



Economic Commission for Europe**Conference of European Statisticians****Group of Experts on Population and Housing Censuses****Twenty-fifth Meeting**

Geneva, 20–22 September 2023

Item 4 (l) of the provisional agenda

**Revising the Conference of European Statisticians Recommendations
for Population and Housing Censuses for the 2030 round:****Dissemination of disaggregated census data****Preliminary progress report of the Conference of European
Statisticians Census Task Force on Dissemination of
disaggregated census data****Note by the Conference of European Statisticians Task Force on
Dissemination of disaggregated census data*¹***Summary*

After the review process for the CES Recommendations was kicked off, this Task Force held three calls and various written exchanges. The work is organised along three strands (general dissemination, statistical disclosure control, geographically referenced outputs). The Task Force reviewed the 2013 survey questionnaire and developed proposals for various question updates and new questions within its topical scope.

Notable topics that are addressed better with the updated questions include: modern dissemination technology and tools including table builders or data visualizations, new developments in statistical disclosure control including noise-based methods and implications from grid outputs or flexible dissemination tools, and recent developments on functional output geographies including grids.

* Prepared by Fabian Bach, European Commission – Eurostat. The views expressed are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission.

Note: The designations employed in this document do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

¹ This document was submitted late due to delayed submission of the paper by the Task Force.



I. Introduction: Task Force scope and work plan

1. The main part of the 2020 CES Recommendations (CESR)² which this Task Force (TF) will review is the ‘dissemination’ subheading, paragraphs 292-307 of the section ‘dissemination, documentation, metadata and archiving’ in Chapter III, field and other operational activities. There are also potential implications for the section ‘Confidentiality and security’, paragraphs 140-148 in Chapter I, Methodology.
2. The TF scope includes:
 - (a) Dissemination formats and products, with their respective rationales, target audiences and specific considerations;
 - (b) Quality assurance in dissemination at high levels of granularity;
 - (c) Implications for confidentiality and disclosure control of different approaches to census data dissemination;
 - (d) Dissemination scheduling;
 - (e) Dissemination of geographically referenced outputs (the Task Force will need to coordinate this part of their review with the Task Force working on geospatial information);
 - (f) Documentation and metadata;
 - (g) Archiving.
3. The TF consists of nine members from: Canada, Eurostat (chair), Integrated Public Use Microdata Series (IPUMS), Ireland, Israel, UNECLAC, UNFPA Kyrgyzstan, UNSD and the UN Women Regional Office for Europe and Central Asia.
4. The TF agreed on a work plan splitting its activities into two broad stages:
 - (1) reviewing relevant questions of the 2013 survey questionnaire (by April 2023);
 - (2) reviewing the mentioned CESR parts based on the analysis of 2023 survey results (i.e. starting autumn 2023 after survey returns). Throughout both stages, TF members organise themselves in three thematic sub-groups to prepare initial ideas to be discussed by the full group:
 - (a) General dissemination, metadata and archiving aspects (led by Ireland);
 - (b) Statistical disclosure control (led by Eurostat);
 - (c) Geographically referenced outputs (led by Canada).
5. This paper reports on the successfully concluded first work stage, namely the questionnaire review. The TF proposed to the Steering Group question updates and new questions in the following 2013 questionnaire sections: Methodology; Technology; Communication and publicity; Documentation, metadata and archiving; Security, confidentiality and disclosure control; Dissemination; Innovations; Quality and coverage; Geographic characteristics. The TF further provided comments and considerations (without having been able to propose question updates) to some elements of the sections: Problems, successes and lessons learned; Costs and benefits.

² https://unece.org/DAM/stats/publications/2015/ECECES41_EN.pdf.

II. Exchange of dissemination experiences across TF members

6. At the second meeting, all TF members had the opportunity to present specific dissemination products or aspects from their own experience. Seven TF members took the floor with the following interventions:

(a) Eurostat presented the main dissemination channels for 2021 EU census outputs, namely the Census Hub³ for comprehensive access to all outputs via a table builder interface, Eurobase⁴ to provide a selection of hypercubes through a data browser, and Statistics Explained articles⁵. Finally, outputs also comprise key indicators on a European 1 km square grid (total population already published⁶);

(b) Ireland gave an overview of its national census dissemination products comprising PX Stat tables (over 1000 of up to 4 dimensions)⁷, electronic publications⁸, small area population statistics incl. census mapping and interactive maps⁹, research micro files¹⁰ and grids;

(c) UNECLAC showed its regional data dissemination platform REDATAM¹¹ with a presentational focus on two products in the census context: harmonised census microdata and migration matrices for the Latin American region;

(d) Canada described its census dissemination products¹² including specific census profiles¹³ and maps¹⁴, before outlining some plans for the future including a new hybrid online table builder;

(e) IPUMS¹⁵ presented its activities and products focussing on internationally comparable releases of census and survey microdata from all world regions;

(f) Israel described its national situation and current challenges, mainly around a profound modernisation of its entire census dissemination system. One of the biggest challenges is how to release data securely;

(g) UNFPA Kyrgyzstan summarised its past and ongoing regional activities, including support for Kyrgyz census publications (e.g. from 2010 round) and current efforts to better cover Washington Group¹⁶ questions on disability statistics and migration.

III. Questionnaire review

7. Key considerations and ensuing questionnaire updates proposed by the TF are presented as follows along the three thematic strands introduced in paragraph 4.

³ <https://ec.europa.eu/CensusHub2> (so far only 2011 EU census outputs; 2021 outputs will be added when available).

⁴ <https://ec.europa.eu/eurostat/data/database> (the 2021 census tab contains some data sets provided voluntarily by EU members, namely population by broad age group and NUTS 3 region and various specific data sets on population with Ukrainian citizenship – in the context of Russia’s unprovoked invasion of Ukraine).

⁵ <https://ec.europa.eu/eurostat/statistics-explained/index.php?search=census>.

⁶ https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_and_housing_census_2021_-_population_grids.

⁷ <https://data.cso.ie>.

⁸ <https://www.cso.ie/en/census>.

⁹ <https://visual.cso.ie/?body=entity/ima/cop/2016&boundary=C03736V04484>.

¹⁰ <https://www.cso.ie/en/census/census2016reports/powscar>.

¹¹ <https://www.cepal.org/en/subtopics/redatam>.

¹² <https://www12.statcan.gc.ca/census-recensement/index-eng.cfm>.

¹³ <https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E>.

¹⁴ <https://www12.statcan.gc.ca/census-recensement/2021/geo/maps-cartes/index-eng.cfm>.

¹⁵ www.ipums.org.

¹⁶ <https://www.washingtongroup-disability.com>.

A. General dissemination, metadata and archiving aspects

8. The TF saw a general trend towards more interactive, web-based dissemination tools since the CESR were last updated. For instance, several countries have introduced interactive table builders as a dissemination tool allowing users to query customised cross-tabulations. At the back end, these tools can be either linked to the microdata (e.g. Australian Table Builder¹⁷ and Bangladesh Bureau of Statistics¹⁸) or to a pre-defined set of very detailed cross-tabulations (e.g. the EU Census Hub, see footnote **Error! Bookmark not defined.**). While the back end and some other design choices may have implications for disclosure control scenarios (see paragraph 16), the TF generally agreed that the resulting interactive and user-friendly front ends of such tools are an important improvement in census data dissemination. Often such tools are combined with interactive data visualization tools.

9. The TF also noted that the modes of census data taking and storage continue to evolve quickly with the digitalisation of societies in many countries. On the one hand, many countries are reducing direct contact with the population by relying more and more on administrative and other data sources. On the other hand, also censuses based fully or partly on a direct survey are increasingly implemented with smart and digital tools.

10. Finally, the TF noted that communication with the user community is an increasingly valued aspect of successful census exercises. This applies to both before the census taking (e.g. stakeholder consultations to inform census planning) as well as after the publication of results (e.g. types of documentation, metadata and quality information provided).

11. Therefore, the TF proposed various updates to the questionnaire to improve the return of information along these lines. Most notably, several questions were extended or newly added to better capture if/how table builders are provided as part of the dissemination; a question was added on whether the census authority has dedicated human resources specifically to data visualization technologies; questions were extended or newly added to adapt to developing modes of census taking (use of administrative data, smart survey response tools); and some questions were proposed on stakeholder consultation, types of documentation, and types of quality indicators provided with the data.

B. Statistical disclosure control aspects

12. The TF acknowledged profound developments in the field since the last CESR update. Notably, there have been many activities across the world to explore and move towards new statistical disclosure control (SDC) methods based on noise, i.e. the application of small perturbations to the microdata or contingency tables in order to protect confidential information.

13. For instance, after the 2011 EU census that saw a proliferation of suppressed data cells in the more complex cross-tabulations, and following initial work e.g. in Australia¹⁹ and other countries, Eurostat organised a project with country experts to develop more efficient SDC methods for the 2021 EU census. This resulted in expert recommendations²⁰ and public software tools²¹ centring around two noise-based methods: pre-tabular targeted record swapping and the post-tabular cell key method. These were selected to not only reduce the amount of suppressed data but also to address specific additional risks from publishing data on non-nested geographies (e.g. 1 km square grid and municipal boundaries).

14. For the 2021 EU census, a significant number of EU countries is expected to apply an SDC setup based on these recommendations. In the wake of the ongoing modernisation of population statistics in the EU, these recommendations will be evaluated and likely

¹⁷ <https://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder>.

¹⁸ <http://redatam.bbs.gov.bd/redbin/RpWebEngine.exe/Portal>.

¹⁹ Thompson G., Broadfoot S., Elazar D. (2013), "Methodology for the Automatic Confidentialisation of Statistical Outputs from Remote Servers at the Australian Bureau of Statistics," [Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality](#).

²⁰ https://cros-legacy.ec.europa.eu/content/harmonised-protection-census-data_en.

²¹ <https://github.com/sdcTools/CensusProtection>.

developed further or possibly complemented by other ideas – for instance, Belgium and France have developed own noise-based methods especially to protect their grid outputs.

15. The U.S. Census Bureau has for the first time published its 2020 census outputs protected by a differentially private²² noise method²³. This received mixed reactions by experts and data users down to grave utility concerns²⁴ and ensuing debates²⁵.

16. Finally, the TF took note of specific SDC challenges when detailed data are released with interactive tools allowing users to customise data queries²⁶. When such tools are considered to be provided as a part of the dissemination programme, the SDC strategy should account for respective attack scenarios.

17. Therefore, the TF proposed various updates to the questionnaire to improve the return of information along these lines. Most notably: (targeted) record swapping, the cell key method and differentially private methods are added as reply options for SDC methods used; questions are added on the use and properties of interactive output tools and SDC implications; and a question is added on specific measures to protect against geographic differencing of non-nested output geographies.

C. Geographical dissemination aspects

18. The TF identified as the most notable development since the last CESR update a strongly increased interest in georeferenced data and analysis. This comprises most notably a widespread trend towards offering census data georeferenced to geographical grids, but also towards offering more powerful and interactive geographical visualization tools such as map builders. New emerging challenges for SDC strategies were also considered (see paragraph 13).

19. The increased availability of georeferenced and smallest-area data and respective analysis tools typically also calls for adaptations to the metadata provided. For instance, any estimation methods used to produce census outputs at the smallest geographical level (e.g. grid) would likely be documented separately. Moreover, countries may publish relevant auxiliary information together with their census outputs to allow more powerful geospatial analyses (e.g. data on road networks or boundaries of fixed or custom defined geographies).

20. Therefore, the TF proposed various updates to the questionnaire to improve the return of information along these lines. Most notably: questions related to small area geographies were updated to cover grids explicitly, including grid resolutions and which statistical units (persons, dwellings, households, etc.) would be available on a grid; some questions were added to capture geographical estimation methods and auxiliary information published together with small-area or grid data.

21. The TF also discussed whether the questions on the definition and use of ‘localities’ for urban/rural delineation should be adapted or complemented to also cover more recent methodologies to delineate cities, urban and rural areas. More specifically, the UN Statistical Commission endorsed at its 51st session in 2020 a new methodology and manual to delineate

²² Dwork C., McSherry F., et al. (2006), “Calibrating Noise to Sensitivity in Private Data Analysis,” in *Theory of Cryptography*, eds. Halevi S., Rabin T., pp. 265–284. Berlin, Heidelberg: Springer Berlin Heidelberg.

²³ Abowd J. M. (2018), “The U.S. Census Bureau Adopts Differential Privacy,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD ’18*. London, UK: Association for Computing Machinery, p. 2867. DOI: 10.1145/3219819.3226070.

²⁴ E.g. Ruggles S., Fitch C., Magnuson D., Schroeder J. (2019), “Differential Privacy and Census Data: Implications for Social and Economic Research,” *AEA Papers and Proceedings*, 109, 403–408.08.

²⁵ E.g. Muralidhar, Domingo-Ferrer (2023), “Legacy Statistical Disclosure Limitation Techniques for Protecting 2020 Decennial US Census: Still a Viable Option,” *Journal of Official Statistics*, Forthcoming, Available at SSRN: <https://ssrn.com/abstract=4472525> or <http://dx.doi.org/10.2139/ssrn.4472525>.

²⁶ E.g. Asghar H. J., Kaafar D. (2020), “Averaging Attacks on Bounded Noise-Based Disclosure Control Algorithms,” *Proceedings on Privacy Enhancing Technologies*, 2020, 358–378.

cities, towns and rural areas²⁷ that essentially derives population clusters from a combined analysis of small administrative areas and population grids. However, the TF decided to forward these considerations to Task Force 11 on geospatial information and small area statistics for censuses, rather than proposing own questionnaire updates in this regard.

IV. Conclusion and next steps

22. The Task Force reviewed the 2013 survey questionnaire in the context of its topical scope and proposed to the Steering Group a number of question updates and new questions along three work strands: general dissemination, metadata and archiving aspects; statistical disclosure control aspects; and geographical dissemination aspects. To the author's knowledge at the date of finalising this report, a large majority of the proposals were taken into the final questionnaire for the upcoming 2023 survey.

23. As soon as the 2023 survey results become available, the Task Force will analyse them and take up the work on reviewing the actual CES Recommendations.

²⁷ <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-02-20-499>.