetic Data Challenge;
- University of the West of England – 10 is the safest number that there's ever been;

- University of Oklahoma – Connecting privacy models and statistical disclosure control methods through bistochastic anonymization;
- University of Oklahoma – Privacy, confidentiality, disclosure: What is the difference?;
- German Institute of Medical Documentation and Information – Assessing the re-identification potential of health care data for people with statutory health insurance in Germany (preliminary results).

D. Topic (iv): Emerging issues

9. This topic was organized by Josep Domingo-Ferrer (Universitat Rovira i Virgili). It included the following presentations:

- Netherlands – Comparing methods of safely plotting variables on a map;
- Italy – Privacy Preserving Set Intersection;
- Chuo University and NTT DOCOMO, INC – The Potential of Anonymization Methods for Creating Detailed Geographical Data in Japan;
- Universitat Rovira i Virgili – Protecting consumer privacy in smart metering by randomized response;
- FAO – Challenges and experiences in anonymizing and disseminating microdata from agricultural surveys in the context of the FAO AGRISurvey program.

E. Topic (v): The framework for confidentiality

10. This topic was organized by Aleksandra Bujnowska (Eurostat) and Janika Tarkoma (Statistics Finland). It included the following presentations:

- The Health Foundation – Understanding personalities in data access decision-making;
- Canada – Successes and Challenges in Increasing Accessibility at Statistics Canada;
- Germany – Evaluation criteria for the selection of a SDC Method;
- Finland – Crisis management – training, practicing and testing;
- Eurostat – Data protection laws and methods in official statistics.

F. Topic (vi): Confidentiality issues of the Census 2020/2021 Round

11. This topic was organized by Eric Schulte Nordholt (Netherlands). It included the following presentations:

- Portugal – A framework for assessing perturbative methods for protection of Census 2021 data at Statistics Portugal;
- Russian Federation – Ensuring data confidentiality of All-Russian Population Census 2020.

G. Topic (vii): Software tools for statistical data confidentiality

12. This topic was organized by Peter-Paul de Wolf (Netherlands), and it included the following presentations:

- Austria and Germany – cellKey - consistent perturbation of statistical tables;
- Norway – Microdata.no - Safe Access to Register Microdata.