UNECE

CONFERENCE OF EUROPEAN STATISTICIANS

Expert Meeting on Statistical Data Confidentiality

1-3 December 2021, Poland

Differential privacy and noisy confidentiality concepts for European population statistics.

Fabian Bach (Eurostat))
fabian.bach@ec.europa.eu

Abstract

First the paper delimits various statistical confidentiality concepts that often get mixed up, such as risk measure, noise distribution, and output mechanism. Then some noise distributions and output mechanisms currently discussed for population statistics and censuses are analysed from typical risk and utility perspectives. We also present methods to infer quantitative limits on generic noise parameters, like noise variance and magnitude bound. The paper finally notes that strictly differentially private approaches appear over-constrained by such a combined risk/utility analysis in typical census scenarios like EU 2021, i.e. there is no straightforward parameter choice to simultaneously ensure acceptable risk and utility properties of the statistical output.

Differential privacy and noisy confidentiality concepts for European population statistics

Fabian Bach*

* Population and migration statistics unit, European Commission, Eurostat, L-2920 Luxembourg, Fabian.BACH@ec.europa.eu.
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.

Abstract. First the paper delimits various statistical confidentiality concepts that often get mixed up, such as risk measure, noise distribution, and output mechanism. Then some noise distributions and output mechanisms currently discussed for population statistics and censuses are analysed from typical risk and utility perspectives. We also present methods to infer quantitative limits on generic noise parameters, like noise variance and magnitude bound. The paper finally notes that strictly differentially private approaches appear over-constrained by such a combined risk/utility analysis in typical census scenarios like EU 2021, i.e. there is no straightforward parameter choice to simultaneously ensure acceptable risk and utility properties of the statistical output.

1 Background

This paper addresses noise-based approaches to statistical confidentiality in official population statistics, including a differential privacy (DP) angle. While there is ample literature introducing DP (e.g. (Dwork, McSherry, et al., 2006; Dwork and Roth, 2014)), a first strict line is drawn here between DP as a *risk measure*, and differentially private (noisy) *output mechanisms* that are engineered to manifestly guarantee a given DP level. However, many other noisy output mechanisms, using bounded or unbounded noise distributions, can be set up to give at least a relaxed DP guarantee too (Dwork, Kenthapadi, et al., 2006; Rinott et al., 2018). For instance, the cell key (CK) method originally proposed by Fraser and Wooton (2005), Marley and Leaver (2011), and Thompson, Broadfoot, and Elazar (2013) can be turned into a (relaxed) DP mechanism (Bailie and Chien, 2019). On the other hand, strictly DP output mechanisms require unbounded noise distributions with infinite tails, which may have particularly negative effects on utility.

This paper aims to first address all these different notions separately, and then to present a consolidated discussion from both risk and utility perspectives. A focus on census-like statistics is chosen because of the current global relevance (2020/2021 census round), and because unweighted counts simplify technical discussions without major loss of generality in the key issues. In the context of some noisy approaches, e.g. based on DP as for the 2020 U.S. census (Abowd, 2018) or on the CK method recommended for the 2021 EU census (Antal et al., 2017), we also outline some analysis that may help setting up noisy mechanisms and parameters for particular output scenarios of official population or census statistics.

2 Concepts and terminology

We give here a short summary of key concepts and terms used throughout the paper. A more comprehensive introduction to each concept is provided in annex A.

The database reconstruction theorem by Dinur and Nissim (2003) states, in a nutshell, that the size of random noise added to statistical output from a given microdata database should scale with the total output complexity (amount of information released). Otherwise there is a risk that the entire input database may be reconstructed accurately from that output; see Garfinkel, Abowd, and Martindale (2018) for an illustrative example.

Risk measures (in the sense used in statistical confidentiality) aim to quantify individual information leakage from statistical outputs. Ideally, risk measures work with minimal or no assumptions on output specifics, thus enabling wide comparability of risks across output classes and protection methods.

Differential privacy (DP) is such a powerful risk measure, first proposed by Dwork, McSherry, et al. (2006) in the wake of the database reconstruction theorem. The concept is appealing from a risk-aware view because it gives a DP guarantee to each individual contributor of a given statistic; cf. annex A.1. More specifically, Eq. (7) defines a strict ε-DP guarantee with a single privacy budget measure ε, and a relaxed (ε, δ)-DP guarantee with a second measure δ quantifying the potential leakage from a strict guarantee.

Noise distributions are probability distributions over the range of the statistical outputs of interest, e.g. non-negative integers in population counts. A noise distribution is designed as part of an output mechanism, which then uses it to draw a dedicated random noise term x to be added to each statistical value in the output. Typically, risk and utility considerations influence the design shape of the distribution, where examples in annex A.2 include manifestly ε-DP distributions and those used by the cell key method (Marley and Leaver, 2011).

Bounded noise comes from a noise distribution with a parameter $E > 0$ such that $\Pr(|x| > E) \equiv 0$, i.e. limiting the magnitude of any noise term x. Note importantly that strict ε-DP, in contrast to (ε, δ)-DP, does not allow $E < \infty$ (see annex A.2). A key goal of this paper is to quantify specific disclosure risks of bounded noise (section 3.1), but also utility problems of *un*bounded noise (section 4).

Same participants–same noise (SPSN) is a principle to decide whether the noise term added to a given output is drawn afresh from the noise distribution, or reused from a lookup table (Fraser and Wooton, 2005; Thompson, Broadfoot, and Elazar, 2013). It was introduced to forestall averaging attacks—which works only to some extent, as discussed in section 3.2.

(Noisy) output mechanisms should be thought of as complex functions taking an entire database as input and returning the complete set of publishable output statistics (e.g. a set of tables, or a sequence of custom queries). Noisy output mechanisms work by injecting random noise in one way or the other but must—importantly—also account for *noise composition* across the entire output, as explained in annex A.3 (cf. scaling law of the database reconstruction theorem). Conceptual dichotomies emerge between

- static mechanisms (entire output fixed before publication) vs. flexible mechanisms (interactive system allowing user-defined queries to some extent);
- manifestly DP mechanisms (using a manifestly DP noise distribution so that the DP composition theorem[1] plays out) vs. other mechanisms (global DP guarantee must be inferred manually[2]);
- risk-driven parametrisations (noise parameters transparently linked to risk measures, e.g. DP measures ε and δ used as parameters) vs. utility-driven parametrisations (noise parameters transparently linked to user interests, like global count-level noise variance V and bound E in the CK method).

3 Various risk aspects

While typical disclosure risks in population statistics relate to small table counts revealing unique personal characteristics, after Dinur and Nissim (2003) the discussion has partly moved to database reconstructions risks. Nevertheless, a personal data breach within a default legal understanding[3] would always entail a (reconstructed) database record to be accurately matched onto a natural person—which seems not straightforward (Ruggles et al., 2019). In any case, noise injection can mitigate these risks, where Theorem 5 of Dinur and Nissim (2003) suggests a scaling of noise size with the complexity t of the statistical output.[4] E.g. Asghar and Kaafar (2020) have recently elaborated two typical attacks exploiting fixed noise size in a flexible output

[1] Theorem 3.16 of Dwork and Roth (2014); cf. section 7 of Rinott et al. (2018) for an application to population count tables.

[2] E.g. Bailie and Chien (2019) compute the (ε, δ)-DP guarantee of a CK setup.

[3] E.g. in the EU context under Regulation (EU) 2016/679 (GDPR, OJ L 119, 4.5.2016, p. 1).

[4] This is sometimes used to argue for DP mechanisms where such scaling is built in through the composition theorem (footnote 1; cf. e.g. Asghar and Kaafar, 2020), but at least the strict ε-DP scaling seems to be overprotective (see annex A.3).

mechanism with scaling complexity. Each of these attacks is now addressed in turn, with the central point being that fixed noise of sufficient size is not a problem for fixed t, i.e. *static* output mechanisms.

3.1 Utility-driven parametrisation: bounded noise

The first attack in Asghar and Kaafar (2020) exploits extreme noise patterns in constrained n-tuples with generic bounded noise protection. Normally the noise bound E should be non-public, so the first step is to disclose the exact value of E.

Revealing the bound The attack of Asghar and Kaafar (2020) relies on m t output 3-tuples of noisy observations with independent noise but respecting a linear constraint. The type of 3-tuples is not important, e.g. a sex breakdown including total count $\{F, M, T\}$ with expectation $\mathrm{E}(F+M-T)=0$ so that $F+M-T$ values are sampling the noise distribution. This gives an estimator for the noise bound

$$\widehat{E} = \lceil \left| \frac{F+M-T}{3} \right| \rceil, \tag{1}$$

where the probability of revealing E correctly from a single 3-tuple is fixed by the noise distribution as $p_1 \coloneqq \Pr[|F+M-T| > 3(E-1)]$.[5] Given p_1, the number of independent 3-tuples needed to infer E at confidence level α is

$$m = \lceil \frac{\log(1-\alpha)}{\log(1-p_1)} \rceil \simeq \lceil \frac{1}{p_1} \rceil \quad \text{for} \quad \alpha = 68\,\% \text{ and } p_1 \ll 1. \tag{2}$$

Results of Asghar and Kaafar (2020) are for uniform noise only, but in general m will depend heavily on p_1 and thus on the particular noise distribution. For instance, in CK-like methods p_1 is fixed by the p-table (Thompson, Broadfoot, and Elazar, 2013) and thus by V and E parameters, which allows to control the required complexity m. Fig. 1 illustrates m over the typical V–E space in a generic CK setup using the p-table tool recommended for the 2021 EU census (De Wolf et al., 2019b).[6]

Note that m converges to the uniform limit for increasing $V > E$ (because the p-table converges to the uniform distribution with maximum variance $V = E(E+1)/3$), but diverges quickly for decreasing $V < E$ (because large noise magnitudes become increasingly unlikely). This suggests that CK setups with moderately large $E \lesssim 10$ and considerably smaller V (e.g. $E = 5$ to 10 and $V = 2$) perform as "quasi-unbounded" noise on attempts to disclose E. In conclusion, Asghar and Kaafar (2020) have argued that E cannot be sufficiently protected, but it was shown above that this depends critically on the noise distribution and relative choice of V and

[5] E.g. uniform noise $\in \{-E, E\}$ gives $p_1 = 20/(2E+1)^3$ by simple combinatorics (Asghar and Kaafar, 2020).

[6] The setup is 'generic' because we use the implemented generic p-table generating algorithm that maximises entropy under the sole constraints of fixed V and E (cf. Giessing, 2016). If p-tables are further tailored to specific needs, e.g. adding more constraints, this may affect p_1 and thus m.

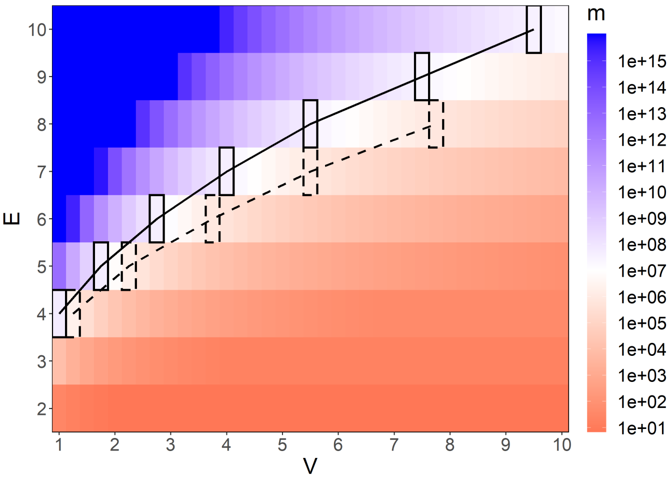

Figure 1: Heat map showing the number m of 3-tuples required to infer E at confidence level $\alpha = 68\,\%$ over the V–E parameter space of CK-like methods (and $p_1 \simeq 1/m$, cf. Eq. (2)). Black boxes highlight the parameter settings where m exceeds the number of independent 3-tuples (i.e. sex breakdowns) available in the 2021 EU census output of Germany (solid) and Malta (dashed).

E: while uniform noise seems E-disclosive, generic bounded noise distributions, manifestly (ε, δ)-DP or not, can be set up to protect E effectively while keeping strong utility guarantees (moderate variance and hard noise bound). But only if E is known accurately one can search the output for extreme noise patterns revealing true counts, e.g. $+E$ on each internal count of an n-tuple and $-E$ on its margin, where again the abundance of such patterns depends on p_1 and thus on V and E. Note finally that such an attack cannot be "aimed" at specific statistics of interest; it is limited to wherever extreme noise patterns happen to occur. Targeted attacks must pursue other strategies, addressed in section 3.2.

Heuristic parameter constraints To generalise this, a heuristic risk constraint can be inferred on the V–E parameter space: to avoid E-disclosure, choose V and E for fixed m (i.e. static output mechanism) such that the E disclosure risk according to Eq. (2) is below $68\,\%$, even when all available 3-tuples are used. The respective contours are added to Fig. 1 for Germany (most independent sex 3-tuples) and for Malta (fewest independent sex 3-tuples). Such a limit can always be set by requiring the noise distribution to satisfy $p_1 \lesssim 1/m$ (static output property). Note that this constraint is very conservative: even if E is disclosed correctly, there are then only $mp_1 = \mathcal{O}(1)$ 3-tuples in the whole output where the noise can be removed.

3.2 All noisy mechanisms: Massive averaging

The second attack of Asghar and Kaafar (2020) aims to remove the noise successively from certain sets of table cells ('histogram reconstruction'). The concept underlying such attack classes is *massive averaging*, to which any noise method with constant variance is susceptible. This is illustrated by the (inverse) Chebyshev inequality:

$$\Pr(|\overline{x}| < \xi) \geq 1 - \frac{\kappa \mathrm{Var}(x)}{\xi^2 t}, \tag{3}$$

	With SPSN			Without SPSN		
Statistic	t	k (DE)	k (MT)	t	k (DE)	k (MT)
total	2775	8.6×10^7	2.1×10^6	6378	9.7×10^7	2.6×10^6
SEX	1376	2.8×10^7	7.1×10^5	3105	3.2×10^7	8.4×10^5
AGE.M	926	3.0×10^6	7.9×10^4	2281	3.4×10^6	9.7×10^4
GEO.L	766	5.4×10^5	5.4×10^5	1734	6.6×10^5	6.6×10^5
SEX×AGE.M	458	9.6×10^5	2.6×10^4	1097	1.1×10^6	3.2×10^4

Table 1: The top five output statistics with the largest number t of IRRs in the 2021 EU census output: apart from the 'total' national population and its 'SEX' breakdown (into F and M as in section 3.1), 'AGE.M' is the medium-detail age breakdown (by five-year bands), and 'GEO.L' is the low-detail geographic breakdown (by NUTS 2 regions); 'SEX×AGE.M' is the two-way breakdown by SEX and AGE.M. k depends on the number of territorial units in the geographic breakdowns, so we list largest (Germany) and smallest (Malta) results.

where $\overline{x}$ is the averaged (unbiased) noise over t independent observations of the same statistic of interest, $\mathrm{Var}(x)$ the variance of each individual noise term, κ a constant factor counting how many outputs had to be summed on average to obtain the statistic (e.g. $\kappa = 2$ for t bi-partitioned observations), and ξ a small parameter (e.g. $\xi = 0.5$ for accurate disclosure of an integer count). The t scaling comes from the averaging, so if $\mathrm{Var}(x)$ is constant in t, the average always becomes accurate for sufficient t. More precisely, Eq. (3) implies that the noise must scale as $\mathrm{Var}(x) \sim t$, i.e. noise of $\mathcal{O}(\sqrt{t})$, to prevent this. This is consistent with the scaling law of Dinur and Nissim (2003) (cf. section 6 of Asghar and Kaafar, 2020).

Counting redundancies Massive averaging works on redundant representations of the same target statistic in the output, but with independent noise such that $\overline{x}$ is random and Eq. (3) plays out. E.g. Asghar and Kaafar (2020) rely on t *user-defined* independent bi-partitions of the same variable. Then the noise of any target statistic broken down t times by these bi-partitions can be removed by averaging over t sums of the bi-partitions. Note that only the SPSN principle (section 2) assumed here to be present requires the use of n-partitions in the first place; without SPSN the adversary could just query t times the target statistic directly. Clearly the scenario of Asghar and Kaafar (2020) is an example of a badly curated flexible output mechanism. The situation is quite different with a *static* output mechanism: in this case the independent redundant representations (IRR) of each output statistic can be counted in advance, and curated if needed.

2021 EU census example To illustrate the above means, one can do the full IRR counting exercise on the complete 2021 EU census output. For any given output statistic targeted by averaging, Eq. (3) suggests $\kappa/t \equiv k/t^2$ as an averaging risk measure, where t is the number of IRRs being averaged and k is the total number of independent counts (i.e. noise terms) contributing to the IRR average. Annex C describes step by step how all IRRs of any statistic contained in a given output can be analysed systematically to count its t and k, with and without SPSN invoked. For illustration, Table 1 lists the top five output statistics by number t of respective IRRs available—showing that the most redundant statistics are totals and

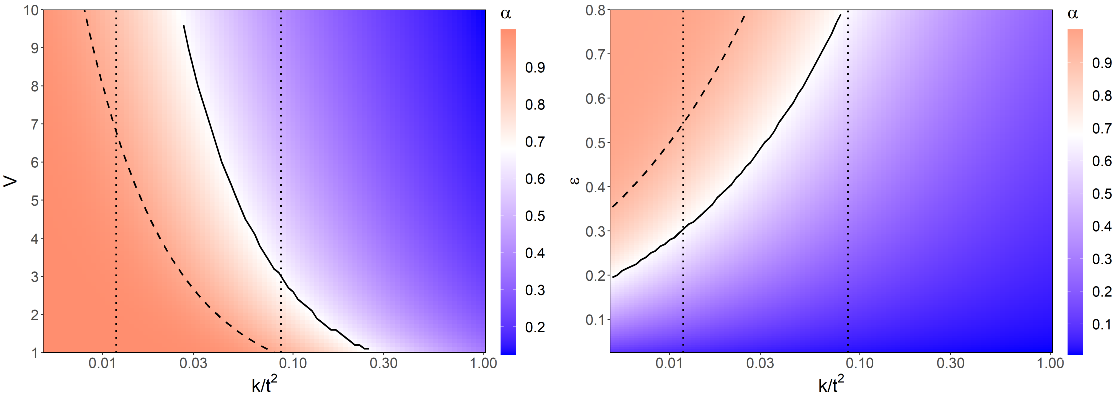

Figure 2: Heat maps of the k/t^2 vs. noise parameter space showing the Gaussian-modelled probability α of averaging a single output count correctly, for the CK variance parameter V (left) and the DP privacy budget ε per single count (right). Both plots also show the Gaussian-modelled $\alpha = 68\,\%$ contour (bold line) and Chebyshev's lower limit at $\alpha = 68\,\%$ (dashed line), as well as (dotted lines) the smallest optimised k/t^2 values found in the 2021 EU census output with SPSN ($k/t^2 \simeq 0.087$) and without SPSN ($k/t^2 \simeq 0.012$).

basic demographic/regional breakdowns, and that SPSN reduces the redundancies (both no surprise). However, the search for the most *risky* target statistics requires an optimisation step (described in annex C.3), because generally the smallest k/t^2 value is found by averaging only a subset of the available IRRs. Here we only give the smallest optimised values found in the 2021 EU census output: with SPSN it is $k/t^2 \simeq 0.087$ (for the one-way AGE.M breakdown of the Luxembourgers), and without SPSN it is $k/t^2 \simeq 0.012$ (for the total population of any Member State).

Heuristic variance constraint Fig. 2 shows the averaging success risk α, computed from a Gaussian model with variance $k/t^2 \times V$ for the averaged noise[7], as a heat map over a plane confronting the count-level variance parameters V (l.h.s.) resp. ε (r.h.s.) with the averaging risk measure k/t^2. Staying on the blue side of the $\alpha = 68\,\%$ contour means that the chance of averaging a target statistics correctly would be less than $68\,\%$, for any k/t^2 value corresponding to that target statistic. The smallest k/t^2 values found in the 2021 EU census output are drawn too, so that a lower bound on V (upper bound on ε) is given by the intersection of the $\alpha = 68\,\%$ contour with the smallest k/t^2 line (with or without SPSN). Accordingly, a CK setup with $V \gtrsim 3$ and SPSN is sufficient to reduce the averaging risk of even the most risky statistic to below $68\,\%$ per count, whereas a DP setup requires count-level $\varepsilon \lesssim 0.3$ (no SPSN). Note that these are very conservative constraints: the chances of obtaining correct averages for n counts of a target histogram (e.g. AGE.M of Luxembourg mentioned earlier) would shrink as α^{n}, and k/t^2 is generally larger. Again the analysis can be applied to any static output: just measure the most risky output statistic with k/t^2 and fix a lower variance bound as done above with Fig. 2 (or reduce complexity to increase k/t^2).

[7] This follows from the central limit theorem, and can be checked numerically by generating test samples of CK noise. It is more conservative than using Eq. (3), which only gives a lower limit on the averaging success.

4 Specific utility aspects of unbounded noise

There are already many studies assessing utility aspects of DP output mechanisms or testing them in statistical applications—e.g. Machanavajjhala et al. (2008), Dwork and Smith (2010), Ghosh, Roughgarden, and Sundararajan (2012), Hsu et al. (2014), Wang, Lee, and Kifer (2015), and Petti and Flaxman (2019). In particular, Rinott et al. (2018) is a key reference for population statistics, but all DP noise distributions there were truncated, so results do not cover tail effects from unbounded noise. On the other hand, the U.S. Census Bureau announced unbounded ε-DP noise for its 2020 census (Abowd, 2018), which triggered severe utility concerns (Ruggles et al., 2019; Santos-Lozada, Howard, and Verdery, 2020). Petti and Flaxman (2019) assessed some utility implications of published test setups, but explicitly left the issue of tail effects open. This section concentrates on such tail effects.

4.1 Parameter setups

Censuses are among the most expensive national statistical exercises, serving a variety of specific research and policy purposes, so ensuring that SDC methods maintain unique census features is critical. Therefore, in contrast to the approach of Rinott et al. (2018) (truncating DP noise and fixing reference ε and δ values for a theoretical comparison), this paper aims to assess actual parameter setups considered in a census context. From a utility perspective, noise properties at the *individual count level* are key for any kind of serious research. While utility-driven parametrisations provide this out of the box (V values and E ranges), an effort has to be made in DP setups to infer the ε spent at the individual count level: a global ε must be split across all outputs, and the methods for doing this in an optimised manner can become complex and rather non-transparent (cf. Garfinkel, 2019). Nevertheless, we attempt to make an educated guess at the ε budget spent on a single output table in the hypothetical U.S. census DP scenario described in Petti and Flaxman (2019).

According to the authors, discrete ε-DP noise is drawn from the two-tailed geometric distribution with a *global* privacy budget $\varepsilon_{\text{global}} \in \{0.25, 0.5, 1.0, 2.0, 4.0, 8.0\}$ (Garfinkel, 2019; Petti and Flaxman, 2019). This global budget is then distributed across six hierarchical geographies (Garfinkel, 2019). Certain optimisations may shift the relative shares away from an even split, but we assume 1/6 for practical purposes as Petti and Flaxman (2019) do. Further intricacies include that noisy total population counts are generated for each geographic level[8] and all further breakdowns are optimised to sum to those totals. The reference also suggests that at each geographic level, 67.5 % of the budget are spent on the more important person aggregate tables. In summary, we assume

$$\varepsilon_{\text{table}} = 67.5\,\% \times 1/6 \times \varepsilon_{\text{global}} \simeq 10\,\% \times \varepsilon_{\text{global}}, \tag{4}$$

[8] Except at State level, where the U.S. Constitution requires the U.S. Census Bureau to publish unperturbed totals (Petti and Flaxman, 2019).

so $\varepsilon_{\text{table}} \in \{0.025, 0.05, 0.1, 0.2, 0.4, 0.8\}$ for tabular (count-level) ε-DP noise. This corresponds to noise sizes at single count level of

$$V \in \{3200, 800, 200, 50, 12.5, 3.125\}, \quad \sqrt{V} \in \{56.6, 28.3, 14.1, 7.1, 3.5, 1.8\}.$$

For comparison, the CK variances tested for the 2021 EU census round are in the area $V \in [1, 5]$ (Antal et al., 2017), so barely touching the above DP range at its most risky end ($\varepsilon_{\text{table}} = 0.8$). Moreover, no tails effects $> E$ are present by definition.

4.2 Demographics at high geographic detail

Accurate demographics at a high geographic detail is one of the key unique census features in many world regions. For instance, the 2021 EU census round will cover ca. 110 000 Local Administrative Unit (LAU) with a total population of roughly 4.5×10^8 people across the whole EU.[9] Coincidentally this matches well with U.S. census outputs at tract level, covering ca. 75 000 geographic units (Garfinkel, 2019) with a total population of 3.3×10^8 people. However, the following analysis is intended solely to discuss effects of a practical ε-DP noise scenario on key EU census outputs. Whether any of the conclusions may apply to tract-level U.S. census outputs depends critically on the correctness of parameter assumptions, Eq. (4), and also on the comparability of population distributions across EU LAUs vs. U.S. tracts.

The statistics of LAUs There is an extreme variety of total population by LAU, with populated units ranging from $\mathcal{O}(1)$ residents (450 LAUs with < 10 people) to 3.3×10^6 residents (Berlin; in total 14 LAUs with $> 10^6$ people). Now the key point is that statistics *across* LAUs is only part of the purpose of these census results; they are also the only source to obtain accurate demographic information on *individual* LAUs. For this purpose, even very unlikely but very large noise outliers can have severe, maybe unacceptable, consequences. Furthermore, if the method of adjusting inner tables to their geographic totals after drawing noise is applied (Petti and Flaxman, 2019), a single large noise outlier on a given small LAU total would systematically and heavily distort all statistics published for that LAU. Therefore, the subsequent focus is on LAUs with counts < 500 illustrated in Fig. 3.

The demographics of LAUs To add a demographic element, we include a sex breakdown into females, males and a total, i.e. SEX $= \{F, M, T\}$ as in section 3.1. This is the spine of all LAU-level person tables in table groups 3 and 8 of the 2021 EU census programme[10]. It also reflects a possible notion of picking more important 'aggregate tables' to which all further breakdowns would then be adjusted (Petti and

[9] The LAU data used for this section are 2011 census outputs from all EU Member States as available at ec.europa.eu/CensusHub2.

[10] 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 (OJ L 105, 21.4.2017, p. 1).

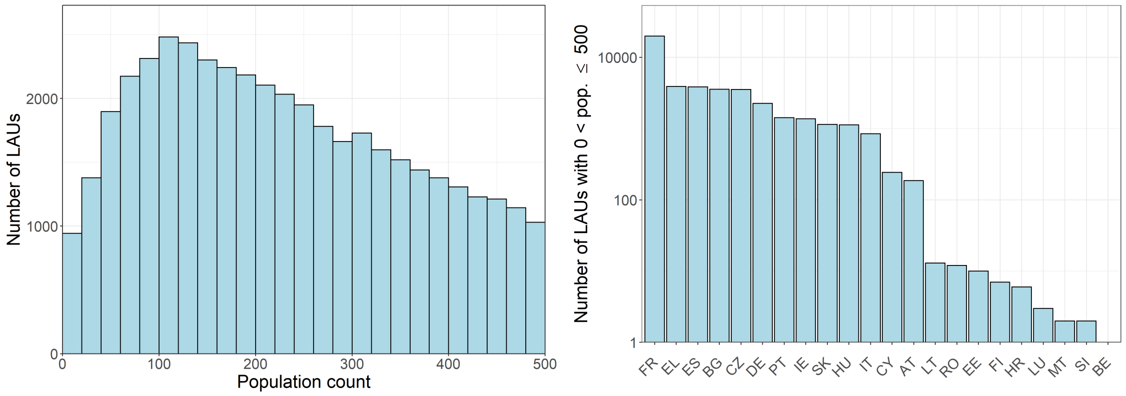

Figure 3: Distribution of populated LAUs with $\leq$ 500 residents across the total population count (left) and across EU Member States (right).

Flaxman, 2019). To cover both large distortions of totals as well as of sex balances, the counts of F, M and T are treated independently. In total, there are $\sim 167\,000$ LAU counts of F, M or $T < 500$ in the 2011 data.

Estimating distortions The basis for ε-DP noise analysed here is the discrete two-tailed geometric distribution, Eq. (10), with an ε range given in Eq. 4. However, in this ε range the discrete distribution already converges well to the continuous $\text{Lap}(1/\varepsilon)$. The cumulative inverse distribution function of $\text{Lap}(1/\varepsilon)$ can be used to calculate the probability for the noise magnitude $|x|$ to exceed a certain threshold E:

$$\Pr(|x| > E|\varepsilon) = \exp\left(-\varepsilon E\right). \tag{5}$$

This probability is plotted in the lower-right of Fig. 4 as a function of ε inside the relevant range, and for $E \in \{20, 50, 100\}$. Now Eq. (5) can be convoluted with the distribution of LAU counts (left plot in Fig. 3) to estimate how many LAU counts in each bin will end up with noise exceeding a given absolute relative error RE = 20, 50 or 100 %. These binned estimates can be tested by actually sampling some noise on the LAU data, and counting occurrences of RE magnitudes above a given threshold. Fig. 4 (left column) overlays the estimates with counts found in the noise-sampled data. Clearly the analytic estimates describe very well the noise-sampled data.

Distortions of single counts Looking now at the actual distortions in the left column of Fig. 4, one finds a sizeable dependence on ε, which is not surprising due to the exponential noise scaling in Eq. (5). In fact, noise distortions of single counts in these LAU statistics may be said to become manageable from $\varepsilon > 0.4$ (and we do not show the upper end of the ε range, $\varepsilon = 0.8$, for this reason). However, for $\varepsilon \lesssim 0.1$ there are many LAU counts expected with $|\text{RE}| > 50\,\%$ or even $> 100\,\%$.

For instance, with $\varepsilon = 0.025$ ($\sqrt{V_{\text{table}}} = 56.6$) there are $1\,648$ observations above 100 affected by $\pm 100\,\%$ or more, and still 87 observations above 200 with $\text{RE} \pm 100\,\%$ or more. Recall that every third of these observations describes a total count, and every 6th a total count with $RE < -100\,\%$, thus wiping out the whole population of that LAU. The largest LAU where this happens is Aragnouet, France, with originally 239 residents (now -7[11]). The situation does improve with $\varepsilon = 0.1$

[11] Negative output counts are a typical consequence of standard DP noise. These may be lifted

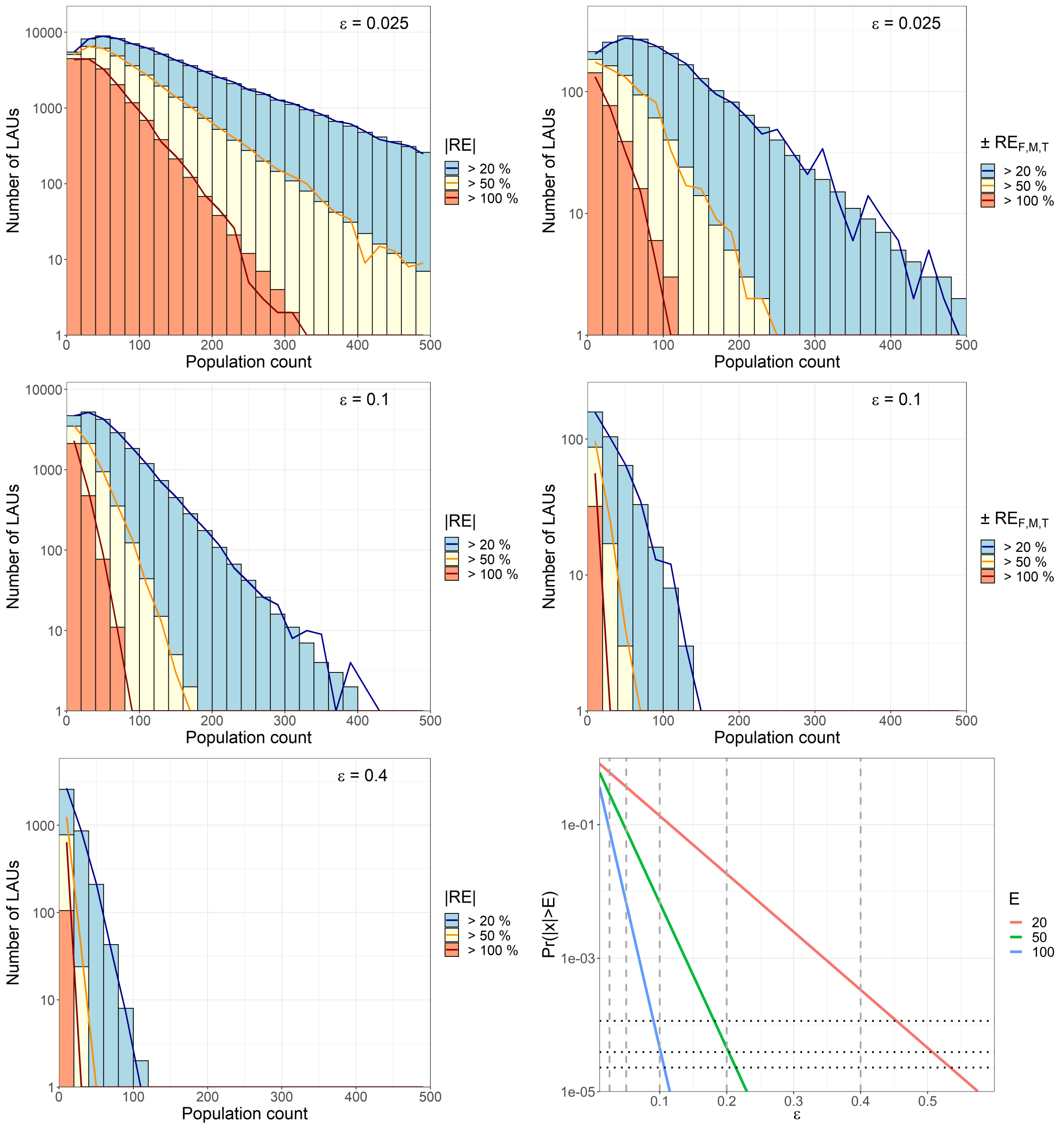

Figure 4: Log-linear estimates for frequencies of relative error (RE) magnitudes exceeding 20 % (blue), 50 % (yellow) and 100 % (orange) occurring in LAU counts, by total count: bins show the analytic estimate obtained from Eq. (5), while lines show the actual distortion frequencies found in the data with noise sampled.
The rows vary $\varepsilon = 0.025$ (top) to 0.1 (middle) to 0.4 (bottom). The left column counts single observations (F, M or T) exceeding a given RE, while the right column counts LAUs where F, M and T all exceed RE in the same direction.
The lower right histogram (F, M and T distorted in the same direction for $\varepsilon = 0.4$) is almost empty and thus replaced by a plot illustrating Eq. (5): log-linear $\Pr(|x| > E)$ as a function of ε with $E = 20$ (orange), 50 (yellow) and 100 (blue). Vertical dashed lines indicate ε choices from Eq. (4), while horizontal dotted lines show 1 over the number of LAUs with $T \leq E = 20$, 50 or 100.

($\sqrt{V_{\text{table}}} = 14.1$), but we still find 122 observations above 40 and 11 observations above 60 with RE±100 % or more. The largest depopulated LAU is again in France, Mélagues with originally 63 residents (now −9).

All these findings on single counts are disconcerting in their own right, but there is an added danger: if the total count is distorted so severely and inner table cells are adjusted to the new total[12], entire LAU populations may disappear from the census output. If inner cells are not adjusted, constraints like $F + M = T$ can be exploited to improve knowledge a bit; e.g. $\widehat{T} = (F + M + T)/2$.

Distortions of entire LAUs Such ad hoc 'repair' estimates exploiting 3-tuple constraints will not always help. This is illustrated by the situation where F, M and T are all distorted in the *same* direction ("broadband distortions"), so the distorted 3-tuple is internally consistent and no ad hoc estimate can improve the user's knowledge. To quantify this, one can count all LAUs affected by such broadband distortions; results are shown in the right column of Fig. 4. For $\varepsilon = 0.025$ there are 28 LAUs above 40 residents and 4 LAUs above 80 with a broadband distortion −100 % or more. The largest such LAU is Landremont, France with $F = 61 \to -8$, $M = 74 \to -26$ and $T = 135 \to -83$. For $\varepsilon = 0.1$, most broadband distortions of ±100 % only occur in the lowest count bin $(0, 20]$, but there is one above: this time Spain, Cidamón with $F = 15 \to -9$, $M = 20 \to -1$ and $T = 30 \to -17$. Broadband distortions ±20 % still occur for 61 LAUs with 100 or more residents. The largest LAU where this happens is Ellend, Hungary with $F = 112 \to 74$, $M = 94 \to 65$ and $T = 206 \to 158$. Even distortions around ±20 % may have significant policy impacts at local level.

Heuristic utility constraint on ε-DP noise Clearly the results above just paint in vivid colours what unbounded count-level noise size $\sqrt{V} \gtrsim \mathcal{O}(10)$ means for the accuracy of individual counts $\lesssim \mathcal{O}(10^2)$. Recall that acceptable accuracy of such counts—at least for the small-area output statistics—is a design requirement for most censuses. This can be turned into a utility constraint on unbounded ε-DP noise distributions by requiring that a given bound E_α is not exceeded at confidence level α for t output counts. The corresponding minimal privacy budget ε_α to be spent on this output is

$$\varepsilon_\alpha(E_\alpha | t, \alpha) > \frac{1}{E_\alpha} \log\left(\frac{t}{1-\alpha}\right) . \tag{6}$$

A reasonable choice could be $E_\alpha = 20$ and t the number of LAUs in an EU Member

to 0, as proposed e.g. by Ghosh, Roughgarden, and Sundararajan (2012) and planned for the 2020 U.S. census (Petti and Flaxman, 2019). However, this generally introduces a (normally small) overall bias to the output and may have other negative impacts on output utility, pointed out by Rinott et al. (2018). In any case, the discussion is not relevant here: all negative counts mentioned in this section can be replaced by 0 without changing any conclusion.

[12] I.e. in this example, noise on T would be fixed but noise on F and M would be post-processed to minimise the violation of the 3-tuple constraint $F + M = T$.

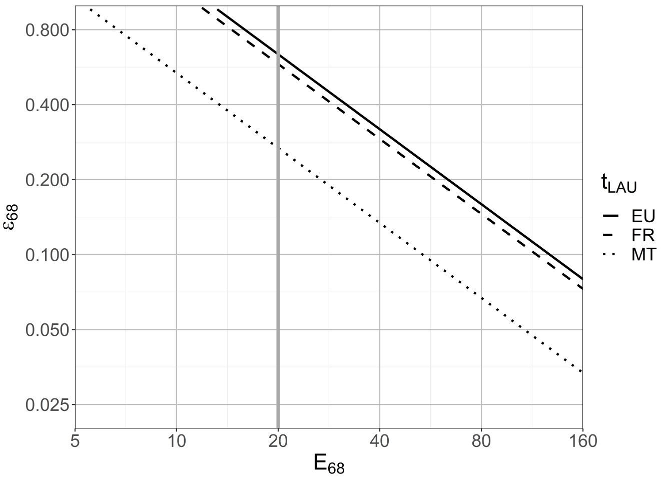

Figure 5: Log-log plot of ε_α as a function of E_α, Eq. (6), for $\alpha = 68\,\%$ and three LAU multiplicities: EU-27 with $t_{\text{LAU}} \simeq 1.1 \times 10^5$, France with $t_{\text{LAU}} \simeq 3.7 \times 10^4$ and Malta with $t_{\text{LAU}} = 68$.

State. This would give a (weak) utility guarantee that no LAU total count is distorted > 20 at c.l. $= \alpha$. Fig. 5 illustrates the implications on count-level ε values. In particular, the previous notion is confirmed that individual $\varepsilon \gtrsim 0.6$ (large countries like France, Germany) resp. $\varepsilon \gtrsim 0.3$ (small countries like Malta) would have to be spent at least on the small-area outputs to maintain unique census utility.

4.3 Discussion

The simple analysis above has shown that tail effects of unbounded noise distributions, such as strictly ε-DP ones, may cause grave distortions at small geographies. This starts to kick in severely around count-level $\varepsilon_{\text{table}} < 0.4$ for most countries ($> \mathcal{O}(10^3)$ LAUs), and points at similar conclusions as in Santos-Lozada, Howard, and Verdery (2020): with unbounded noise it is very difficult to maintain a certain minimum utility per individual small area unit, for *every* small area unit in the output. Of course, one could now enter the game of redistributing privacy budgets between geographies or between other statistics, but such fine-tuning is beside the point because it is extremely difficult, on principle, to avoid large tail distortions in an output corner that may turn out to be critical. There is an obvious solution that avoids any of these concerns or haggling of budgets between output statistics: don't use unbounded noise for population statistics where single-count accuracy is key.

5 Risk vs. utility for upcoming censuses

To integrate the findings of sections 3 and 4 for the scope of the 2021 EU census, the parameter constraints of Figs. 1, 2 and 5 can be combined into a global picture of the generic noise parameter space: Fig. 6 illustrates that utility-driven parametrisations using individual count-level variance V and noise bound E can be set up within a range that avoids all risk/utility constraints assessed in this paper (e.g. $V \simeq 2$ to 3 and $5 \lesssim E \lesssim 10$). On the other hand, risk-driven approaches such as strictly ε-DP mechanisms with unbounded noise are severely constrained by combining risk (mas-

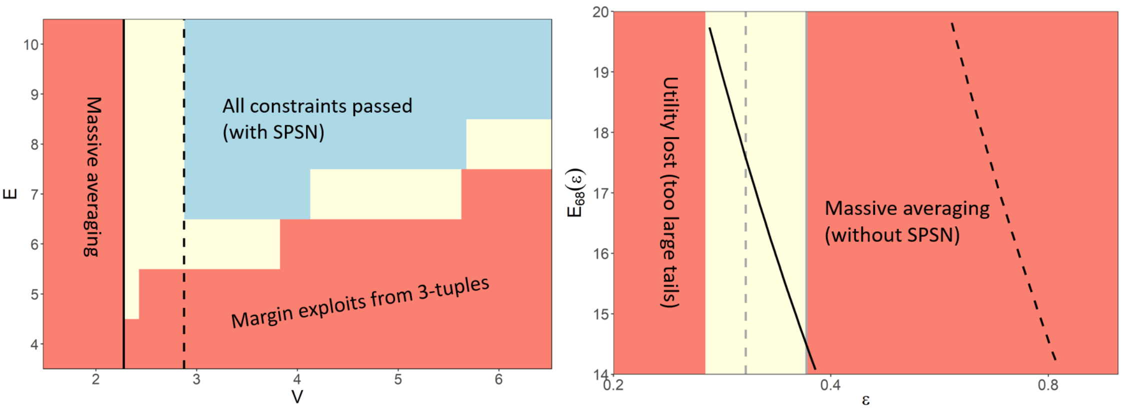

Figure 6: Generic noise parameter space highlighting regions that survive all risk/utility constraints (blue/yellow): the utility-driven generic V–E plane (left) and the risk-driven (count-level) ε range (right). Yellow regions are not excluded; they rather indicate that such setups may work in certain circumstances, or with slightly relaxed constraints. Note that the ε range (right) is a one-parameter space, where a utility constraint is taken from Eq. (6) (note different E scales). The SPSN principle is assumed to be invoked on the left, but not on the right (DP default). On the right, no blue region survives all constraints conservatively: averaging (grey lines showing ε for smallest k/t^2 dashed and for next-smallest solid) and $E_\alpha < 20$ at $\alpha = 68\,\%$ (black lines showing the $E_\alpha(\varepsilon)$ curve for Malta solid and for France dashed). When relaxing certain constraints (e.g. $E_{68} \gtrsim 20$ and/or slight averaging vulnerability), a small yellow band $\varepsilon \in [0.27, 0.37]$ remains.

sive averaging) and utility (small-area accuracy) considerations. In particular, only a narrow window around count-level $\varepsilon \simeq 0.3$ remains with acceptable compromises.

Global constraints as in Fig. 6 do depend on the exact (static) output, but in general such constraints can always be obtained systematically from the static output structure. This is what makes the risks controllable: if no satisfying parameter setup is found, the output can be curated to relax the constraints. While DP as a *risk measure* may contribute to an assessment of appropriate noise amounts, the flexibility of ε-DP *mechanisms* is heavily limited with just a single parameter (the privacy budget). It is the presence of a second parameter—the noise bound E, or δ in (ε, δ)-DP mechanisms—that adds flexibility to arbitrate risk vs. utility.[13]

6 Conclusions

While traditional SDC methods in population statistics mainly focused on small counts at high risk of direct re-identification (e.g. suppressing those, etc.), powerful theoretical results of Dinur and Nissim (2003) have meanwhile shown that entire microdata databases can often be reconstructed accurately from too detailed output tables, thus exposing rare records even if small output counts were treated. These results suggest that random noise methods are the most effective counter-measure, where the amount of noise should scale with output detail. This scaling rule implies

[13] For instance, CK offers some margin to increase noise by tuning V while keeping E fixed. In ε-DP mechanisms one can only reduce ε, which often has severe consequences on utility.

a first important notion: flexible output mechanisms (where the complexity is not fixed a priori) require some kind of noise scaling and are thus much harder to realise within reasonable risk and utility constraints. On the other hand, static output mechanisms (pre-fixed complexity) allow for a diligent curation, including controlling risks and assessing risk/utility trade-off to fix a static noise amount. Unless imposed by external constraints, a move from a static to a flexible output mechanism should be considered only with utmost care.

Differential privacy (DP) is a useful concept to quantify risk irrespective of a particular output scenario, and hence to compare risk levels consistently between various SDC approaches (Dwork, McSherry, et al., 2006; Dwork and Roth, 2014). DP risk measures may thus contribute to a broadly based SDC assessment. Moreover, DP provides for automatic noise scaling with output complexity, as required by flexible output mechanisms. However, this paper suggests that the complexity scaling of DP noise levels is over-protective for increasingly complex outputs, so DP inferences on absolute noise levels should be handled with care, especially with complex static outputs.

DP risk measures should be distinguished from DP output mechanisms, where the latter may give strict ε-DP or relaxed (ε, δ)-DP privacy guarantees with ε the total privacy budget spent on the entire output. However, strictly ε-DP mechanisms must employ unbounded random noise distributions, while relaxed (ε, δ)-DP or not manifestly DP mechanisms can have bounded distributions. It is shown that in static output scenarios, typical generic risks such as margin exploits and massive averaging are controllable with bounded noise, (ε, δ)-DP or not. Conversely, the unbounded noise of strictly ε-DP mechanisms may lead to severe utility damage when the noise amount is tuned up to evade averaging risks. More generally, the fact that ε-DP mechanisms only have a single parameter costs a lot of flexibility.

Censuses are big national investments for comparably narrow purposes, not necessarily to answer any question any user may have on any characteristics of any subpopulation. This suggests a static output mechanism with a utility-driven parametrisation, which allows to maximise utility within purpose scope while controlling risks carefully. Finally, if particular SDC mechanisms jeopardise unique census features, they are bluntly unfit for the purpose. For the scope of the 2021 EU census round, bounded noise from the cell key method (recommended in the European Statistical System (Antal et al., 2017)) is found suitable to protect outputs in a controlled way: The generic parameter space (noise variance and magnitude bound) is constrained by different risk or utility requirements, but various setups remain feasible. Such setups can obtain a relaxed (ε, δ)-DP guarantee, if needed. On the other hand, strictly ε-DP mechanisms are severely constrained, with only a small parameter window remaining for a possibly acceptable compromise. It seems strict ε-DP guarantees are overpriced (in utility) at least for census-like scenarios.

A Extended preliminaries and concepts

A.1 Differential privacy: a risk measure

Differential privacy was first proposed by Dwork, McSherry, et al. (2006), in the wake of the database reconstruction theorem. In plain words, its paradigm is that every query result (output) should be robust against addition to, or removal from, the input database of any single record, e.g. picking one record and removing it from the database should not significantly change any outputs (hence *differential* privacy). This is the individual privacy guarantee, and its immediate attraction is formulated by Dwork (2011): "*Any mechanism satisfying this definition addresses all concerns that any participant might have about the leakage of his or her personal information, regardless of any auxiliary information known to an adversary: Even if the participant removed his or her data from the dataset, no outputs (and thus consequences of outputs) would become significantly more or less likely.*"

There are various mathematical definitions of differential privacy, so we repeat here the most generic one, introducing both strict as well as relaxed (or approximate) differential privacy in one go as Dwork, Kenthapadi, et al. (2006): given two neighbouring input databases d and d' that differ exactly in one record, any mechanism $\mathcal{M}(\cdot)$ acting on the universe of input databases to generate outputs must fulfil

$$\Pr(\mathcal{M}(d) \in S) \leq e^{\varepsilon} \Pr(\mathcal{M}(d') \in S) + \delta \tag{7}$$

for all subsets $S \subseteq \text{Range}(\mathcal{M})$ to be *δ-approximately ε-differentially private* or short (ε, δ)-DP, where ε and δ are parameters establishing the differential privacy level. For $\delta \to 0$, Eq. (7) reduces to a definition of *strictly ε-differentially private* or short ε-DP mechanisms.

The definition implies that, for any single output $s \in \text{Range}(\mathcal{M})$—singleton S in Eq. (7)—with nonzero probability on d, the probability to obtain s from d' should also be nonzero for the mechanism to be possibly ε-DP. This suggests some kind of noise injection applied by $\mathcal{M}$ as an option to comply with Eq. (7). While noisy $\mathcal{M}$s are discussed in more detail in annex A.3, it is important to note here that ε-DP or (ε, δ)-DP are attributes or qualifiers of any given $\mathcal{M}$, thus measuring the individual information leakage from any thinkable output. Therefore, ε and δ are handy risk measures to compare different output scenarios and noise mechanisms, as done e.g. by Rinott et al. (2018).

Finally, it is interesting to note how differential privacy embraces the *fundamental law of statistical disclosure control* (formalised by Dwork and Naor, 2010 as *rigorous impossibility of Dalenius's privacy goal*), which essentially states that *any* provision of useful statistical information *necessarily* entails a nonzero trailing risk of disclosing information on some individuals[14], i.e. that the trade-off between privacy and utility

[14] Interestingly, including potentially individuals that did not even contribute to the statistics, as pointed out e.g. by Dwork (2011).

is fundamental and not contingent. With the privacy budget parameter ε, differential privacy provides a transparent and intuitive "turning knob"—but no immediate guidance on how exactly to adjust it.

A.2 Noise distributions: bounded or unbounded?

Recall that the discussion is confined to outputs representing unweighted person counts, or sets of such counts (e.g. contingency tables). Then the most generic output mechanism $\mathcal{M}(\cdot)$, in the sense of annexes A.1 and A.3, returns an ordered k-tuple of frequencies representing the answers to k individual counting queries passed to $\mathcal{M}$. Further let $\widetilde{\mathcal{M}}(\cdot)$ denote an *exact* output mechanism without any noise injected, so that $\text{Range}(\widetilde{\mathcal{M}}) = \mathbb{N}_0^k$. Then by noise distribution we mean the probability distribution underlying the process of drawing an additive (pseudo) random noise term $x \equiv (\mathcal{M} - \widetilde{\mathcal{M}})(d)$ for $k = 1$ and any given d. Among the popular options are e.g. Laplace, Gaussian, or entropy-maximising distributions, which may come in various flavours and with auxiliary constraints, but many properties can be captured by just two generic attributes: the noise variance $\text{Var}(x)$ and its magnitude bound $|x| \leq E$ ($\leq \infty$). Here we just give a crude classification based on the DP categories introduced in annex A.1.

ε-DP noise distributions manifestly comply with Eq. (7) for any possible singleton S (single output count) with $\delta = 0$. It is easy to show (Dwork, 2011) that e.g. the Laplace distribution

$$\text{Lap}\,(\Delta/\varepsilon) : \; x \sim \frac{\varepsilon}{2\Delta} \exp\left(-\frac{\varepsilon|x|}{\Delta}\right) \tag{8}$$

with $\text{Var}(x) = 2(\Delta/\varepsilon)^2$ fulfils this requirement, where Δ is the *global sensitivity* of $\widetilde{\mathcal{M}}$ defined as

$$\Delta := \max_{d,d'} \sum_{i=1}^{k} \left| \widetilde{\mathcal{M}}(d)_i - \widetilde{\mathcal{M}}(d')_i \right| \tag{9}$$

with i running through output k-tuple indices. Clearly for $k = 1$ and unweighted person counts, $\Delta = 1$ and $x \sim \text{Lap}(1/\varepsilon)$ in this case.[15] Now this distribution is over $\mathbb{R}$, so that $\text{Range}(\mathcal{M}) = \mathbb{R}^k$ which may return non-integer person counts. This can be lifted this by using the discrete two-tailed geometric distribution (Ghosh,

[15] Apart from unweighted counts, the issue with Δ is that it is generally hard to obtain, and arbitrarily difficult for some queries on weighted or magnitude data: e.g. Bambauer, Muralidhar, and Sarathy (2014) argue that in an average income query the global sensitivity is theoretically driven by the highest-income person in the world, because the query result must be robust also against addition of that person to the database. Naturally such a Δ drives the noise through the roof and renders all outputs useless. On the other hand, capping Δ arbitrarily dilutes the individual privacy guarantee.

Roughgarden, and Sundararajan, 2012)

$$x \sim \frac{1 - \exp(-\varepsilon)}{1 + \exp(-\varepsilon)} \exp(-\varepsilon|x|), \tag{10}$$

which gives $\text{Range}(\mathcal{M}) = \mathbb{Z}^k$ and approximates to $\text{Lap}(1/\varepsilon)$ for $\varepsilon \ll 1$.

Note finally that noise distributions, continuous or discrete, must be unbounded to be ε-DP. To see this, assume bounded noise with $\Pr(x > E) = 0$. Then in Eq. (7), choose without loss of generality $d > d'$ and $s = \widetilde{\mathcal{M}}(d) + E$ (i.e. $k = 1$). Thus, $\Pr(\mathcal{M}(d') = s) = 0$ as $s - \widetilde{\mathcal{M}}(d') = E + 1$ and the inequality requires $\delta \geq \Pr(\mathcal{M}(d) = s) > 0$ to hold, which contradicts $\delta = 0$.

(ε, δ)-DP and other noise distributions In a sloppy manner, most noise distributions that are not ε-DP are (ε, δ)-DP: If a distribution fails the strict ε-DP requirement, Eq. (7) with $\delta = 0$, a $\delta > 0$ can usually be found to establish (ε, δ)-DP. In particular, unbounded noise distributions can usually be truncated to give (ε, δ)-DP, where δ depends on the resulting probability distribution close to its discontinuity (Rinott et al., 2018). Also for cell key noise of Thompson, Broadfoot, and Elazar (2013), taking variance $\text{Var}(x) \equiv V$ and noise bound $|x| \leq E$ as input parameters, an (ε, δ)-DP level can be inferred (Bailie and Chien, 2019). However, the issue is not about finding a δ but about dealing with its value: clearly it should be $\delta \ll 1$ but how small exactly? For instance, $\delta < 1/n$ is stated in Dwork and Roth (2014), but higher values are also discussed in Rinott et al. (2018). It is also argued there that often the choices of δ (and ε) are policy decisions, not statistical decisions.

Bounded vs. unbounded noise Why select an unbounded noise distribution? As shown above, if a strictly ε-DP output mechanism is ultimately desired, the underlying noise distribution must be unbounded. Moreover, Asghar and Kaafar (2020) recently claimed that a tight noise bound poses additional disclosure risks. However, sections 3 and 4 will argue that unbounded noise may come at too high a price on utility, while the additional risks of bounded noise can be controlled.

A.3 Noisy output mechanisms

Noise distributions handle the special case of a single scalar output, i.e. a single call to $\mathcal{M}$ with $k = 1$. In contrast, a generic (noisy) output mechanism denotes a more powerful and complex $\mathcal{M}$ that ideally accounts automatically for *noise composition* across all outputs. In particular, from a DP perspective an ε-DP noise distribution does not automatically constitute an ε'-DP output mechanism with $\varepsilon = \varepsilon'$; as argued below, an additional layer of output curation or privacy budget management is needed. But before outlining various output mechanisms, some prerequisites are introduced.

Disjoint outputs: histograms and tables are lists of counts breaking down the input database into sub-populations, i.e. (using notation introduced in annex C.1) a single call to an $\mathcal{M}: d \mapsto T_A$ where $A = \{a_i\}$ is a list of disjoint breakdown categories contained in d and obviously $k = |A|$. It is easy to see from Eq. (9) that still $\Delta = 1$ despite $k > 1$ output counts, because addition or removal of a single record in d can only change the single count $\mathcal{M}(d)_i \equiv T_{a_i}$ where that record contributes.[16] In DP literature, this class of queries is called histogram queries (e.g. Dwork, 2011), and properties transcend to multi-dimensional table outputs $\mathcal{M}: d \mapsto T_{\mathbf{A}}$ with $\mathbf{A} = \{A_i\}$ and $k = \prod_{i=1}^{|\mathbf{A}|} |A_i|$ but $\Delta = 1$ still.

Same participants–same noise (SPSN) is a principle to decide whether the noise term added to a given output count is drawn afresh or reused from a lookup table (Fraser and Wooton, 2005; Thompson, Broadfoot, and Elazar, 2013). It was introduced to forestall averaging to some extent, as per Chebyshev's inequality the probability that t redundant noisy observations average to the true count converges to 1 for increasing t (cf. Eq. (3)). If the noise is looked up instead to be always the same for the same question asked, averaging over redundancies is less straightforward (but still possible, as shown in section 3.1).[17]

Static output mechanisms: non-interactive $\mathcal{M}$ With notions of table outputs and SPSN at hand, a static output mechanism without SPSN is defined here as

$$\mathcal{M}: d \mapsto \cup_{I\in\{1\cdots M\}} TI_{\mathcal{P}(\mathbf{A}_I)} \tag{11}$$

returning M tables and all marginals in a single call ($\mathcal{P}(\mathbf{A})$ is the power set of $\mathbf{A}$). Think of $\mathcal{M}$ returning the entire set of $M = 103$ population tables in the 2021 EU census programme[10] at once. If SPSN is invoked,

$$\mathcal{M}: d \mapsto \cup_{I\in\{1\cdots M\}} T_{\mathcal{P}(\mathbf{A}_I)} \tag{12}$$

containing only unique cross-tabulations in the table programme, which is a much smaller set than in Eq. (11). Then the M predefined tables can be put together for publication from the unique outputs (e.g. collecting internal cells and all corresponding marginals). This also shows how a static output mechanism can be combined with a more flexible user interface (table builder): Just expose the unique outputs as building blocks available for user-defined tables.

The advantage of a static output is that the curator has full control over the output and so decides about the acceptable amount of redundancies and internal constraints (the levers for typical disclosure attacks described in section 3) before publication. Moreover, k and Δ for a DP mechanism can simply be counted from the redundancies in the output set; the full exercise is carried out in annex C, and k and Δ are given by Eq. (18) (without SPSN) resp. Eq. (19) (with SPSN).

[16] Note that we decided in annex C.1 to suppress all categories a_i = total; if the total count was included in A, $\Delta = 2$.

[17] The intricacies of defining SPSN discussed in Rinott et al. (2018) are not relevant here, because noise independence is established contextually through the variable attributes $\mathbf{a}$ of a given cell count $T_{\mathbf{a}}$.

Flexible output mechanisms: interactive $\mathcal{M}$ It is sometimes argued (justly) that static outputs cannot provide the full richness of d, at least not without running into the database reconstruction theorem (e.g. Dwork, McSherry, et al., 2006; Dwork, 2011). For the scope of this paper, a flexible output mechanism is again defined as $\mathcal{M}: d \mapsto T_{\mathbf{A}}$, but this time with interactive elements: the cross-classification $\mathbf{A} = \{A_i\}$ as well as variable breakdowns $A_i = \{a_{ij}\}$ are (at least to some extent) customisable, and the user is allowed to call $\mathcal{M}$ t times with a series of questions $\{\mathbf{A}_t\}$. For each $\mathbf{A}_t$ the noise can be curated but a priori, t is unknown. Such an output mechanism without any noise curation across t is highly susceptible to various attacks described in Asghar and Kaafar (2020), and discussed again in section 3.2. DP may guide a way out, but at a costly price, as shown below. It seems that, in certain scenarios such as censuses, static outputs are preferable because the objectives are clear and limited, so that outputs can be curated accordingly.

Manifestly ε-DP output mechanisms In static outputs with an ε-DP noise distribution, the global privacy budget ε is automatically distributed correctly across all outputs by virtue of Δ. Customised distributions of ε are possible, using e.g. i calls to $\mathcal{M}$ with ε_i each. Then the whole output is still ε-DP with $\varepsilon = \sum_i \varepsilon_i$ through the DP composition theorem (Dwork and Roth, 2014; Rinott et al., 2018). This is still a static mechanism, and it is what the U.S. Census Bureau plans for its 2020 census outputs (Abowd, 2018; Garfinkel, 2019).

Flexible mechanisms can also be ε-DP through composition but are more tricky, because ad hoc rules must be applied to distribute a global ε across t outputs. Two approaches are usually proposed (e.g. Dwork, 2011): either cap t per user at some value, or spend ε iteratively as $\varepsilon_i = \varepsilon/2^i$, so that $\varepsilon = \sum_{i=1}^{\infty} \varepsilon_i$. Both have serious disadvantages: a cap is always arbitrary, while the iterative noise explodes quickly: $\sqrt{\mathrm{Var}(x)} \simeq \mathcal{O}(10^3)$ for $\varepsilon = 1$ and $i = 10$ (and assuming $\Delta = 1$ in each case). Moreover, in both cases, adversaries may try to refresh their ε budget somehow.

Finally, strict ε-DP mechanisms seem to scale too steeply with complexity: For a global ε-DP guarantee on an output of complexity t, DP composition requires that the global privacy budget be split between all outputs, e.g. as ε/t. Then e.g. the $\mathrm{Lap}(1/\varepsilon)$ noise, Eq. (8), scales as

$$\sqrt{\mathrm{Var}(x)} = \sqrt{2}\frac{t}{\varepsilon} \sim t \tag{13}$$

and not as $\sim \sqrt{t}$, which would be sufficient from Dinur and Nissim (2003). Therefore, ε-DP mechanisms relying on these noise distributions will over-protect outputs of increasing complexity.[18] In any case, the situation is more complicated when the

[18] Theorem 3.20 of Dwork and Roth (2014) provides for improved composition scaling $\sim \sqrt{t}$, but the resulting global guarantee is relaxed (ε, δ)-DP only, not strictly ε-DP. Similarly, Gaussian DP noise scales $\sim \sqrt{t}$ (Dwork, Kenthapadi, et al., 2006; Dwork and Rothblum, 2016), but also there the guarantee is only (ε, δ)-DP.

noise distribution is not DP-parametrised (i.e. ε and δ are calculated ad hoc on each single output, rather than being noise input parameters): Then the privacy budget spent must be calculated manually across all outputs to obtain a global ε-DP or (ε, δ)-DP guarantee.

Risk-driven vs. utility-driven parametrisations A risk-averse statistics curator may find it desirable to publish only outputs that provide a global ε-DP guarantee. Manifestly ε-DP output mechanisms may thus be said to follow a *risk-driven parametrisation*, because the ε parameter makes the selected level of risk/utility trade-off transparent. However, such risk-driven DP parametrisations have their own pitfalls: as noted in Rinott et al. (2018), the total privacy budget (value of the global ε) is not easily fixed from statistical and/or disclosure control decisions, and the noise scaling is over-protective as argued above in Eq. (13). Moreover, as shown in annex A.2, unbounded noise is unavoidable if strict ε-DP is sought. In effect, risk-driven parametrisations may generally be said to provide high transparency and control on the privacy side, at the cost of low transparency and control on utility characteristics.

On the other hand, a *utility-driven parametrisation* is much closer to the amount of noise actually injected to each single output count. For instance, if the noise variance V and its bound E are the input parameters, like in the cell key method, users get a very transparent idea of what happened to the data: they know the typical noise size $\pm\sqrt{V}$ and that each individual count is at most $\pm E$ off.[19] This gives strong utility guarantees on the output, which a risk-driven parametrisation simply cannot provide. We argue in section 3.1 that risks are well controllable for a static (census-like) output with utility-driven parametrisation. However, to complete the dichotomy to the risk-driven view above, utility-driven parametrisations do generally have drawbacks in that their corresponding privacy guarantees are not obvious or straightforwardly accessible. For instance, Rinott et al. (2018) and Bailie and Chien (2019) outline procedures how to translate utility-driven parameters (like count-level noise variance and tail cut-off) into corresponding global (ε, δ)-DP guarantees, where naturally the DP composition theorem (Dwork and Roth, 2014) is central.

B ESS recommendations for harmonised 2021 EU census protection

Within the legal framework for EU censuses[20], statistical confidentiality and SDC measures are under the responsibility of the national statistical authorities of the

[19] The E parameter is usually not disclosed exactly, because its knowledge gives additional—theoretical—disclosure risks. However, a vague communication e.g. that $E \lesssim 10$ is still very useful (cf. section 4.2).

[20] Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses (OJ L 218, 13.8.2008, p. 14).

Member States, so that confidential data must not be submitted to Eurostat.[21] This means that Eurostat will not receive, maintain or process any confidential data (or personal data in GDPR sense) related to the upcoming 2021 EU census round. This has lead to various European Statistical System[22] projects facilitated by Eurostat and aimed at developing common ESS recommendations for harmonised methods (Antal et al., 2017) and tools (De Wolf et al., 2019a) to protect census outputs at the national level. These recommendations include the cell key (CK) method as a particular mechanism implementing bounded noise (Meindl and Enderle, 2019).

During the project work, considerable efforts were made to provide more flexible and accessible SDC tools (De Wolf et al., 2019b) but also to assess disclosure risks vs. utility and improve the general methodology (Giessing, 2016; Enderle, Giessing, and Tent, 2018; Enderle, Giessing, and Tent, 2020). Also the findings of this paper (see section 5) suggest that the ESS recommended CK method stands as a good practice, if applied correctly and consistently, featuring in particular superior utility properties in various output scenarios with still small and controllable disclosure risks.

Small count threshold parameter Note that the CK method formally takes a third parameter (in addition to variance V and noise bound E), namely a threshold j_s for the smallest non-zero count that may occur in the output (i.e. no output count will be > 0 and $\leq j_s$). This may be desirable for some national output curators, e.g. for historical or cultural reasons. The CK method implements j_s in a consistent manner without introducing biases, as opposed to naïve noise truncation ≥ 0 (Ghosh, Roughgarden, and Sundararajan, 2012; Petti and Flaxman, 2019). However, Rinott et al. (2018) argue that a $j_s > 0$ actually increases risks and may loosen a corresponding (ε, δ)-DP guarantee. Having a $j_s > 0$ or not does not affect any of the arguments of this paper significantly, so we just fixed $j_s = 0$ for all analyses presented.

The European census grid is a notable new output from the 2021 EU census round.[23] Several key census indicators will be published for each cell of a pan-European 1 km^2 grid, which poses specific new disclosure risks (see Bach, 2019 for an overview). However, main risks relate to the possible combination of different non-nested small areas (e.g. Costemalle, 2019), where again random noise is considered effective (Antal et al., 2017). For the scope of this paper the grid output is not important: the additional number of 3-tuples is negligible,[24] and it does not add

[21] See Articles 2(5) and 4(3) of Regulation (EU) 2017/712, footnote 10.

[22] The joint body of Eurostat and the national statistical institutes of all EU countries and Iceland, Liechtenstein, Norway and Switzerland. It is responsible for the development and quality assurance of official European statistics.

[23] 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 (OJ L 296, 22.11.2018, p. 19).

[24] The IRR counting in section 3.2 does not include the 1 km^2 grid output (cf. annex B), but

any redundancy to any of the other output statistics because the grid does not overlap with administrative geographic breakdowns and it has no ‘total’.

C Massive averaging in static outputs

C.1 Output structure

In static output scenarios, all output statistics are fixed in advance. In census-like scenarios, these are typically contingency tables which cross-tabulate several categorical variables with pre-fixed, finite sets of values (variable breakdowns). The tables may contain margins, i.e. there may be a hierarchical structure inside the variable breakdowns (some values are contained within other values). In particular, let each variable breakdown contain a value “total”, which denotes the union of all other (“internal”) values of that variable. Finally, a table cell is a distinct combination of variable values, where the cell value (without noise injection) is the number of microdata records that are characterised by the combination of variable values defining the cell.

Let A denote a variable breakdown represented by a finite set of discrete values $a \in A$, where generally $A = \{\cdot, \text{total}\}$ and $|A|$ the cardinality of A. Let $\mathbf{A}$ denote an m-tuple of variable breakdowns $\{A_i\}_m$, so that $|\mathbf{A}| = m$, and let $\mathbf{a} \in \mathbf{A}$ denote an m-vector of variable values $\{a_i\}_m$ with $a_i \in A_i \ \forall \ i \in \{1 \cdots m\}$. Then $T_{\mathbf{A}}$ represents the m-dimensional cross-tabulation of $\mathbf{A}$ (a table), and $T_{\mathbf{a}}$ a single table cell thereof. Furthermore, $T_{\mathbf{A}',\mathbf{a}''}$ represents the $|\mathbf{A}'|$-dimensional sub-table of $T_{\mathbf{A}}$ obtained by fixing variables $\mathbf{A}''$ to values $\mathbf{a}''$, and $\mathbf{A}' \,\dot\cup\, \mathbf{A}'' = \mathbf{A}$. Now suppress all $a_i = \text{total}$, such that $T_{\mathbf{A}'} \subset T_{\mathbf{A}}$ represents the $|\mathbf{A}'|$-dimensional marginal table for $\mathbf{a}'' = \{\text{total}\}_{|\mathbf{A}''|}$. Thus the table $T_{\mathbf{A}}$ consists of

$$|\mathcal{P}(\mathbf{A})| = \sum_{i=0}^{m} \binom{m}{i} = 2^m$$

sub-tables, including itself, all marginal tables and the total margin T, where $\mathcal{P}(\mathbf{A})$ is the power set of $\mathbf{A}$.

On the other hand, we can sum all table cells characterised by $\mathbf{a} = \{\mathbf{a}_{m-1}, a_m\}$ over all values $a_m \in A_m$ to obtain an independent redundant representation (IRR) of the marginal cell $T_{\mathbf{a}_{m-1}}$, or in short-hand notation

$$\sum_{i=1}^{|A_m|} T_{\{\mathbf{a}_{m-1}, a_{m,i}\}} \coloneqq T_{\mathbf{a}_{m-1}}{}^{\mathbf{A}_m} \tag{14}$$

and $T_{\mathbf{A}_{m-1}}{}^{\mathbf{A}_m}$ the respective marginal table. Thus, generally, $T_{\mathbf{A}'}{}^{\overline{\mathbf{A}'}}$ is an IRR of $T_{\mathbf{A}'}$ for any possible disjoint partition of $\mathbf{A} = \{\mathbf{A}', \overline{\mathbf{A}'}\}$. Then the set of all IRRs of a

the additional number of complete sex 3-tuples (one per grid cell) is negligible for both section 3.1 and 2: e.g. for Germany the grid would provide another 3.6×10^5 3-tuples, a 1 % effect.

target cell $T_{\mathbf{a}'}$ (or entire target table $T_{\mathbf{A}'}$) is given by $\mathcal{P}(\overline{\mathbf{A}'})$ as $\{{T_{\mathbf{a}'}}^{\mathbf{A}_i}\}_{\mathbf{A}_i \in \mathcal{P}(\overline{\mathbf{A}'})}$, or short ${T_{\mathbf{a}'}}^{\mathcal{P}(\overline{\mathbf{A}'})}$.

Example: Table 9.2 of the 2021 EU census programme[10] is defined as

$$9.2: \quad \text{GEO.M} \times \text{SEX} \times \text{AGE.M} \times \text{YAE.H}, \tag{15}$$

where GEO.M is the geographic breakdown of medium detail (NUTS 3), SEX $= \{\text{F}, \text{M}, \text{total}\}$ is the sex breakdown, AGE is the medium age breakdown (5-year bands) and YAE.H is the highly detailed breakdown of year of arrival in the reporting country (by single years). This table is expressed as $T_{\mathbf{A}}$, where $\mathbf{A} = \{\text{GEO.M}, \text{SEX}, \text{AGE.M}, \text{YAE.H}\}$, and $|\mathbf{A}| = 4$ so there are 16 possible sub-tables or subsets of $\mathbf{A}$. For instance, let $\mathbf{A}' = \{\text{GEO.M}, \text{SEX}\}$. Then $\overline{\mathbf{A}'} = \{\text{AGE.M}, \text{YAE.H}\}$ and

$$\mathcal{P}\left(\overline{\mathbf{A}'}\right) = \{\emptyset, \{\text{AGE.M}\}, \{\text{YAE.H}\}, \{\text{AGE.M}, \text{YAE.H}\}\}.$$

Thus, we find for the table of interest $T_{\{\text{GEO.M},\text{SEX}\}}$ four IRRs inside $T_{\mathbf{A}}$: $T_{\{\text{GEO.M},\text{SEX}\}}$ itself (the trivial marginal table obtained from $\emptyset \in \mathcal{P}(\overline{\mathbf{A}'})$) as well as ${T_{\{\text{GEO.M},\text{SEX}\}}}^{\{\text{AGE.M}\}}$, ${T_{\{\text{GEO.M},\text{SEX}\}}}^{\{\text{YAE.H}\}}$ and ${T_{\{\text{GEO.M},\text{SEX}\}}}^{\{\text{AGE.M},\text{YAE.H}\}}$.

C.2 Averaging attacks

Let the static output consist of M predefined tables $\cup_{I \in \{1 \cdots M\}} TI_{\mathcal{P}(\mathbf{A}_I)}$. This fixes the entire universe of independent output counts, or table cells before publication, including all hierarchical constraints and redundancies outlined in section C.1. It is thus possible to identify, for any given target count $TI_{\mathbf{a}}$ inside the universe, all IRRs contained in the universe, i.e. all ${TJ_{\mathbf{a}}}^{\mathcal{P}(\overline{\mathbf{A}_J})}$ from all $J \in \{1 \cdots M\}$ with $\mathbf{A} \subset \mathbf{A}_J$ and $\{\mathbf{A}, \overline{\mathbf{A}_J}\} = \mathbf{A}_J$. The corresponding set is

$$\mathcal{T}_{\mathbf{a}}(\mathbf{A}) := \cup_{J \in \{1 \cdots M\} \text{ where } \mathbf{A} \subset \mathbf{A}_J} {TJ_{\mathbf{a}}}^{\mathcal{P}(\overline{\mathbf{A}_J})},$$

A global constraint using all available redundant representations of the target count is thus given by

$$TI_{\mathbf{a}} = \frac{1}{|\mathcal{T}_{\mathbf{a}}(\mathbf{A})|} \sum_{i=1}^{|\mathcal{T}_{\mathbf{a}}(\mathbf{A})|} \mathcal{T}_{\mathbf{a}}(\mathbf{A})_i. \tag{16}$$

When the output is protected by noise injection as described in annex A, each independent output cell will get independent noise from a given distribution with variance V. Note that the SPSN principle introduced in annex A.3 leads to *dependent* noise on all $TI_{\mathbf{A}}$ and $TJ_{\mathbf{A}}$ with $I \neq J$ and $\mathbf{A} \subseteq \mathbf{A}_I \cap \mathbf{A}_J$. Thus, with SPSN we can drop table indices I so that the entire output universe is just $\cup_{I \in \{1 \cdots M\}} T_{\mathcal{P}(\mathbf{A}_I)}$. Defining a set that consists only of unique variable combinations disjoint from $\mathbf{A}$

$$\mathcal{U}(\mathbf{A}) := \cup_{I \in \{1 \cdots M\} \text{ where } \mathbf{A} \subset \mathbf{A}_I} \mathcal{P}(\overline{\mathbf{A}_I}),$$

the complete set of IRRs of $T_{\mathbf{a}}$ is just ${T_{\mathbf{a}}}^{\mathcal{U}(\mathbf{A})}$, and Eq. (16) reduces to

$$T_{\mathbf{a}} = \frac{1}{|\mathcal{U}(\mathbf{A})|} \sum_{i=1}^{|\mathcal{U}(\mathbf{A})|} {T_{\mathbf{a}}}^{\mathcal{U}(\mathbf{A})_i}. \tag{17}$$

This does *not* hold for noise mechanisms ignoring the SPSN principle (such as generic DP mechanisms), where independent noise is drawn for any $TI_{\mathbf{A}}$ and $TJ_{\mathbf{A}}$ with $I \neq J$, even though $\mathbf{A}$ is fixed, and Eq. (16) holds. Therefore, depending whether the principle is enforced or not, Eq. (16) or Eq. (17) can be used to average over all independent representations of the target count available in the output to obtain an estimate $\widetilde{T_{\mathbf{a}}}$. Note that in a static output scenario, no other independent representations of the target can be generated by the user: the amount of independent noise is predefined by the complexity of $\{TI_{\mathbf{A}_I}\}_{I\in\{1\cdots M\}}$ resp. $\{T_{\mathbf{A}_I}\}_{I\in\{1\cdots M\}}$.[25]

C.3 Disclosure risks

Eq. (3) relates the overall success probability of averaging the target count correctly to the (constant) variance V of the noise applied to each single independent output, where k is the number of independent outputs being summed and t is the number of independent representations of $T_{\mathbf{a}}$:

$$k_{\mathbf{a}} = \sum_{I\in\{1\cdots M\} \text{ where } \mathbf{A}\subset\mathbf{A}_I} \sum_{i=1}^{|\mathcal{P}(\overline{\mathbf{A}_I})|} \prod_{j=1}^{|\mathcal{P}(\overline{\mathbf{A}_I})_i|} \left|\mathcal{P}(\overline{\mathbf{A}_I})_{ij}\right| \quad \text{and} \quad t_{\mathbf{a}} = |\mathcal{T}_{\mathbf{a}}(\mathbf{A})| \tag{18}$$

resp.

$$k_{\mathbf{a}} = \sum_{i=1}^{|\mathcal{U}(\mathbf{A})|} \prod_{j=1}^{|\mathcal{U}(\mathbf{A})_i|} |\mathcal{U}(\mathbf{A})_{ij}| \quad \text{and} \quad t_{\mathbf{a}} = |\mathcal{U}(\mathbf{A})|. \tag{19}$$

However, Eq. (3) gives a *lower* limit on the averaging success probability, whereas an *upper* limit would be required from a protection point of view. Therefore, to assess averaging risks conservatively we continue the analysis under the assumption that the cumulated noise on the averaged estimate $\widetilde{T_{\mathbf{a}}}$ is again Gaussian with variance

$$\mathrm{Var}(\widetilde{T_{\mathbf{a}}}) = \frac{k_{\mathbf{a}} V}{t_{\mathbf{a}}^2}, \tag{20}$$

so that the averaging success probability $\alpha_{\mathbf{a}} = \Pr(|\widetilde{T_{\mathbf{a}}} - T_{\mathbf{a}}| < 0.5)$ can be calculated exactly. This approximation follows from the central limit theorem and performs well in the CK and Laplace noise scenarios discussed in this paper.[7]

[25] This is significantly different from the scenario of Asghar and Kaafar (2020), where the user can submit many queries asking for *custom* bi-partitions of any available variable; in our nomenclature this is equivalent to custom-defining $A = \{a_1, a_2\}$, which is impossible by assumption of this section (static output).

	With SPSN			Without SPSN		
#	MS	$T_{\mathbf{A}}$	opt. k/t^2	MS	$T_{\mathbf{A}}$	opt. k/t^2
1	LU	AGE.M	0.0867	all	total	0.0118
2	CY	AGE.M	0.0884	all	GEO.L	0.0170
3	MT	AGE.M	0.0916	all	AGE.M	0.0170
4	EE	AGE.M	0.108	all	SEX	0.0234
5	all	total	0.112	all	GEO.L×AGE.M	0.0237

Table 2: The top five smallest optimised k/t^2 values found for EU Member States (MS), with and without SPSN.

Eqs. (3) and (20) indicate that the ratio $k_{\mathbf{a}}/t_{\mathbf{a}}^2$ is a suitable measure of averaging risk, because it fixes $\alpha_{\mathbf{a}}$ for given V. However, generally the average over all available IRRs as in Eq. (16) or Eq. (17) does not give the lowest (i.e. most risky) $k_{\mathbf{a}}/t_{\mathbf{a}}^2$, because some IRRs require very many internal table cells to be added, thus increasing $k_{\mathbf{a}}$ and the corresponding summed noise amount disproportionately. This can be accounted for with a simple optimisation:

1. Sort all available IRRs from low to high individual $k_{\mathbf{a}}$ (fixing I and i in Eq. (18) resp. i in Eq. (19) and carrying out the product).

2. Start averaging iteratively, including new IRRs with increasing $k_{\mathbf{a}}$ one at a time, as long as the aggregate $k_{\mathbf{a}}/t_{\mathbf{a}}^2$ of the average decreases.

3. The first time the aggregate $k_{\mathbf{a}}/t_{\mathbf{a}}^2$ increases, discard the last IRR added and return the previous average as the optimal (i.e. most risky) one.

The risk of disclosure from averaging any target count $T_{\mathbf{a}}$ is thus fixed by the static output complexity $\{TI_{\mathbf{A}_I}\}_{I\in\{1\cdots M\}}$ resp. $\{T_{\mathbf{A}_I}\}_{I\in\{1\cdots M\}}$, as well as constant noise variance V. This means an output curator can control for it by either reducing output complexity or increasing V.

C.4 Results for the 2021 EU census scenario

The 2021 EU census programme (i.e. output table set) consists of $M = 103$ three to six-dimensional tables[10] cross-tabulating counts of natural persons by 32 different variable breakdowns.[10] While $t_{\mathbf{a}}$ is fixed for every $\mathbf{a}$ by the table set, for any target cell $T_{\mathbf{a}}$ without a geographic attribute $k_{\mathbf{a}}$ generally depends on one or more geographic breakdowns and thus on the reporting country, cf. Table 1. As expected, Table 2 shows that the smallest k/t^2 values (optimised as described in annex C.3) are found for the smallest countries, where geographic margins contributing to the independent representations have the smallest k weights. Furthermore, Fig. 7 shows the distribution of k/t^2 across all available $T_{\mathbf{A}}$ whose IRRs do not depend on any geographic breakdown (mostly where $\mathbf{A}$ contains itself a geographic breakdown, which is never crossed with another geographic breakdown in any of the output tables), i.e. those k/t^2 that are equally valid for all reporting countries.

Note the difference of almost an order of magnitude in Table 2 between the smallest k/t^2 when enforcing SPSN, and the smallest k/t^2 when ignoring it. Also in the distribution in Fig. 7, a sizeable share of the output statistics tends to have

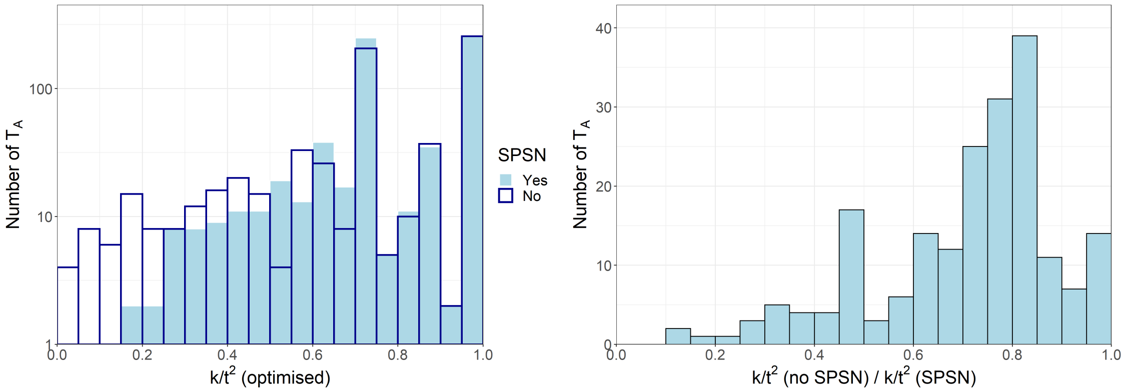

Figure 7: Histograms of the k/t^2 distribution across all $T_{\mathbf{A}}$ without geographic dependence in their averages, with and without SPSN (left), and the ratio of k/t^2 with and without SPSN for each of the $T_{\mathbf{A}}$ (right, ratio of 1 suppressed).

smaller k/t^2 without SPSN than with it.[26] While the difference is not too big for the majority of target $T_{\mathbf{A}}$, a systematic attack could first remove the noise from the most vulnerable $T_{\mathbf{A}}$ and then use those to reduce successively reduce the noise on subsequent $T_{\mathbf{A}}$ averages. Hence a conservative approach would fix V such that *every* $T_{\mathbf{A}}$ (even with the smallest k/t^2) is sufficiently unlikely to be averaged correctly. Section 3.2 discusses implications of this approach on generic noise parameter ranges when protecting 2021 EU census outputs.

References

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

Antal, L. et al. (2017). *Harmonised protection of Census data*. https://ec.europa.eu/eurostat/cros/content/harmonised-protection-census-data_en. Accessed on 26 Jan 2021.

Asghar, H. J. and Kaafar, D. (2020). "Averaging Attacks on Bounded Noise-based Disclosure Control Algorithms". *Proceedings on Privacy Enhancing Technologies* 2020.2, pp. 358–378. DOI: 10.2478/popets-2020-0031.

Bach, F. (2019). "Statistical Disclosure Control in Geospatial Data: The 2021 EU Census Example". *Service-Oriented Mapping: Changing Paradigm in Map Production and Geoinformation Management*. Ed. by J. Döllner, M. Jobst, and P. Schmitz. Cham: Springer International Publishing, pp. 365–384. DOI: 10.1007/978-3-319-72434-8_18.

[26] Among the two very frequent k/t^2 values apparent in Fig. 7, $k/t^2 = 0.75 \equiv 3/4$ stems from IRR sets that only contain the SEX breakdown ($k = 2$) and the trivial margin ($k = 1$) with total $k = 3$ and $t = 2$, which often happens for $T_{\mathbf{A}}$ representing very rare breakdowns. Similarly, $k/t^2 = 1$ reflects all the uniquely occurring $T_{\mathbf{A}}$, including all the internal cells of full tables $T_{\mathbf{A}_I}$, which never occur elsewhere as a margin so that their IRRs consist only of themselves ($k = 1$ and $t = 1$).

Bailie, J. and Chien, C.-H. (2019). "ABS Perturbation Methodology Through the Lens of Differential Privacy". *Joint UNECE/Eurostat work session on statistical data confidentiality*.

Bambauer, J., Muralidhar, K., and Sarathy, R. (2014). "Fool's gold! An illustrated critique of differential privacy". *Vanderbilt J. Entertain. Technol. Law* 16, pp. 701–755.

Costemalle, V. (2019). "Detecting geographical differencing problems in the context of spatial data dissemination". *Statistical Journal of the IAOS* 35.4, pp. 559–568. DOI: 10.3233/SJI-190564.

De Wolf, P.-P. et al. (2019a). *Perturbative confidentiality methods*. https://ec.europa.eu/eurostat/cros/content/perturbative-confidentiality-methods_en. Accessed on 26 Jan 2021.

— (2019b). *SDC Tools - User Support and sources of tools for Statistical Disclosure Control*. https://github.com/sdcTools. Accessed on 26 Jan 2021.

Dinur, I. and Nissim, K. (2003). "Revealing Information while Preserving Privacy". *Proceedings of the Twenty-second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pp. 202–210. DOI: 10.1145/773153.773173.

Dwork, C. (2011). "A Firm Foundation for Private Data Analysis". *Commun. ACM* 54.1, pp. 86–95. DOI: 10.1145/1866739.1866758.

Dwork, C., Kenthapadi, K., et al. (2006). "Our data, ourselves: Privacy via distributed noise generation". *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, pp. 486–503.

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

Dwork, C. and Naor, M. (2010). "On the Difficulties of Disclosure Prevention in Statistical Databases or The Case for Differential Privacy". *Journal of Privacy and Confidentiality* 2.1. DOI: 10.29012/jpc.v2i1.585.

Dwork, C. and Roth, A. (2014). "The Algorithmic Foundations of Differential Privacy". *Foundations and Trends in Theoretical Computer Science* 9.3–4, pp. 211–407. DOI: 10.1561/0400000042.

Dwork, C. and Rothblum, G. N. (2016). "Concentrated Differential Privacy". arXiv: 1603.01887 [cs.DS].

Dwork, C. and Smith, A. (2010). "Differential Privacy for Statistics: What we Know and What we Want to Learn". *Journal of Privacy and Confidentiality* 1.2.

Enderle, T., Giessing, S., and Tent, R. (2018). "Designing Confidentiality on the Fly Methodology - Three Aspects". *Privacy in Statistical Databases - UNESCO Chair in Data Privacy, International Conference, PSD 2018, Valencia, Spain, September 26-28, 2018, Proceedings*. Ed. by J. Domingo-Ferrer and F. Montes. Vol. 11126. Lecture Notes in Computer Science. Springer, pp. 28–42. DOI: 10.1007/978-3-319-99771-1_3.

Enderle, T., Giessing, S., and Tent, R. (2020). "Calculation of Risk Probabilities for the Cell Key Method". *Privacy in Statistical Databases - UNESCO Chair in Data Privacy, International Conference, PSD 2020, Tarragona, Spain, September 23-25, 2020, Proceedings*. Ed. by J. Domingo-Ferrer and K. Muralidhar. Vol. 12276. Lecture Notes in Computer Science. Springer, pp. 151–165. DOI: 10.1007/978-3-030-57521-2_11.

Fraser, B. and Wooton, J. (2005). "A proposed method for confidentialising tabular output to protect against differencing". *Monographs of Official Statistics: Work Session on Statistical Data Confidentiality*, pp. 299–302.

Garfinkel, S. L. (2019). "Deploying Differential Privacy for the 2020 Census of Population and Housing". *JSM 2019 Session: Formal Privacy - Making an Impact at Large Organizations*.

Garfinkel, S. L., Abowd, J. M., and Martindale, C. (2018). "Understanding Database Reconstruction Attacks on Public Data". *Queue* 16.5, pp. 28–53. DOI: 10.1145/3291276.3295691.

Ghosh, A., Roughgarden, T., and Sundararajan, M. (2012). "Universally utility-maximizing privacy mechanisms". *SIAM Journal on Computing* 41.6, pp. 1673–1693.

Giessing, S. (2016). "Computational Issues in the Design of Transition Probabilities and Disclosure Risk Estimation for Additive Noise". *Privacy in Statistical Databases - UNESCO Chair in Data Privacy, International Conference, PSD 2016, Dubrovnik, Croatia, September 14-16, 2016, Proceedings*. Ed. by J. Domingo-Ferrer and M. Pejic-Bach. Vol. 9867. Lecture Notes in Computer Science. Springer, pp. 237–251. DOI: 10.1007/978-3-319-45381-1_18.

Hsu, J. et al. (2014). "Differential Privacy: An Economic Method for Choosing Epsilon". *2014 IEEE 27th Computer Security Foundations Symposium*. Vol. 2014-January, pp. 398–410. DOI: 10.1109/CSF.2014.35.

Machanavajjhala, A. et al. (2008). "Privacy: Theory meets Practice on the Map". *IEEE 24th International Conference on Data Engineering (ICDE)*, pp. 277–286. DOI: 10.1109/ICDE.2008.4497436.

Marley, J. K. and Leaver, V. L. (2011). "A Method for Confidentialising User-Defined Tables: Statistical Properties and a Risk-Utility Analysis". *Int. Statistical Inst.: Proc. 58th World Statistical Congress (Session IPS060)*, pp. 1072–1081.

Meindl, B. and Enderle, T. (2019). "cellKey - consistent perturbation of statistical tables". *Joint UNECE/Eurostat work session on statistical data confidentiality*.

Petti, S. and Flaxman, A. (2019). "Differential privacy in the 2020 US census: what will it do? Quantifying the accuracy/privacy tradeoff". *Gates Open Research* 3. DOI: 10.12688/gatesopenres.13089.2.

Rinott, Y. et al. (2018). "Confidentiality and Differential Privacy in the Dissemination of Frequency Tables". *Statistical Science* 33, pp. 358–385. DOI: 10.1214/17-STS641.

Ruggles, S. et al. (2019). “Differential Privacy and Census Data: Implications for Social and Economic Research”. *AEA Papers and Proceedings* 109, pp. 403–08. DOI: 10.1257/pandp.20191107.

Santos-Lozada, A. R., Howard, J. T., and Verdery, A. M. (2020). “How differential privacy will affect our understanding of health disparities in the United States”. *Proceedings of the National Academy of Sciences* 117.24, pp. 13405–13412. DOI: 10.1073/pnas.2003714117.

Thompson, G., Broadfoot, S., and 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.*

Wang, Y., Lee, J., and Kifer, D. (2015). “Revisiting differentially private hypothesis tests for categorical data”. arXiv: 1511.03376 [cs.CR].