United Nations ECE/CES/GE.41/2022/3

Economic and Social Council

Distr.: General
23 June 2022

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Twenty-fourth Meeting
Geneva, 21−23 September 2022
Item 5 of the provisional agenda
Transitions in census methodology; plans, experiences and innovations

Perturbative Methods for Census 2021 tables within the European Framework

Note by Statistics Netherlands*

Summary

Censuses have been an essential part of the statistical programme of national statistical institutes for many decades. The European perspective has become an important additional dimension to the dissemination of all the different census results. Combining the national census results obviously requires some coordination and harmonization. As a first step harmonization of the design of the census output tables for the European dimension was established. This clearly makes it easier to combine the census tables from different countries. However, since member states used different disclosure control methods to protect the privacy of their inhabitants, the advantages of harmonized table design were not as big as anticipated. Thus, harmonization of the statistical disclosure control methods used was needed as well. Two European projects were launched to propose such a harmonized approach to disclosure control of census tables. It is not mandatory for member states to use the proposed methods, but if many countries would use them, it would definitely improve the comparability of the European census tables. This paper is partly based on the results of the two European projects that aimed at developing and implementing a harmonized approach. Moreover, we will pinpoint some remaining issues to consider when the proposed methods will be used for the 2021 European Census.

* Prepared by Eric Schulte Nordholt and Peter-Paul de Wolf.

I. Introduction

1. The year 2021 is European Census year. This implies that all Member States of the European Union (EU) have to conduct a Population and Housing Census with a reference day in 2021 (Census Day). This is an important means to harmonize European census results. Moreover, all EU countries will publish a set of harmonized tables to make comparisons possible. This set of linked high dimensional tables gives a precise description of the people living in the EU and their housing situation. This table set is called the European Census 2021 hypercubes. Additionally, for the first time a set of grid square tables is mandatory for the European Census 2021.

2. The experience of the Census 2011 in Europe demonstrated that further harmonization is required to make the data more internationally comparable. Different European countries applied substantially different methods to protect their Census 2011 tables which hampered the ability to compare the results across countries to a great extent. It is legally not possible to prescribe exactly how the European Census 2021 hypercubes have to be protected. However, by exchanging information on best practices and recommending protection methods for these tables an important step forward can be made. In this paper these recommendations are presented. The recommendations are based on experiences in many individual countries inside and outside the EU. Much progress was made in two European projects. The key results of these projects are described in this paper.

3. In section II a historical perspective is sketched. The proposed protection methods are described in section III. Section III also discusses combining these methods. Section IV describes how the methods have been assessed so far in the Netherlands. The paper ends with a conclusion in section V. The conclusion includes some remarks on issues that may be encountered when using the proposed methods in practice.

II. Historical perspective

4. Censuses are important data collections. All Member States of the European Union conduct population and housing censuses to provide comprehensive data about their populations. European Union Member States must provide census data to Eurostat, the statistical office of the European Union. Eurostat compiles European level census data based on the data provided by the Member States. In most Member States, census data can only be released if measures have been taken to prevent disclosing information on individual respondents. Therefore, Statistical Disclosure Control (SDC) is an important step before census data are released. In Hundepool et al. (2012) more information can be found on Statistical Disclosure Control in general. In this paper, we discuss the specific aspect of protecting detailed and linked census tables.

5. For the European Census 2001, the legal underpinning was lacking. Only a gentlemen’s agreement was reached stating that all Member States would do their best to provide Census tables to Eurostat. Clearly, this was not a sufficiently strong basis to produce all tables required of all Member States.

6. The situation improved for the European Census 2011 thanks to the introduction of the European Census Act (European Commission, 2008). As delivering census microdata to Eurostat met legal hurdles, the concept of census hypercubes was introduced at that time. Hypercubes are high dimensional tables. From these tables many simpler tables can be derived for publication purposes. At the same time the format in which the tables had to be provided changed from Excel to SDMX.

7. It is possible to apply regular non-perturbative SDC methods to census tables. These methods include table redesign, global recodes and local suppressions. European Census table lay-outs were fixed to facilitate better comparisons between data of different countries. In the Census 2011 this was introduced using the mandatory new format of SDMX. However, these fixed table lay-outs also implied that table redesign and global recodes were no longer options to protect the tables against disclosure of individual information, whilst in the Census 2001 these two methods were heavily applied by several countries.

8. It may seem a feasible alternative to apply local suppressions to protect a high dimensional table. Indeed, it is also often feasible to protect a few linked tables simultaneously. However, the set of hypercubes of the Census 2011 was far too large and complex to be protected by local suppressions in an optimal way. By optimal we mean obtaining a set of suppressions such that none of the primary unsafe cells can (approximately) be recalculated while the hypercubes maintain a sufficient amount of information to remain useful to the users.

9. The problem of how to protect a set of census hypercubes properly was acknowledged and a Task Force on Census Statistical Disclosure Control was created in 2008. The work of this Task Force was complicated as no real hypercubes yet existed and countries were legally not allowed to share their old census microdata. Moreover, as it is the responsibility of the countries to protect their own census tables, it is not possible to mandatorily prescribe how the census hypercubes had to be protected. As a result, in the end all countries protected their Census 2011 hypercubes in their own way. An additional complication was the fact that representatives of different countries had different ideas on what information was sensitive and needed to be protected. This result was disappointing.

10. The harmonization intended by the Task Force should have led to more comparable results across countries, but the practice became that the hypercubes made available in the Census Hub (see https://ec.europa.eu/CensusHub2/) were not always comparable across countries due to a wide variety of different protection methods. Some countries did not protect their tables at all, many countries introduced missing values in sensitive cells and lots of other cells to protect the sensitive cells and other countries added noise in different ways to protect their hypercubes.

11. Two sensitive variables with many categories were heavily sacrificed in the Census 2011 hypercubes of the Netherlands. For the variables country of birth and country of citizenship only aggregates were published and individual country information was suppressed when the data were transformed into SDMX format and published in the Census Hub. It is clear that this protected the sensitive information, but the information loss was huge and led to many additional table requests in later years. Such requests were normally accepted if no high dimensional tables were asked for, as high dimensional tables would still lead to disclosure of individual information. Clearly, Statistics Netherlands was not keen on repeating this approach for the Census 2021 tables. Similarly, many other countries were not satisfied with their own SDC approach for the Census 2011 tables either.

12. It was clear that the process of producing European Census results had to be improved further. The European Census Act (European Commission, 2008) is the legal basis for the European Censuses of both 2011 and 2021. For the Census 2021 in addition four implementing regulations have been made to specify precisely what the Member States have to deliver (European Commission, 2017a, 2017b, 2017c and 2018). Moreover, in order to try to prevent a similar undesired situation on the protection of the hypercubes for the European Census 2021, two Specific Grant Agreements (SGAs) under Framework Programme Agreement (FPA) № 11112.2014.005-2014.533 were signed in recent years. Results of these two SGAs are discussed in the following two sections.

III. Proposed methods

A. Introduction

13. The first SGA mentioned in section II (№ 11112.2016.005-2016.367) started in September 2016 and lasted for one year. The statistical offices of six European countries (Finland, France, Germany, Hungary, Netherlands and Slovenia) were involved and Statistics Netherlands acted as project leader. During this project on ‘Harmonised protection of census data in the European Statistical System (ESS)’ a survey was conducted among the countries of the ESS about the protection of their census tables. Obviously, this questionnaire included questions on legal, methodological and technical aspects. The aim of the SGA on “Harmonised protection of census data in the ESS” was to provide recommendations for the protection of the 2021 census tables. Such recommendations could only be made properly if

the national (legal) situations were taken into account and therefore a number of questions were asked about these situations. Additionally, questions were asked on the countries' evaluations of their protection methods of the 2011 census hypercubes and on the use of grid cells in (national) census hypercubes. In the end, response of 33 European countries (27 out of the then 28 Member states and 6 of the 7 Accession countries) were received.

14. Note that the recommendations made in this SGA did not imply legal obligations to the countries in the ESS. The aim of this SGA was to provide guidance in obtaining well-protected census tables that are easy to compare among countries.

B. Conclusions drawn in the SGA

15. National laws that apply to the publication of census results are often vague on how and what to protect. The Census 2021 hypercubes is a set of linked high dimensional tables. This implies that many census table cells will have a very low value or will even be 0. This implies that individual information can be relatively easily disclosed from the Census 2021 hypercubes. This clearly shows that taking no disclosure control measure at all would be unlawful for all ESS countries. Similarly, even though the notion of sensitivity differs between countries, the general consensus is that the most problematic census variables seem to be country/place of birth (POB) and country of citizenship (COC). Especially the most detailed level of these variables (individual countries) could aid in disclosing individual information in the hypercubes in which they appear.

16. In the survey, many countries mentioned that post-tabular methods are not popular. However, in our view, without post-tabular methods it will be virtually impossible to properly protect the census hypercubes. This is related to what most countries are aware of: when protecting census hypercubes one should also look at national census tables. Indeed, even if both the European hypercubes and the national tables are protected properly *on their own*, the combination of the publications is not necessarily safe. As a consequence, if for example the national tables are published first, the European hypercubes have to be protected conditionally on the already published information. A similar situation would arise when considering other (national) demographic publications.

17. In the Census 2021 new kinds of tables are added to the European hypercubes: tables on grid squares of 1 km × 1 km. These tables are not detailed in content (for each of these tables only one characteristic is included), but detailed in structure (the number of grid squares is in all countries much larger than the number of municipalities). Moreover, grid squares and regional distributions are non-nested variables. This implies that countries have to check whether information about individuals can be disclosed by crossing these grid squares with municipalities (LAU), the most detailed level of region in the European hypercubes. Other levels of geography (Country, NUTS1, NUTS2 and NUTS3) in the hypercubes are combinations of LAUs. Moreover, these levels are nested, i.e. they follow a hierarchical structure. The grid squares are the only geographical variable that is not nested in this hierarchical structure.

18. Member States produce and provide census hypercubes for Eurostat. The variables of each hypercube and their categories are harmonized across the countries. Therefore, the Member State data can be combined into European-level data. However, Member States can apply SDC methods of their own choice and the differences across countries in the methods used might have a negative effect on the quality of European-level data. Eurostat aims at harmonizing the SDC methods across the Member States in order to increase data quality. The more Member States that apply the recommended SDC methods, the more harmonized the European-level data can become.

19. Classical non-perturbative methods like global recoding and cell suppression are for different reasons not a solution to protect the European census tables. To make comparisons between countries possible the table formats are fixed and cannot be altered. Therefore, global recoding is not an option. Applying cell suppression to such a large set of high dimensional linked tables in an optimal way is practically impossible. Theoretically, it would be possible to apply cell suppression with a lot of over-suppression to make the set of tables safe. This would however lead to huge information loss which is unacceptable from a user's

point of view. Another problem is the management of risk of disclosure by differencing between hypercube and grid-level data, adding further to the complexity of cell-suppression-based protection concepts.

20. A harmonized method should offer some flexibility so that countries can easily adapt it to their specific needs and expectations regarding an acceptable level of residual disclosure risk on the one hand and an acceptable level of information loss on the other hand. The method should be adaptable by changing parameters and should consist of separate modules that can be used in combination. The idea was then born to include modules for pre-tabular perturbation, as well as modules for post-tabular perturbation.

21. Therefore, the project team decided to select the pre-tabular method of Targeted Record Swapping and the post-tabular Cell Key Method in which noise is added to the table cells. The following two subsections briefly introduce the suggested methods. The parameters of both methods are not fixed; Member States can decide on them. Both methods do not lead to suppressed data, therefore Member State data, if treated by these methods, can be combined into European-level data.

22. If many Member States use the same method − though perhaps in different flavours − this will help to prepare European-level data in a more straightforward way. Unlike with cell suppression, with the perturbative methods proposed, data will be available for all hypercube cells. This will be a great advantage for all users and will greatly improve the comparability of the data across countries.

23. In order to maintain consistency between European and national data releases, Member States are encouraged to apply the same SDC method to all kinds of data releases. If, however, another method is employed to protect the national release data, Member States should check and eventually develop variants that avoid residual disclosure risks that might arise when users compare the European hypercube data to national releases.

C. Targeted Record Swapping

24. Record swapping is a pre-tabular SDC method and as such it is applied to the microdata prior to constructing the census hypercubes. The general idea of record swapping is that pairs of records are selected and the values of certain variables are swapped between the records. The selection of the records is usually done in such a way that some specific analytic properties are maintained and the introduced bias is minimized.

25. The Targeted Record Swapping (TRS) as suggested for the Census 2021 hypercubes is based on an approach developed by the UK Office for National Statistics (ONS). Some small adjustments had to be made because the ONS implementation was targeted at the specific UK situation and the SGA was aiming at a general implementation applicable to all Member States.

26. TRS is applied at household level in the sense that only complete households will be swapped, not individual persons. This is one way to prevent changing the distribution of household characteristics too much. Additionally, only the geographical variables will be swapped. This way, dependencies within households will not be affected too much.

27. Generally speaking, TRS can be described in the following way (the levels of geography mentioned are assumed to be nested):

(a) At each level of geography, determine the households with a disclosure risk above a certain threshold;

(b) Then start at the coarsest level of geography available;

(c) Determine ‘similar’ households (i.e. households that have the same values on some specified household characteristics, but not necessarily on other characteristics) in other regions at the same geographical level to obtain a set of donor households;

(d) Randomly select one of the donor households;

(e) Swap the geographical variables of the household at risk with those of the selected donor household (i.e. swap the same for all members/records of the households);

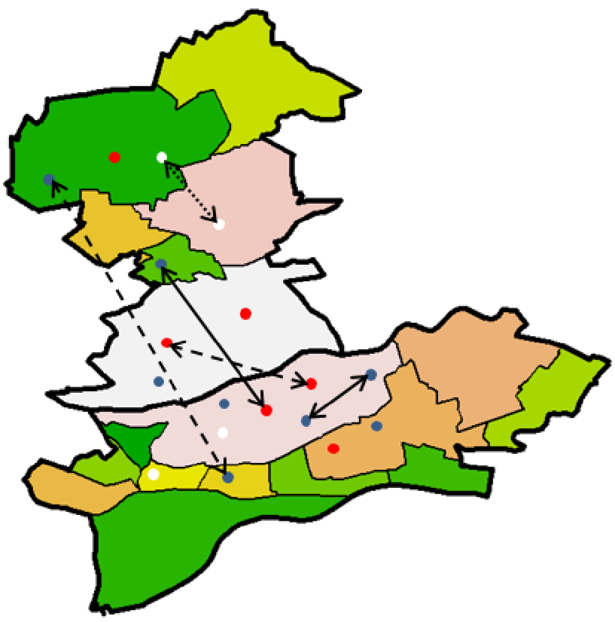
(f) Go to the next detailed level of geography and repeat from step 3 until the most detailed level is reached;

(g) If the percentage of swapped records is below a predefined threshold once the most detailed level of geography is dealt with, swap additional households randomly at the most detailed level of geography until the desired swap rate is reached.

28. An additional restriction that is imposed on the above-mentioned iterative process is that a household cannot be swapped twice. Furthermore, it is clear that this process leads to swapping of all households that are at risk. This implies that it will be much more difficult to identify individual information after the swapping has been applied.

29. Figure 1 visualizes the application of TRS. The figure displays two geographical levels, represented by thick borders (coarsest level) and coloured areas (detailed level). The coloured dots represent households, where similar households have the same colour. The arrows with a solid line show swaps that are not allowed: swaps of non-similar households, or swaps of households within the same area at the same geographical level. The arrows with dashed lines show possible swaps when working at the coarsest level, the arrows with dotted lines show a possible swap when working at the most detailed level.

Figure I
Illustration of TRS



Note: Solid arrows show swaps that are not allowed, dashed arrows swaps at the coarsest geographical level, dotted arrows swaps at the most detailed level.

30. After TRS is applied, the census hypercubes can be calculated. This will result in hypercubes where cell counts may differ from the "original" cell counts due to the swapped households and their corresponding household members.

D. Noise addition using a Cell Key Method

31. Noise addition using a Cell Key Method (CKM) is a post-tabular method and as such is applied to the constructed set of census tables. CKM does thus not change the underlying microdata and only effects the census tables. CKM is based on the method as introduced by the Australian Bureau of Statistics (ABS), see e.g. Fraser and Wooton (2006). Their method builds on so-called "cell keys" to ensure that the random noise added to a specific cell will always be exactly the same, irrespective of the particular census hypercube it appears in. We

have slightly adjusted the method proposed by ABS, in the sense that we are a bit more flexible in assigning cell keys.

32. To ensure the consistency of the added noise between different hypercubes, the process of assigning cell keys to cells should be consistent from the start. To that end, so-called "record keys" are assigned to the records in the microdata underlying all census hypercubes. That is, a random number is assigned to each individual of the population. Whenever a cell of a hypercube is constructed, the number of records that fall into that cell is calculated, and the record keys of those records produce a cell key which will be used to select the noise to be added. That way, the randomness of the record keys determines the randomness of the noise, whereas the deterministic character of calculating the cell keys will ensure the consistency across the different hypercubes.

33. Generally speaking, CKM can be described in the following way:

(a) Assign a uniformly [0,1) distributed number to each record in the census microdata;

(b) Determine a so-called "p-table" that defines the noise distribution;

(c) When aggregating the microdata to a census hypercube, for each cell additionally calculate in a deterministic way the cell key using the record keys of the records that fall into that cell;

(d) Use the cell key together with the cell value to determine from the p-table the noise to be added to the cell;

(e) Add the noise to the cell value.

34. The cell keys are calculated as follows: add the record keys of all records that fall into that particular cell, then take the fractional part of the result as cell key. That way the cell keys are also uniformly [0,1) distributed values. These values can then be used to draw from the distributions in the p-table. Essentially the cell keys are taken as arguments of the inverse distribution to obtain a realization of the noise.

35. The p-tables suggested for use for frequency count tables (i.e. for the census hypercubes) have some specific parameters that can be set:

(a) The variance of the added noise, denoted by V;

(b) The maximum value for added noise, denoted by D and allowing for a noise distribution on $\{-D, -D+1, \dots, -1, 0, 1, \dots, D-1, D\}$;

(c) The minimum positive cell value allowed after adding the noise, denoted by $j_s + 1$.

36. Additionally, the distribution in a p-table should be such that negative cell counts are impossible while the expectation of the added noise is zero for each cell. As a consequence, zero cells (cells with a count of zero) are not allowed to be perturbed into positive counts. Moreover, as zero cells are sometimes known to have no contributors, changing these cell values would not contribute to the protection of the tables.

37. Note that j_s could for example be used to prevent the occurrence of values of 1 and 2 in frequency count tables, as some national laws require. However, this parameter still allows that a positive cell count can be perturbed into the value zero. If for j_s the value 2 is chosen, then the set of possible cell values after applying CKM will be $\{0, 3, 4, \dots\}$.

E. Combining the two methods

38. Even though the TRS and CKM methods are both recommended to be used to protect the census hypercubes in a harmonized way, different EU Member States still have some freedom on how to use these methods. Not only can they choose different parameter values, they can also decide to use only one of the methods or a combination of both methods. Indeed, given the different confidentiality rules applicable in different European countries as well as differences in the size0 of these countries, it was advisable to recommend not just a single

method. However, by limiting the number of recommended methods, it will be easier for Eurostat as well as other users to compare protected census statistics between countries.

39. An advantage of combining the two methods would be that the parameters used for each method may be set less strict as compared to a situation where only one of the methods is used. Moreover, CKM specifically aims at protection against differencing whereas TRS introduces uncertainty in general, but essentially on record level.

VI. Assessing the methods

40. During the second SGA (№ 2018.0108) preliminary releases of tools to implement the methods were made publicly available and Member States were invited to test them and provide feedback. Unfortunately, only a limited number of Member States actually did provide feedback. Their feedback mainly concerned installation issues and conceptual questions, e.g., how to choose the parameters. More (research) work on the choice of adequate parameter values is currently being conducted in several European countries.

41. We are aware of several assessments of the proposed methods. Here we briefly describe the one by Statistics Netherlands (SN).

42. Since SN conducts their census using administrative data, they were able to produce a test census dataset that was more recent than the 2011 census. The dataset used for assessing the two implemented methods was based on 2017 population data. SN's intention was to use a combination of TRS and CKM. In their view, the main protection against differencing would come from CKM. However, if CKM was to deliver enough protection on its own, the parameters would probably have to be taken relatively strict. This would hamper the utility quite severely.

43. TRS and CKM each have their own impact on utility and disclosure risk. By combining the two methods, the burden on the parameters could be divided over both methods. This allows for not too strict parameter settings for each individual method. Moreover, the effects on utility could also be split over both methods. Based on these considerations, SN only assessed the application of a combination of TRS and CKM.

44. Since the publication of hypercubes based on grid cells was new, SN concentrated their assessment on this type of tables. They considered not only differencing risk between tables, but also between different kinds of geographical variables. E.g., they considered the differencing between grid cells and LAU regions.

45. Several variants of the combination of TRS and CKM were considered using different values for the parameters. At the time of writing this paper, several European countries have decided to use TRS, CKM, or a combination to protect their Census 2021 hypercubes, but no final conclusions have been drawn from the assessment of SN. However, the research conducted so far at SN has taught us that for census tables a small swap rate for the TRS (e.g., 1%) where at least all records at risk are swapped will contribute significantly to the protection in the sense that it is much more difficult to judge whether a found disclosure is a real disclosure. Higher swap rates would lead to severe information losses. For CKM it was found that even with not very small variances (e.g., 2 or 3) the information loss need not be so high. These lessons have been used to find suitable parameter values in using a combination of TRS and CKM.

V. Conclusions

46. Harmonizing the way census tables are presented has been on the agenda of Eurostat for quite some time. The recent developments concerning the Statistical Disclosure Control of the European census hypercubes seems promising. Several countries have the intention to use TRS or CKM as their disclosure control method for the Census 2021 hypercubes. Assessments of the methods by Member States suggest that TRS should not be used on its own because the remaining disclosure risk is still too high. CKM could be used on its own, but by combining it with TRS it seems that the utility loss due to CKM can be mitigated substantially.

47. Even though we are confident that for the Census 2021 a much more harmonized approach will be used by many EU Member States, there are still some issues that deserve attention when using perturbative methods.

48. Firstly, the connection with other data releases. In many countries, not only will European census hypercubes be published, but standard outputs such as national census tables and other demographic tables as well. These additional outputs are obviously linked with the European census hypercubes if they have the same reference day as is the case in a number of countries. Whenever the national tables and the other demographic tables are protected with different methods (i.e., not by applying TRS or CKM), the different releases could lead to the undesired situation in which tables, which *on their own are correctly protected*, could be combined to disclose individual information. A straightforward idea to try to circumvent this problem would be to use the same methods in other releases as well. To be more precise, when using TRS the same perturbed microdata set should be used (at least, the same swaps should be present) and when using CKM the same record keys should be used. This might be possible as long as the additional output is purely based on the same microdata. However, for certain additional releases, census data is combined with non-census data. This would make it difficult to correctly apply the original perturbations.

49. Nowadays it is becoming more and more popular for (accredited) researchers to conduct analyses on microdata sets that are made accessible by NSIs. For them it would be difficult to protect their output with TRS and CKM. These methods are aimed at protecting frequency count tables. However, researchers do not necessarily produce that kind of output: they may consider more complex model estimation or combine the census data with other data to produce magnitude tables.

50. A second issue concerns the communication of the perturbed results. For data users, it should be made clear that the published tables are still “valid” tables. For the general user, the publication of non-additive tables should be explained. This could be done in a similar fashion as explaining that tables with rounded figures are sometimes non-additive. For the more experienced users of published data, the added uncertainty due to TRS and CKM should be quantified. However, as discussed in Enderle et al. (2020), knowledge about the maximum perturbation in CKM could lead to increased disclosure risk.

51. Not only should the utility be explained, it should also be clear to the general public that the perturbed tables are indeed protecting their privacy. NSIs themselves do not yet know how to choose the parameters to balance the risk and the utility in an optimal way. This shows that it will be even more difficult for the general user to grasp the idea that the perturbed tables are still useful while at the same time the publication of these tables does not breach their privacy.

52. A third issue concerns the choice of the method parameters. As discussed in section IV, some NSIs have recently tried to assess this issue. Eurostat has also contributed to this discussion (Bach 2021). In our view, the exact choice will depend on the local situation of the NSIs: Member State population size, cultural differences, etc. To aid future assessments we would like to draw attention to two recent publications. The first is “How to select noise parameters for the Cell Key Method?” by Giessing et al. (2021). This paper specifically aims at quantifying the (remaining) disclosure risk. A second paper by Ricciato et al. (2021) discusses an open-source tool for experimenting with noise-based perturbation schemes, specifically aiming at quantifying the utility of the output.

53. Despite the above mentioned issues we still think that the proposed harmonization of the protection of the European Census 2021 hypercubes is a big step forward. It is the best one can do when there is no possibility to prescribe mandatory use of certain protection methods.

References

Bach, F., 2021, “Differential Privacy and Noisy Confidentiality Concepts for European Population Statistics”, in *Journal of Survey Statistics and Methodology*, 2021 (smab044, https://doi.org/10.1093/jssam/smab044).

Enderle, T., Giessing, S. and Tent, R., 2020, "Calculation of risk probabilities for the cell key method," in J. Domingo-Ferrer and Muralidhar, K. (Eds.), *Privacy in Statistical Databases*, pp. 151−165. New York: Springer-Verlag, LNCS, volume 12276.

European Commission, 2008, Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses. *Official Journal of the European Union*, L218, pp. 14−20.

European Commission, 2017a, Commission Implementing Regulation (EU) 2017/543 of 22 March 2017 laying down rules for the application of Regulation (EC) No 763/2008 of the European Parliament and of the Council on population and housing censuses as regards the technical specifications of the topics and of their breakdowns. *Official Journal of the European Union*, L78, pp. 13−58.

European Commission, 2017b, Commission Regulation (EU) 2017/712 of 20 April 2017 establishing the reference year and the programme of the statistical data and metadata for population and housing censuses provided for by Regulation (EC) No 763/2008 of the European Parliament and of the Council. *Official Journal of the European Union*, L105, pp. 1−11.

European Commission, 2017c, Commission Implementing Regulation (EU) 2017/881 of 23 May 2017 implementing Regulation (EC) No 763/2008 of the European Parliament and of the Council on population and housing censuses, as regards the modalities and structure of the quality reports and the technical format for data transmission, and amending Regulation (EU) No 1151/2010. *Official Journal of the European Union*, L135, pp. 6−14.

European Commission, 2018, Commission Implementing Regulation (EU) 2018/1799 of 21 November 2018 on the establishment of a temporary direct statistical action for the dissemination of selected topics of the 2021 population and housing census geocoded to a 1 km^2 grid. *Official Journal of the European Union*, L296, pp. 19−27.

Fraser, B. and J. Wooton, 2006, "A proposed method for confidentialising tabular output to protect against differencing," in *Monographs of Official Statistics*. Work session on Statistical Data Confidentiality, Eurostat Office for Official Publications of the European Communities, Luxembourg, pp. 299−302.

Giessing, S., Enderle, T. and Tent, R., 2021, *How to Select Noise Parameters for the Cell Key Method?*, presented at NTTS 2021, 9 – 11 March 2021, extended abstract at https://coms.events/NTTS2021/data/x_abstracts/x_abstract_10.docx.

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Schulte Nordholt, E., Spicer, K., de Wolf, P.P., 2012, *Statistical Disclosure Control*. Wiley series in Survey Methodology, John Wiley & Sons, Ltd, ISBN: 978-1-119-97815-2.

Ricciato, F., Stocchi, M., Bach, F., Bujnowska, A. and Kloek, W., 2021, *An open-source tool for experimenting with noise-based perturbation schemes*, presented at NTTS 2021, 9 – 11 March 2021, extended abstract at https://coms.events/NTTS2021/data/x_abstracts/x_abstract_105.pdf.