UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS
Expert Meeting on Statistical Data Confidentiality
26-28 September 2023, Wiesbaden

An overview of used methods to protect the European Census 2021 tables

Marieke de Vries*, Manca Golmajer**, Reinhard Tent***, Sarah Giessing*** and Peter-Paul de Wolf*
* Statistics Netherlands, ** Statistics Slovenia, *** Destatis

mm.devries@cbs.nl, Manca.Golmajer@gov.si, Reinhard.Tent@destatis.de, Sarah.Giessing@destatis.de, pp.dewolf@cbs.nl

Abstract

In the European projects “Harmonized protection of census data in the ESS” and “Open source tools for perturbative confidentiality methods” it was suggested to use Targeted Record Swapping and/or the Cell Key Method to protect the census 2021 tables. At the end of 2022 the first tables were sent to Eurostat, containing information at the level of 1km x 1km grid cells. In this paper we will discuss the outcomes of a workshop dedicated to discuss the SDC methods actually used to protect the grid cells, the intended methods to be used to protect the census hypercubes and the consequences for other publications of population statistics.

1 Introduction

At its meeting in November 2022 the European Expert Group on statistical disclosure control (EG SDC) has selected topics for guidelines to be developed by a present Centre of Excellence on SDC (CoE on SDC[1]). One of the selected topics is 'SDC for population statistics'. The intended guidelines are expected to include lessons learned from application of methods developed for census 2021 data statistics. These methods include targeted record swapping (TRS) and cell key method (CKM). Moreover, the guidelines should include recommendations for improvements.

As a first step, the CoE on SDC organized a virtual workshop on Statistical Disclosure Control of Census-data in May 2023, inviting Census experts as well as SDC experts. The aim of the workshop was to share knowledge and experience on topics like:

- the status of Census grid data protection at European NSIs
- plans for the Census hypercube data protection
- coordination of Census and population data protection
- feedback from data users in countries where SDC methods have been employed while releasing Census statistics

Participants were invited to give short presentations on special problems and solutions, work in progress, interesting examples from their institute's experience on the listed topics or any other relevant topic. Additionally, a survey with questions addressing the aforementioned topics was sent out to the participants in advance. The results of that survey were shortly discussed during the workshop.

Our paper provides an overview on the survey results in Section 2 and summarizes the presentations prepared by some of the participants in Section 3. Section 4 explains special consistency issues in a situation when Census and population data protection must be coordinated. The paper concludes with a summary and outlook on plans for further work to derive the requested guidelines on statistical disclosure control for population statistics.

2 Results of the Survey

A survey was sent out to all participants of the workshop on protection of census data. Some countries did not participate, and were later asked to fill in the survey. We have received replies from 30 countries: DE, AT, SI, MT, EL, CH, HR, IS, FR, BG, FI, PL, CY, LV, IE, RO, HU, LU, EE, NL, IT, PT, SK, BE, SE, CZ, ES, DK, LT, NO. The following section consists of a short overview of the main results of the survey. For an overview of the questions asked during the survey, see Appendix 4.1.

2.1 Methods

At the time of the workshop, 11 countries had only completed the production phase for the grid data cells, and two countries had only finished the production phase for the hypercubes. Just one country had completed both. Countries were asked which methods they applied for statistical disclosure control, specifically about the use of the Cell Key Methods and Targeted Record Swapping. Not all countries applied both methods. In fact, half of all respondents do not use TRS at all, while 40% do not apply CKM. Some countries use their own methods while some do not protect their data as they are allowed by national law to release unprotected tables.

[1] The CoE on SDC is co-funded by the European Commission by means of grant agreement number 899218, 2019-BG-Methodology. For more information on this grant and the work of the CoE on SDC, please consult: https://ec.europa.eu/eurostat/cros/content/centre-excellence-statistical-methods-and-tools_en.

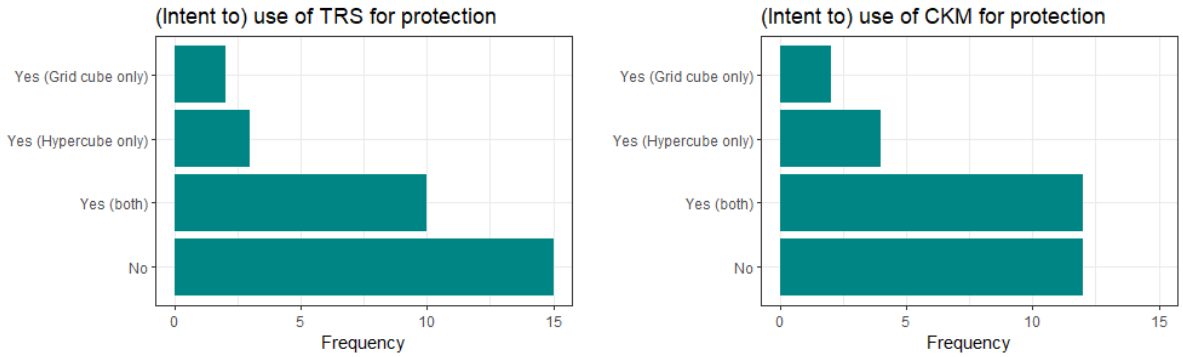

Figure 1: Use of, or intent to use, Targeted Record Swapping and the Cell Key Method in European countries.

Figure 2 shows that use of the dedicated statistical disclosure control software mu-argus and tau-argus are the least popular for testing and/or production. The R-packages *sdcMicro*, *RecordSwapping*, *ptable* and *CellKey* are more popular, for both testing and production. Some countries used none of the aforementioned methods, or used other tools as supplements: multiple countries mentioned having implemented TRS or a comparable method using SAS, or having built their own R scripts and packages for CKM.

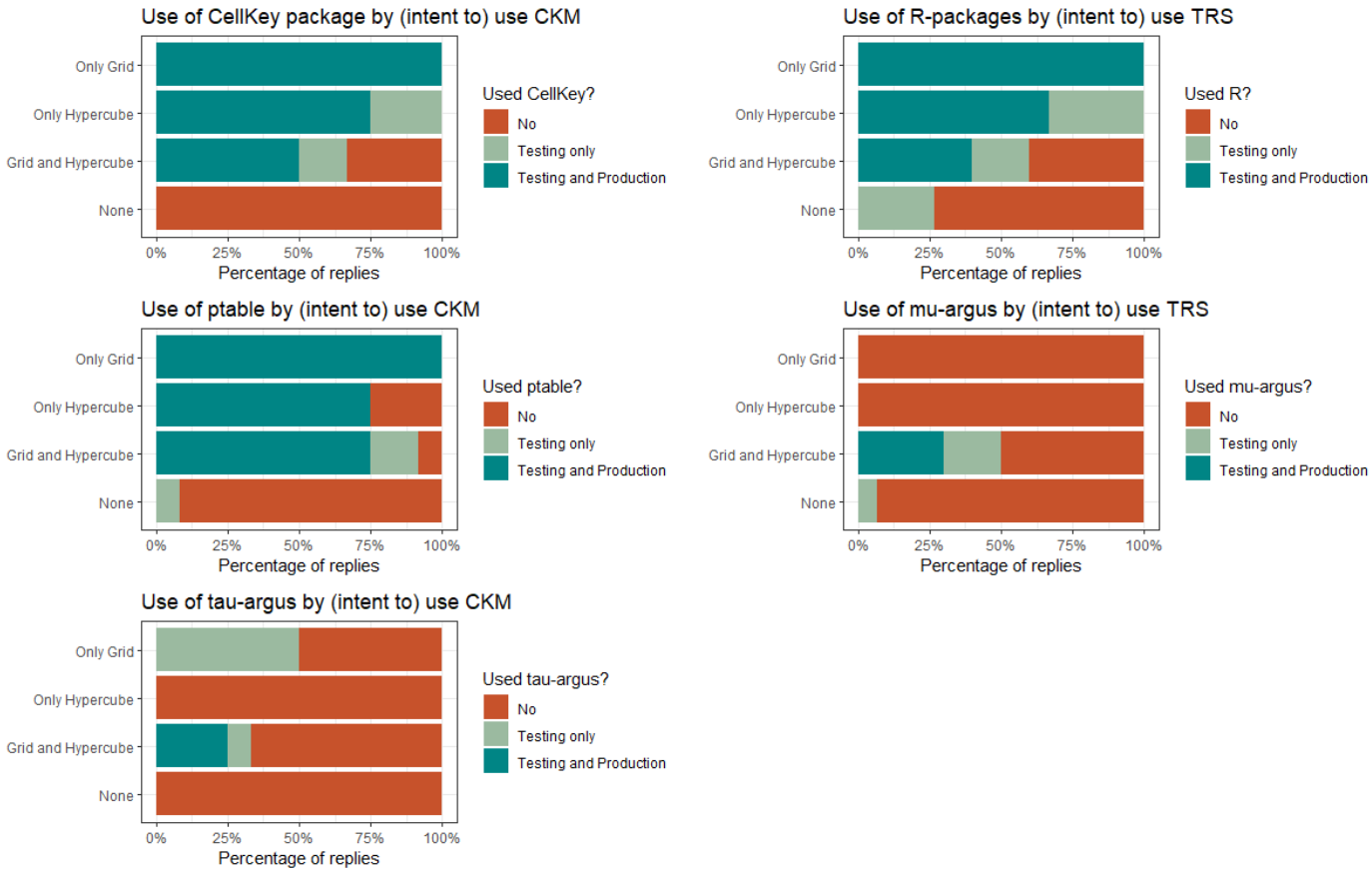

Figure 2: Use of software for testing and production of grid data and/or hypercube, differentiated by whether countries had previously used, or intended to use in the future, either the Cell Key Method or Targeted Record Swapping.

Some examples from countries using other methods include:

- Norway is not using either CKM or TRS, but making use of their own methodology: *Small Count Rounding*. This method can be used with the R-package SmallCountRounding, created and used by Norway[2].
- Belgium is also using their own methodology called *geographical perturbation*, implemented in SAS. Here grid cells are protected based on the aggregated data by virtually moving households from cells that need to be

[2] For more information on Small Count Rounding, see: https://cran.r-project.org/package=SmallCountRounding

protected to another geographical location for which the cell does not need to be protected. Restrictions are placed on which relocations can be made, such as always moving within the same municipality and if possible the same neighbourhood.

- Spain has applied a rounding method for their national dissemination where the data is randomly rounded to multiples of three, and is considering the same method for Eurostat delivery.

When asked about the methods with which countries found suitable parameters, most used a combination of methods. Data-driven experimentation was the most popular.

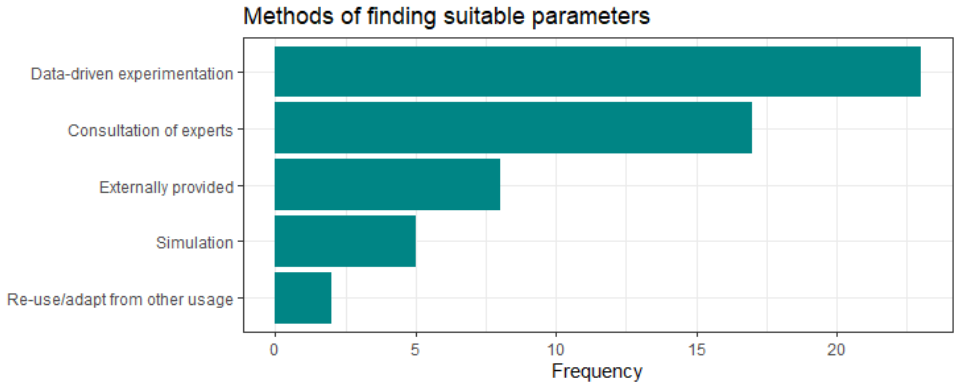

Figure 3: Overview of which methods are used when finding suitable parameters for Cell Key Method and Targeted Record Swapping.

2.2 Communication

At the time of the survey, not all countries had yet finished their strategy for communication. Most of the replies in 'Other' are because of this. Seven of the 30 countries offer user specific explanations of the protection method, with different levels of depth or detail for different user group.

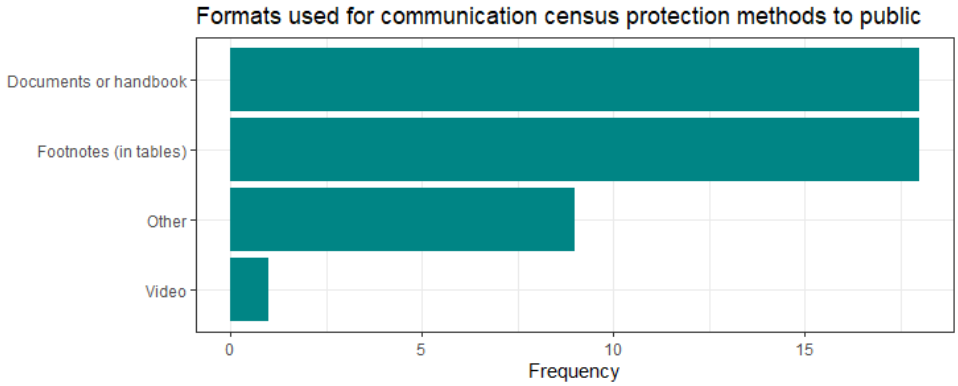

Figure 4: Formats used for communication to the public about statistical disclosure control methods used in the protection of census data.

Only three countries partially release their parameters: Luxembourg released the maximum noise and variance for CKM, Portugal released the risk threshold for TRS and France is considering releasing their swapping rate for TRS. All other countries either do not apply the methods or decided not to make the parameters public. Some countries have committed to publishing details about the used protection methods, but do not include any parameter choices in this communication. Germany has published quality indicators.

About three quarters of the countries (23 respondents) were aware of the recommendations on communication offered by the Centre of Excellence on Statistical Disclosure Control, of which 10 had applied them. Feedback on the protection methods has been sparse so far; the few comments that have been received by countries are mostly apprehensive or negative about the distortion of the data, although some feedback has been positive.

2.3 National Publication

Not all respondents knew whether the level of detail between their national census publication and the EU census hypercubes and grid data was different: 23% replied they were unsure. Most countries (60%) however said there was a difference in level of detail.

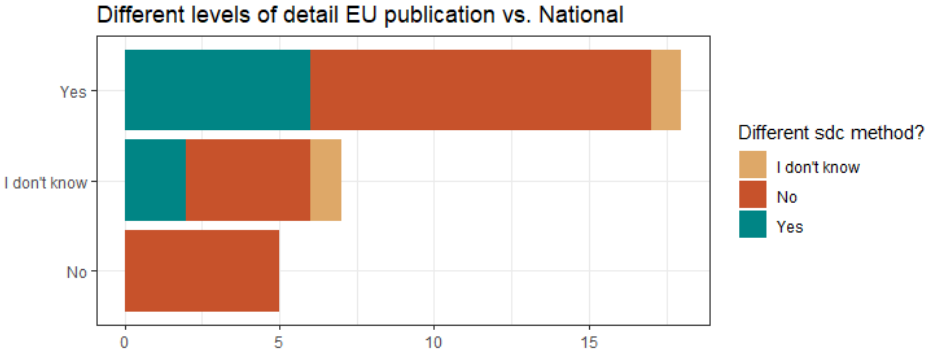

Figure 5: Response to the questions "Is the level of detail different between your national Census publication and the EU Census hypercubes and Grid Data?" and "Do you use different protection methods for your national publication and EU Census Hypercubes or Grid Data?"

Of the 18 countries affirming, only six applied or partly applied different methods for the protection of national and EU census publications. In some cases, cell suppression is used. In others, aggregation or detail levels were not published in cases where they might create issues with disclosure. In a few cases where respondents were unsure of any difference in detail, there were still differences in protection between EU and national publications.

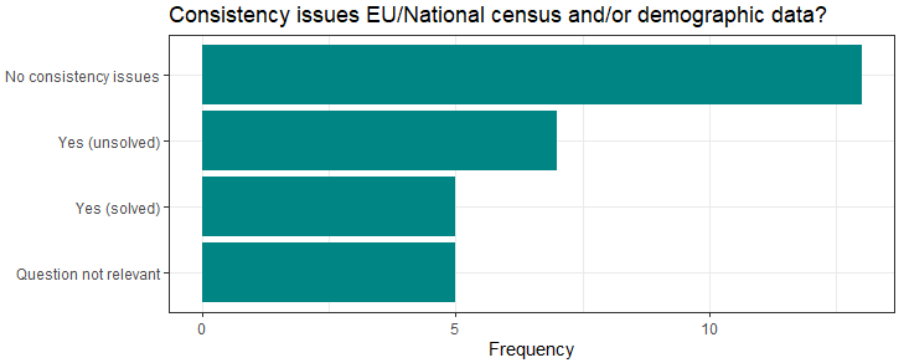

Figure 6: Responses to "Do you have encountered consistency issues regarding national and European level census and/or demographic data?"

The majority of countries do not seem to have encountered consistency issues regarding national and European level census data, or regarding census and demographic data. Five countries replied they had issues, but that those issues were solved. Another six countries do however still struggle with consistency issues. In Sections 3 and 4, we expand upon consistency issues as experienced at Statistics Slovenia.

2.4 Challenges

Countries report more challenges with the methods, concept and code development of disclosure control for census data than computational challenges.

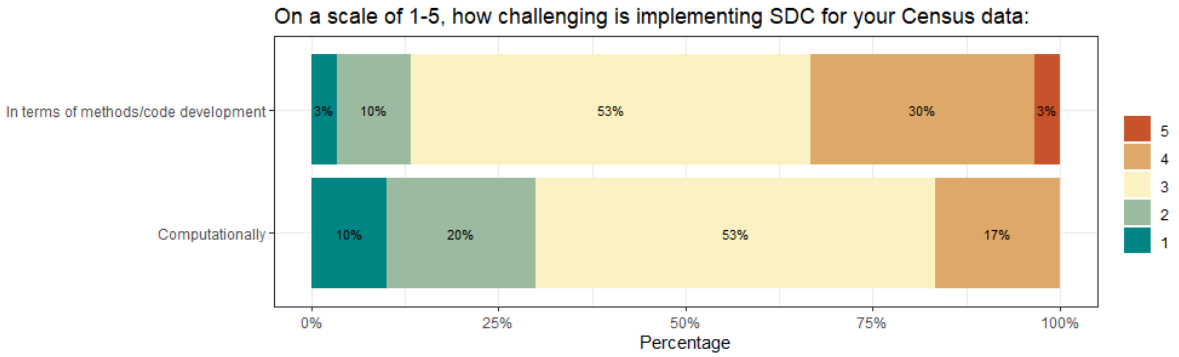

Figure 7: Answers to the questions "On a scale of 1-5, how challenging is implementing disclosure control for your Census data in term of concepts / methods / code development?" and "On a scale of 1-5, how challenging is implementing disclosure control for your Census data computationally?"

Some countries specified exactly what difficulties they faced during the process of protecting the census data. Many of the difficulties seem to stem from issues with knowing what choice of parameters is best. Some countries found this to be complicated due to issues with interpreting and comparing information loss measures for different sets of parameters. Other difficulties encountered during the process were:

- A difference in the national grid and Eurostat grid. This can cause issues with differencing, and therefore requires extra protection.
- Technical difficulties with installing packages from GitHub for testing statistical disclosure control methods.
- Restrictions imposed due to national requirements, such as population values of administrative regions needing to be published unperturbed, which complicated the implementation of statistical disclosure control methods.
- More complexity in real-world census data compared to examples found in literature.

3 Presentations during workshop: description of census data protection for a sample of European countries

During the workshop, five countries elaborated on their approach to statistical disclosure control for the 2021 Census data. In the following Section we summarise the presentations of four countries: France, Ireland, Portugal and Slovenia. Additionally, we give a summary of the protection methods used in Germany.

3.1 France

INSEE has a special method of data collection. Instead of a survey on a specific date, the French census is a rolling procedure where every year a sample of the data is collected. The EU Census release is derived from these data and is seen as an additional release for INSEE, which consists of more detailed and more sensitive data, to their national disseminations. Especially, the usage of grid data is a new type of release for French census data. The disclosure control process of these two different products is performed separately.

For grid data, INSEE uses some kind of coarsening and cell suppression technique called the 'Multilevel Grids Method', which is implemented in a package called *gridy*. For the Multilevel Grids Method the tiles that make up the grid are grouped, such that risky tiles build larger groups together with other tiles, such that these groups are above a predefined threshold. The use of several grids helps to constraint the search of partners in a certain area. The algorithm proceeds from the coarsest level (a grid which doesn't contain any tile below the threshold) to the finest level (1km or 200m tiles). In a second step each subtotal at the group level (for example the number of males) is broken down in the tiles of the group, proportionally to the population of each tile.

Opposed to that, for hypercubes Targeted Record Swapping will be used. For this technique, depending on a threshold k, a cell value is declared as risky if there are less than k observations contributing (and hence these corresponding observations need to be swapped). INSEE swaps individuals and not households, since none of the households or dwellings hypercubes contain sensitive information. Hence, there's no loss of information on

them, while swapping only individuals. This might lead to some minor inconsistencies between hypercubes on individuals and hypercubes on households, which INSEE is willing to accept.

3.2 Ireland

The Irish population census was carried out as a traditional paper census in April 2022. The Central Statistics Office (CSO) of Ireland decided to apply Targeted Record Swapping to the final microdata file as one aspect of statistical disclosure control.

The data CSO collected are the de facto count of all persons present on census night. Now TRS will be carried out, without those households which had visitors staying on census night or which had usual residents who were absent that night. For the specification of the TRS method the census team decided to swap households within a close geographical distance. The CSO didn't use an existing implementation of the TRS method but carried out the swapping in SAS in order to tailor the program to their specific needs. Unlike INSEE, CSO decided to swap entire households of the same size. However certain dwelling level characteristics remain with the geography variables e.g. it wouldn't make sense to swap detached dwelling descriptors with an apartment (such as Year Built, Water Supply or Number of Floors). For communal establishments, like nursing homes and educational establishments, the swapping algorithm was extended to include swapping individuals between establishments of the same type. Here the algorithm takes into account the gender and age of each individual to prevent illogical swaps (like a woman in an all-male prison). In contrast to INSEE, by now CSO didn't make use of *k*-anonymity to define which households need to be swapped. Instead the selection is based on a random probability sampling. To this end a seed was included at the request of the census team so they can replicate the process if needed.

Tests showed that the swapping rates of the CSO's implementation are similar to what other NSIs have reported. Yet there are future plans, to expand the program, by including *k*-anonymisation techniques to select risky households. CSO also plans to assess other software such as mu-argus and R packages.

3.3 Portugal

In particular, the loss of additivity of the tables and its high operational demands are major weaknesses of the Cell Key method. In contrast to that, being a pre-tabular method, Targeted Record Swapping only needs to be applied once to the microdata base. Since all tabulations must then be obtained from the same perturbed microdata, evaluations are cheap to perform.

Yet TRS also has its limitations. For example, frequencies relating to geographical levels higher than the geographical hierarchy considered in the application of the method (e.g. NUTS II, NUTS I and national total levels) are not disturbed - and therefore not protected. Hence Statistics Portugal has done some research to find out if it is possible to get around the limitations of TRS.

The results showed the effect of the TRS is neither evident nor substantial with respect to cells with low frequencies: depending on the frame and the municipality under analysis, the TRS can lead to a decrease, maintenance or even an increase in the number of cells with low frequencies. It turns out, however, that as a rule, the number of cells with actual low frequencies (i.e. whose frequency is also low in the original data) is always smaller than the number of cells with low frequencies in the perturbed data. This fact reflects the main effect of the TRS regarding data protection: the increased uncertainty associated with the analysis of the information provides, in itself, some degree of protection.

With regard to the risk of attribute disclosure, Statistics Portugal finds that in the generality of the tables this risk is already very low from the outset. TRS had no significant effects on this type of disclosure in the analysed tables and municipalities.

3.4 Slovenia

Unlike the other NSO's presented here, the Statistical Office of Slovenia (SURS) already delivered their grid data as well as the annual demographic tables to Eurostat. For protection of grid data, the Cell Key Method is

used and implemented in R. Record keys are generated in R using the *runif* function (random uniform distribution) and the perturbation table is generated in R using the *ptable* package. The record keys and the perturbation table are the same for different grids. Grid data for Eurostat have already been sent to Eurostat. The values and the protection are final for all 13 statistics on 1 km^2 grids.

Slovenia also publishes their own grid data in a special application called STAGE. These include several statistics for grids with length size 1 km, 500 m, and 100 m. The 1 km^2 grids for SURS do not perfectly overlap with the 1 km^2 grids for Eurostat, as they have different boundaries. However, the disclosure risk due to differences between grids for SURS and for Eurostat is very low.

SURS has also already published their own demographic tables. In general, tables at a higher geographical level can be more detailed and they can have more dimensions. Some parts of tables are suppressed on the whole (e.g. if a settlement has 1-9 inhabitants, then information about their gender and age is suppressed regardless of the cell frequencies). Secondary suppression is also used (e.g. if only one settlement in a municipality is suppressed, then another settlement is chosen for secondary suppression). In comparison to the previous census, some additional protection was done: this was done mostly by aggregation, and in some cases by suppression.

Annual demographic tables for Eurostat have already been sent to Eurostat. No protection is used, taking into account how detailed these tables are. Protection of annual demographic tables for Eurostat and tables for SURS is consistent.

Hypercubes for Eurostat haven't been protected yet. The hypercubes are very detailed, so they need additional protection. Some parts of hypercubes are the same as in tables for SURS or annual demographic tables for Eurostat. For the parts of hypercubes that are more detailed, CKM has already been tested. SURS found that only applying CKM is not enough and can lead to disclosure.

3.5 Germany

The Federal Statistical Office of Germany (Destatis), together with the State Statistical Offices, decided to solely rely on the Cell Key Method to protect all of their disseminated census data. Since data shall be disseminated via an online database and since CKM is a post tabular disclosure control method, there is the necessity for tools to access the original microdata, build requested tables and to apply perturbation, before disseminating the data. Hence, Destatis developed a java application which performs the actual perturbation and runs on a dedicated server. Now, this application cannot only be called from the public census database, but can also be accessed by authorized users from statistics software like SAS or R, for which suitable interfaces have been created. Having such a centrally managed tool and interfaces guarantees that every perturbation will be performed consistently and with the same parameters.

Additional features to the method have been added in order to e.g. improve the accuracy of perturbed ratios and in order to deal with disclosure issues.

The reference date for the data collected by the German Federal and state Statistical Office is May 15th 2022, whereas the reference date for the EU census tables is December 31st 2021. This means the data have to be backcasted. To this end, monthly migration data from demography statistics are used. Unfortunately, this leads to the problem, that the reference date for the EU census data coincides with the reference date of the German annual demography tables. So far, for those tables it was not intended to apply the Cell Key Method. What impact this will have on disclosure risks is an open issue at the moment, which still needs to be investigated.

4 The issue of annual demographic tables and their connection to Eurostat Census Hypercubes

In addition to the Eurostat Census Hypercubes, each European country disseminates annual demographic tables. Since the SDC method used for an NSI’s annual demographic tables might be different from the methods recommended by Eurostat for census data, this might lead to some issues concerning disclosure risk.

Specifically, these issues occur when the exact breakdown combinations are published for the same reference time in both statistical products.

Imagine for example that for the annual demography population stock tables, only coarsening is used. Hence, all the published data are original data. Furthermore, assume the Cell Key Method is used for protecting the census tables. If now there are characteristic combinations that occur in both tables, one can directly compare the original data with the perturbed ones and by this retrieve additional information about the perturbation parameters (like maximum possible deviation). An attacker could use this knowledge when attempting to disclose other perturbed values in the Census disseminations as well, particularly very detailed cells with low person counts.

One country that had to deal with this issue is Slovenia. Slovenia has already sent their annual demographic tables to Eurostat. Here no protection is used, taking into account how detailed these tables are. On the other hand, hypercubes for Eurostat have not been protected yet. These hypercubes are very detailed, so they need additional protection. Some parts of the hypercubes are identical to tables for Statistics Slovenia or annual demographic tables for Eurostat; for these parts, the same protection will be used in the interest of consistency. See Appendix 3.2 for an overview of tables that share variables. The use of CKM for the parts of the hypercubes that are more detailed was tested. The conclusion was that CKM only is not enough and can lead to disclosure, so another method besides CKM will be considered (e.g. TRS).

The Working Group on Population and Housing Censuses has covered this issue before. In a paper on the Harmonised protection of census data from May 2022[3], usage of TRS, CKM and their combination is analysed. Eurostat recommends the usage of both TRS and CKM and, if external consistency is to be established, as a practical solution, it is recommended to impute the overlapping cell values from the annual tables into the census hypercubes.

5 Summary and Outlook

In the Census 2021 it was advised to apply two methods: TRS and CKM. This paper has provided a first analysis of results of a survey sent out to participants of a workshop on protection of census data. The results prove that in half of the 30 respondent's countries TRS is used (or intended to be used) on at least one of the Census data products. For CKM this holds for even 60%. To find out, how many countries use (or plan to use) the recommended combination of both methods, a more detailed analysis of the responses will be needed. But clearly, it is the first time that so many European countries apply these techniques, and will thus be able to deliver the harmonized, detailed outputs requested by the European regulations while still addressing disclosure control in a way considered appropriate in each individual country.

Other interesting results of the survey are that the implementations of CKM and TRS as R-packages were quite popularly used, for both testing and production. Still, although supported by these (or other) packages, participants consider method and code development the most challenging part of implementing SDC for their Census data. Regarding communication with the users of the data, most countries will rely on documents and footnotes. As of yet, not much user feedback has been received.

Looking at more detailed reports from some countries presented in section 3, an interesting observation is that although two countries may 'nominally' use the same SDC method, implementation may be quite different, not only because different parameters are being used. For example, in France individuals are swapped and not households. In Ireland it is the other way around. Slovenia and Germany both use CKM. While in Germany CKM is applied to all Census data output products, in Slovenia it is used only to protect grid data and the European Census hypercubes. However, both countries have to deal with consistency issues regarding demography data released for the exact same reference date.

[3] *Harmonised protection of census data - issues for 2021*, Doc. CENS/2022/May/10 prepared by Eurostat for the meeting of the Working Group on Population and Housing Censuses, 16-17 May 2022

Now that we approach the publication of all high dimensional Census 2021 tables, it is time to think about the follow-up. Based on the shared experiences so far, the intended guidelines on ‘SDC for population statistics’ should discuss the lessons learned and make suggestions for improvements of the applied methods. Moreover, they should discuss whether the new techniques and experiences can also be used for other statistics. It is logical in particular to continue the current work in the future ESOP (European Statistics on Population and Housing) context where census and demography are combined. Especially the current inconsistencies in census and demographic statistics, and their protection, should be addressed. Guidelines how to apply the new protection methods might also be helpful for several future statistics.

6 Appendix

6.1 Survey Workshop on Statistical Disclosure Control of Census data

Question	*Options*
Which country are you joining us from?	*open question*
Methods: How sure are you about the method(s) you are going to use?	1-5
Production: Has the production phase of the protected tables (Census Hypercubes or Grid Data) - to be delivered to Eurostat - been completed?	Yes (Census Hypercubes only) Yes (Grid Data only) Yes (Census Hypercubes and Grid Data) No
Methods: Do/Did you use (or intend to use) TRS?	Yes (Census Hypercube only) Yes (Grid Cube only) Yes (both) No
Methods: Do/Did you use (or intend to use) CKM?	Yes (Census Hypercube only) Yes (Grid Cube only) Yes (both) No
Tools: Do/Did you (intend to) apply TRS using Mu-Argus?	Yes, for testing Yes, for production Yes, for testing and production No
Tools: Do/Did you (intend to) apply TRS using the R-packages (sdcMicro or recordSwapping)?	Yes, for testing Yes, for production Yes, for testing and production No
Tools: Do/Did you (intend to) apply CKM using Tau-Argus.	Yes, for testing Yes, for production Yes, for testing and production No
Tools: Do/Did you (intend to) apply CKM using the R-package cellKey?	Yes, for testing Yes, for production Yes, for testing and production No
Tools: Do/Did you (intend to) use the R-package ptable to design the CKM parameters?	Yes, for testing Yes, for production Yes, for testing and production No
Tools: Do/Did you use another (freely) available tool?	*open question*

Tools/Methods: How did/will you find suitable parameters?	data-driven experimentation simulation consultation of experts externally provided (from literature, national law) re-use / adapt from usage in other statistics
Tools/Methods: Did you or your colleagues participate in the ESTP Course "Statistical disclosure control (SDC) methods and tools for census 2021"?	Yes No I/We didn't know about that course.
Communication: Which format(s) or channel(s) do you use to communicate Census data protection methods to the public?	footnotes (in tables) documents or handbooks wiki videos podcasts other
Communication: Please specify "other" if you have used other format(s)/channel(s).	*open question*
Communication: Do you offer user specific explanations of the protection method (with different levels of depth or detail for different user groups)?	Yes No
Communication: Do you make the parameters of your SDC methods available to the public?	Yes, fully Yes, partially Yes, to selected user groups only No
Communication: Which of the TRS parameters are or will be made public?	[TRS] swapping rate [TRS] k-anonymity [TRS] risk threshold None of the mentioned parameters We don't apply TRS
Communication: Which of the CKM parameters are or will be made public?	[CKM] Maximum Noise [CKM] Variance [CKM] Probabilities for no change [CKM] ptable None of the mentioned parameters We don't apply CKM
Communication: What other parameters or further information do you publish?	*open question*
Communication: Do you know about the recommendations offered by the Centre of Excellence on Statistical Disclosure Control?	Yes (we applied the recommendations) Yes (but we elaborated our own different communication strategy or applied another methods)

	No
Communication: Did you get any feedback from prospective (external) user groups for your planned protection method(s)?	*open question*
Communication: How was the feedback?	1-5
National Publication: Is the level of detail different between your national Census publication and the EU Census hypercubes and Grid Data?	Yes No I don't know
National Publication: Do you use different protection methods for your national publication and EU Census Hypercubes or Grid Data?	Yes (please specify in the next question) No I don't know
National Publication: Please, specify why you have used or will use different methods.	*open question*
National Publication: Do you have encountered consistency issues regarding national and European level census and/or demographic data?	Yes (unsolved) Yes (solved) No consistency issues Question not relevant
Experience: On a scale of 1-5, how challenging is implementing disclosure control for your Census data in term of concepts / methods / code development?	1-5
Experience: On a scale of 1-5, how challenging is implementing disclosure control for your Census data computationally?	1-5
Experience: If there are difficulties encountered in practice for Census SDC, please tell us.	*open question*

6.2 Connection between hypercubes for Eurostat and annual demographic tables for Eurostat

List of hypercubes can be found in explanatory notes on EU legislation on the 2021 population and housing censuses (2019 edition)[4]. Hypercubes are divided into 41 groups. They contain:

1. **Groups 1-32: Total population**
2. Groups 33, 35: Number of all private households
3. Groups 34, 36: Number of all families
4. Groups 37, 38: Number of all conventional dwellings
5. Groups 39, 40: Number of all occupied conventional dwellings
6. Group 41: Number of all living quarters

List of annual demographic tables can be found in files Eurostat_Demography_UNIDEMO1260_RY2020.xls and Eurostat_Demography_UNIDEMO862_RY2020.xls. Annual demographic tables are divided into 12 groups:

1. **Group P: Population**
2. Group SLA: Surface
3. Group B: Live births

[4] File KS-GQ-18-010-EN-N.pdf, p. 170-175, https://ec.europa.eu/eurostat/documents/3859598/9670557/KS-GQ-18-010-EN-N.pdf/c3df7fcb-f134-4398-94c8-4be0b7ec0494

4. Group L: Legally induced abortions, late foetal deaths
5. Group D: Deaths
6. Group X: Infant deaths
7. Group M: Marriages
8. Group S: Divorces
9. Group I: Immigrants
10. Group E: Emigrants
11. Group A: Acquisition of citizenship
12. Group LS: Loss of citizenship

Tables can be mandatory, voluntary or partially mandatory.

Hypercubes from groups 1-32 (Total population) are linked to annual demographic tables from group P (Population). Other groups of hypercubes (33-41) are not linked to annual demographic tables (households, families, dwellings, living quarters are not sent to Eurostat every year).

The following 19 variables appear in Hypercubes from groups 1-32:

Variable content	**Variable name**	**Variable exists in annual demographic tables**
Age	AGE	YES (age)
Current activity status	CAS	NO
Country of citizenship	COC	YES (citizenship)
Educational attainment	EDU	YES (educational attainment)
Family status	FST	NO
Place of usual residence	GEO	YES (region of residence)
Housing arrangements	HAR	NO
Household status	HST	NO
Industry (branch of economic activity)	IND	NO
Legal marital status	LMS	YES (legal marital status)
Size of the locality	LOC	NO
Location of place of work	LPW	NO
Occupation	OCC	NO
Country/place of birth	POB	YES (country of birth)
Place of residence one year prior to the census	ROY	NO
Sex	SEX	YES (sex)
Status in employment	SIE	NO
Ever resided abroad and year of arrival in the country since 1980	YAE	NO
Ever resided abroad and year of arrival in the country since 2010	YAT	NO

Below, variable names from hypercubes are used also for variable names from annual demographic tables because of simplicity, although variables from annual demographic tables can be a little different (e.g. different categories).

Out of the 19 variables, only seven variables appear also in annual demographic tables: GEO, SEX, AGE, COC, POB, EDU, and LMS.

- All tables in group P have variables GEO, SEX, AGE. Only P07 does not have SEX and AGE.
- Only tables in group P that have GEO.N. (only country level) can have other variables besides SEX and AGE:
    - Variable COC appears only in tables P03, P04, P07.
    - Variable POB appears only in tables P05, P06, P07.
    - Variable EDU appears only in table P09.
    - Variable LMS appears only in table P08.
- The lowest geographical level in group P is in table P02, which is similar to GEO.M. x SEX. x AGE.M.

In order to check linked protection between a group of hypercubes and annual demographic tables, it is not necessary to check all hypercubes and all tables in group P, it is enough to check the combinations below:

Group of hypercubes	**Hypercubes for linked protection**	**Annual demographic tables for linked protection**
1	1.1	P08
2	2.1	P02, P08
	2.3	P02
3	3.1	P02
	3.3	P02, P08
4	4.3	P01, P09
5	5.3	P02, P09
6	6.2	P02, P09
7	7.2	P02
8	8.1	P02, P03
	8.2	P02, P05
9	9.1	P03, P05, P07
	9.3	P02, P05
10	10.1	P02, P03
	10.2	P02, P05
11	11.1	P02, P03
12	12.2	P02, P05
	12.3	P02, P03, P05, P07
13	13.1	P02, P05
	13.2	P02
	13.3	P02, P05
	13.4	P02, P03
	13.5	P01, P07
14	14.1	P02, P03

	14.2	P02, P05
15	15.1	P02, P05, P09
	15.4	P02, P03, P05, P07
16	16.1	P02, P03
	16.2	P02, P05
17	17.1	P02, P03
18	18.2	P02, P05, P09
	18.3	P02, P03, P05, P07
19	19.1	P02, P05, P09
20	20.1	P02, P03
	20.2	P02, P05
21	21.2	P02, P08, P09
22	22.1	P02, P09
23	23.2	P02, P09
24	24.2	P02, P08
25	25.1	P02, P05, P08
	25.2	P02, P03, P08
26	26.3	P02, P03, P05, P07
27	27.1	P02
28	28.1	P02
29	29.1	P02, P05, P08
30	30.1	P02, P03, P08
31	31.1	P02, P05
	31.3	P02, P05, P09
32	32.1	P02, P03
	32.3	P02, P03, P09