**Statistical confidentiality: New initiatives in the European Statistical System**

Fabian Bach, Eurostat, fabian.bach@ec.europa.eu

Wim Kloek, Eurostat, wilhelmus.kloek@ec.europa.eu

Aleksandra Bujnowska, Eurostat, aleksandra.bujnowska@ec.europa.eu

**Abstract**

*The protection of confidential information has a huge impact on how statistical data can be published and used for analysis, which makes it a key aspect of data quality. This paper presents new methods and tools currently being investigated in the ESS in order to publish more – and more useful – data without compromising statistical confidentiality. It covers new methodological and IT developments, where concrete use cases demonstrate their impact on data quality. For instance, a promising methodological direction is random noise: several ESS use cases at different maturity stages are presented, including recommendations for the harmonised protection of 2021 EU Census data. Another direction is to reflect at a more fundamental level where protection is needed. Several ideas will be presented along this line.*

**Keywords:** Statistical confidentiality, data protection, disclosure control, random noise

1. Introduction

Data are considered confidential when they allow statistical units (persons, businesses, organisations) to be identified, either directly or indirectly, thereby disclosing individual information (Regulation (EC) 223/2009 Article 3). Therefore, *statistical disclosure control* (SDC) methods are required to reduce the risk of disclosing information on the statistical units before data can be published. These methods are usually based on restricting reducing the information or modifying the data before publication (Regulation (EC) 557/2013 Article 2). Eurostat is actively involved in the methodological advancement of SDC methods for the European Statistical System (ESS)[[1]](#footnote-1), as well as in the maintenance and development of SDC software tools.[[2]](#footnote-2)

In general there is a distinction between the protection of the privacy of individual persons and the protection of the business interests of companies and organisations. Section 2 focusses on new SDC methods based on data perturbation (adding noise), which are most suitable for categorical data tabulating individuals – i.e. the social statistics regime – while business statistics often consist of quantitative data. On the other hand, businesses can also have an interest in the *disclosure* of their data if it implies also the publication of statistical information on their market, so that they can position themselves. Thus, section 3 presents three ideas which are of a completely different nature than the technical solutions of section 2: they rather question the need for protection in the first place, under certain circumstances. Finally, section 4 concludes the paper.

1. New SDC Methods based on perturbation
	1. Modernisation needs for statistical disclosure control

When it comes to mitigating residual disclosure risks in public tables aggregated from confidential microdata, the traditional approach used in statistical offices is cell suppression, i.e. risky cells in tables to be published are identified and their value is not published ("suppressed"). While such primary suppression is generally easy to carry out (and to understand by the user) there are various well-known drawbacks, e.g.

* secondary suppression: especially in complicated tables with many cross-tabulated dimensions, many inherently safe cells need to be suppressed to avoid that primary suppressions can be re-calculated by subtracting the sum of all other cells in a breakdown from the public breakdown total;
* consistency: when many tables/data products are published from the same source microdata, it is often difficult (or impossible) to apply suppression in a consistent manner across all products, i.e. to avoid that suppressed information can be reconstructed by comparing various public data;
* data analysis is seriously hampered by suppression.

Despite these flaws, suppression is still rather common in many offices and statistical domains. In Eurostat, where suppressed cells are usually flagged with 'c' and published with the value ':', this leads to various issues:

* human intervention is often needed for (primary/secondary) suppression, e.g. feedback loops between Eurostat and data providers or data users (depending on the specific production process in a given statistical domain);
* EU totals cannot be published if one or very few Member States report a national aggregate as confidential, because secondary suppression is ruled out by the obligation to publish transmitted (non-confidential) data;
* 'c' flags are sometimes misused / misunderstood by data providers to hold back unreliable data, considered less accurate due to methodological issues (low statistical significance, number of observations, etc.)
* tables with confidential cell cannot be analysed directly, but need some pre-processing by the user.

For these reasons, Eurostat aims to support the modernisation of confidentiality treatment, where a focus is on more efficient and automatable methods.

* 1. Confidentiality on the fly as a general concept

Indeed there was lively development over the recent years: various statistical offices have been experimenting with random noise methods to overcome cell suppression and increase utility of their data for users. For instance, the Australian Bureau of Statistics (ABS) has already changed its publication channel for census and most other population data products[[3]](#footnote-3), while Statistics New Zealand (Stats NZ) has proposed a variant to protect magnitude tables in business statistics[[4]](#footnote-4). Various other countries (among them e.g. Canada, the United Kingdom and Norway) are currently testing respective methods and implementations in various statistical domains.

Under the headline "confidentiality on the fly", and based on several ongoing pilot projects Eurostat plans to develop a generic and versatile instance of automatised confidentiality via random noise – ultimately implemented into a user-friendly front-end tool. The resulting software should complement Eurostat's data access services by innovative ad-hoc request tools, but ideally also be modularised and flexible enough for sharing with other statistical offices in the ESS.

Once implemented, "confidentiality on the fly" will have various striking advantages – both to **data providers**, i.e. Eurostat production units dealing with applicable microdata:

* automatic and consistently safe confidentiality treatment out of the box (no human intervention / secondary suppression needed);
* protection setup details (type / size of noise) adaptable to specific data needs;
* EU totals always publishable;
* custom table builder service available out of the box;

as well as to **data users**:

* one central access tool for custom tables on all kinds of applicable statistical domains where Eurostat has microdata (table builder);
* no more 'c' flags / suppressions, i.e. more data available with well-defined and quantifiable noise intervals.

The methodological basis, current state and envisaged next steps of "confidentiality on the fly" within Eurostat are outlined in the following sections. Some drawbacks resp. reservations typically brought up against the proposed method should also be mentioned – these will be addressed at the end of the next section, after the method itself has been introduced.

* 1. Cell key method

The general term random noise denotes the random application of small perturbations to the data – either directly to the microdata or to aggregated frequencies/magnitudes in output tables. In the latter case, it typically consists of a predefined probability distribution for the noise to be applied to the data, combined with a mechanism to randomly draw from that noise distribution in a consistent manner. Particular variants include the "cell key method" originally proposed by the ABS. A generic implementation of this method involves two to three distinct steps:

1. *Cell key module*: assigns a fixed random number, or "cell key", to each cell of the output table. The ABS variant ensures by design that the noise applied to a specific cell will always be exactly the same, even if the cell appears in different tables – this is meant by consistency of the method.
2. *Noise module*: draws the noise for each output table cell from the predefined noise probability distribution, based on the cell's random number (cell key). The overall noise magnitude is controlled through this probability distribution.
3. *Additivity module* (optional): obviously random noise breaks the additivity of the output tables. Additivity can be restored in a post-processing step, but this spoils consistency across several tables. Hence, Eurostat resp. the relevant statistical domains would need to assess carefully which property they consider more important.

See the annex for an illustration on an example table.

While the ABS variant applies additive noise to frequency tables (e.g. ±1, ±2, etc.), which is appropriate for person counts in the census and can be straightforwardly extended to survey tables (see section 2.4.2), Stats NZ proposes multiplicative noise (data rescaling e.g. around ±10%) applicable to *microdata*, for instance in business statistics. Both variants essentially share the cell key module above, while differing in their noise module. Thus it would be straightforward to integrate both variants as operating modes in the same random noise software architecture. In either case, the global statistical noise variance applying to each cell value of the output table is a pre-defined method parameter: this is a strong advantage especially for experienced users interested in quantitative data analytics including reliability estimations.

However, there are also two main drawbacks often formulated against the various flavours of random noise added to output tables: on one hand, many users seem to expect that tables are exactly additive, which can be reinstated here only at the cost of some loss of consistency – see method module (3) above. On the other hand, there are concerns that users may perceive the applied noise as a disproportionate damage to data precision, or that the general concept of noise variances may be too complicated/obscure for the "average use". Both issues can and should be addressed by essentially the same set of good arguments:

* the amount of noise required for a sufficient protection (and hence level of non-additivity) is almost always smaller than the numerous other sources of data inaccuracies, such as measurement, processing or statistical errors;
* the perturbation created by random noise is controlled by parameters, and it has been demonstrated[[5]](#footnote-5) how to set up parameters in an optimal way to minimise the information loss at a given protection level;
* as mentioned above, a pre-defined public global variance improves data utility especially for quantitative analytics.

Hence, well-prepared, clear and – where appropriate – didactic communication about general data accuracy vis-à-vis random noise effects is essential here.

* 1. Data pilots

Before translating the methodology described above into some kind of multi-purpose software, it is reasonable to validate the random noise approach in several piloting statistical domains. This should prove the versatility and fitness of the method variants for different domain requirements. Moreover, it fosters exchange and collaboration between methodological and domain experts at an early stage, thus facilitating a common understanding of the scope and gains from the outset.

* + 1. 2021 EU Census

Between 2016 and 2017, Eurostat hosted a dedicated project "Harmonized protection of census data in the ESS" to develop recommendations for confidentiality treatment during the next EU Census round in 2021. The project recommends a combination of two perturbation methods for general use in the ESS to protect census hypercubes as well as 1 km square grid data, namely

1. targeted record swapping: each risky household is paired with a similar household, then the geographic locations of each pair are swapped;[[6]](#footnote-6)
2. an adapted version of the ABS cell key method.

Extensive documentation, including SAS code and first test results, can be found on the project's CROS portal page.[[7]](#footnote-7) The project results show that the methods are in principle appropriate for census data, and can be implemented in a straightforward way. Member States can (and are invited to) perform own test to assess if the method is suitable for production in their situation.

While not technically integrated with the "confidentiality on the fly" project, this use case thus represents a valuable data pilot for the methodology, i.e. on-the-fly protection of table data in production using the cell key method. Moreover, the imminent method implementation for interested NSIs may provide first insights and lessons learned for the planned development in Eurostat. Finally, if many NSIs adopt the method for their 2021 census production (which currently seems likely), this will also increase acceptance in the greater social statistics community within the ESS.

On final note, **data geocoded to geographical grids** are coming more and more into the focus of social statistics. In particular, the ESS members intend to publish 13 harmonised key variables from the 2021 EU Census on a common European 1 km square grid.[[8]](#footnote-8) This pilot will prepare the stage for a planned recurring/evolving grid data component of the ESS post-2021 census strategy. While the project mentioned above took particular care to include these grid data in their recommendations, it currently seems that standard configurations of the cell key method cannot fully handle specific user requirements for grid data. For instance, the method may still distort the *distinction* between populated and unpopulated areas to an unacceptable degree. In their 2021 pilot exercise, the ESS members foresee to deal with this requirement through a metadata-based (flag) approach.[[9]](#footnote-9)

* + 1. Ad-hoc tables from Labour Force Survey data

From a methodological point of view, a first step towards this generalization is the currently ongoing Eurostat effort to test the cell key method for Labour Force Survey (LFS) data. More specifically, Eurostat provides a service for ad-hoc table extractions from LFS data, where users can request tables with weighted results (unweighted results are not provided). In this process, all table cells based on 3 or less unweighted sample observations are flagged with 'c' and suppressed in the output. If the share of 'c' flags in a request is below a certain threshold the table is delivered, otherwise Eurostat rejects the request and asks the user to revise it (feedback loop). Possibly in combination with same basic rules for extraction requests (e.g. limits on table dimensions and/or total number of cells), the cell key method could render this service considerably more efficient, ultimately making the described feedback loop with users obsolete.

While the cell key method has originally been developed for population census type data, implying no weights or weights close to 1, the generalisation to weighted data seems straightforward. In fact, the ABS version (footnote 3) already contains a solution: apply the noise to the *unweighted* number of observations in each cell and then multiply the result with the average weight per observation of this cell.

In November 2017, Eurostat conducted a first consultation with some key LFS users to find out if cell key results based on the ABS idea would meet their requirements: while several testers had technical and/or methodological questions that could be clarified by Eurostat, no one raised fundamental concerns obstructing the applicability or utility of the method from a user perspective. Finally, in December the approach was also presented to the Working Group on Labour Market Statistics, where the general reaction was tentatively positive. A small number of Member States expressed concerns about the protection level and lack of experience in production, and recommended further consultations.

From an IT perspective, the Eurostat Units concerned (LFS and Methodology) jointly developed a deployment plan to roll out the method as a part of the existing software for producing ad-hoc table extractions. After the aforementioned positive test phase, the deployment is currently work in progress. As a result, the method will be available soon to protect ad-hoc tables from those Member States that wish it.

* + 1. Possible further pilots

With these first promising results, we propose to build further test cases covering more statistical domains and output types. Optimally this will increase acceptance of random noise methods on the part of data providers as well as data users, and ultimately prepare the floor for the potential roll-out of new Eurostat software solutions for data access/discovery services.

* 1. Table Builder as a new web service

While the cell key method is quite straightforward to apply, it is clear from the generic algorithm that the method is closely linked with the actual procedure of building the output table, i.e. counting records in the microdata with certain attribute combinations defined by each table cell and formatting these counts into an output table. The added value of "confidentiality on the fly" will thus be maximized when integrated with a similarly automatized table builder tool. Such a tool would need to be linked directly to the underlying microdata, while providing a front-end user interface for requesting custom tabulations according to some predefined rules. ABS already published such a tool for census and various other statistical domains (see footnote 3), supporting basic functionalities free of charge and more powerful functionalities for (paying) subscribers. The United Kingdom's Office for National Statistics (ONS) recently reported progress towards a similar tool for their 2021 census results.[[10]](#footnote-10)

If the methodological pilots continue to evolve positively, Eurostat plans to develop a stand-alone software solution, integrating random noise methods with a table builder framework while addressing the particular needs of its statistical domains and data users. One important input to this will be the deliverables of a new project "Open source tools for perturbative confidentiality methods" in the Centre of Excellence on SDC (see footnote 2). In particular, two key deliverables in this context will be:

1. *cell key method implementations* in some existing stand-alone confidentiality tools maintained by the ESS (τ-Argus and sdcTable, linked in footnote 2) including one or more options for weighted data – expected still 2018;
2. *technical specifications* for a generic and user-friendly table builder web service incorporating confidentiality on the fly, accounting for user needs (e.g. graded access for public / scientific / internal users) as well as particularities of different statistical domains (e.g. weighted vs. unweighted counts, frequencies vs. magnitudes) – expected early 2019.
3. Other Solutions for Business Statistics

The use of the cell key method with skewed quantitative data (notably in business statistics) is not obvious. Statistics New Zealand proposes for instance multiplicative noise (quantitative *microdata* rescaling, e.g. around ±10%) for business statistics, but other versions of multiplicative noise for magnitude tables are thinkable. Anyway, the perturbations necessary for sufficient protection will damage to some extent the patterns in the data and may thus undermine the usefulness. So are there other solutions available to avoid suppression in the case of skewed quantitative data? This section will present three ideas.

* 1. Passive confidentiality

By default statistical institutes are cautious and always take measures to protect data with a disclosure risk (e.g. due to small numbers or to dominance). In the case of passive confidentiality, protection measures are taken only at the request of a business. At the EU-level passive confidentiality is currently only applied in statistics on trade in goods[[11]](#footnote-11); some national statistical institutes apply passive confidentiality also in other domains.

The advantage of passive confidentiality is that it significantly reduces information loss due to suppression of confidential cells. The disadvantage is that businesses are less protected. They might not have noticed the regime of passive confidentiality or they might have miss-judged the risk involved in the disclosure of their information. All this depends a lot on the quality of communication to the businesses and on the quality of internal communication inside the businesses. At the EU level passive confidentiality can only be introduced by the European law where its use has to be properly justified by the need for detailed information and an analysis of the potential harm to the businesses concerned. The harm is less, for instance, if the same or similar information is already publicly available from administrative sources.

Passive confidentiality is used for the survey as a whole, but it is perfectly possible to apply it in a selective way: for some variables, for certain size-classes or for certain economic activities. Selective application might be more difficult to communicate and to manage.

* 1. Use of waivers[[12]](#footnote-12)

A waiver is an explicit permission of respondents to disclose their data. Waivers can be applied in very different ways. Permission could be asked after the data collection to allow the publication of specific cells. The advantage is that the requests can really be targeted on the businesses that matter most for the availability of data. The disadvantage is that it comes late in the statistical production process, so that the collection of waivers could cause delays in the publication.

Permission can also be asked during data collection. The advantage is that it requires just one contact; it is also clear and explicit what the permission concerns. It is also an advantage that the permission can remain directly linked to the data throughout the statistical production process. The disadvantage is that the permission only concerns one data collection wave.

Finally, permission can be asked before data collection. The advantage is that it could concern several waves of the same survey. It could even concern a set of surveys. The disadvantage is that the information on the existing waivers should be well managed and shared. This requires staff, procedures and tools.

Waivers can be applied in a systematic (covering all or a broad base of businesses) and in a targeted way (covering the most significant). The systematic approach has the biggest overall impact. Which approach is the most cost-effective depends on the details of the organisation and on available tools.

The use of waivers can reduce the number of suppressed cells significantly. Experiments in the domain of industrial production statistics with a detailed classification by type of product show reductions in the number of confidential cells from 5 to over 40 percent points.

* 1. Expiration date

In most statistical systems the confidentiality of data has no expiry date; data once defined as confidential, remain confidential. There are exceptions, for instance in the UK the microdata from the census becomes publicly available after 100 years.[[13]](#footnote-13) In France microdata from business surveys is no longer confidential after 25 years (and microdata from household surveys after 75 year or 25 years after the death of the person concerned).[[14]](#footnote-14)

A difference between the two examples is that in the UK the publication of the microdata is actively prepared, whereas in France no active dissemination is foreseen. Envisaging the publication of microdata after a long time period has some disciplining effect on archiving the data and metadata.

Old microdata are mainly interesting for research. Most other types of users (notably policy and market analysts) are mainly focussed on the near future and use data on the recent past as proxy or as basis for extrapolation. Currently hardly any microdata on businesses are available for research, whereas microdata from the key social surveys are available, generating a significant body of research.

In case of active publication of microdata after the expiration data, what should happen with the related aggregate tables that were published by the statistical institutes? Current production processes have not been designed to do the revision after for instance 25 years. Revision will be very expensive or nearly impossible and there will be little demand for it. Moreover, a revision would disclose the protection method, making it easier for possible intruders to disclose other information.

Introduction of an expiration date for microdata on businesses seems to be a good approach to making some microdata available to researchers. It builds on the idea that the harm that can be done to the business interests of the companies concerned quickly reduces over time (the delays could dependent on the specific domain). Past data on businesses are also inherently protected because of continuous restructuring of the entities (change of ownership, merger, take-over, etc.), which creates considerable uncertainty of the exact delineation of the unit.

The introduction of an expiration date for microdata on businesses will not be realised in the short run. It requires:

* discussion with stakeholder groups;
* preparation in the statistical production environment;
* possibly legal changes.

At the EU level legal changes might not be necessary if access is provided to recognised research organisations for research purposes. But the actual availability still depends on the approval by the data owners (the national statistical authorities). Several data owners could have legal restrictions at the national level. Moreover, for most statistical domains in business statistics Eurostat currently only collects aggregate data from the Member States; the microdata would have to be added.

1. Conclusions

Protection of statistical data against disclosure through suppression limits the usability. For instance, tables with suppressed cells cannot be analysed directly with standard tools, but require some assumption on the suppressed cell values: different users will make different assumptions and produce different results. The key advantage of additive noise as a protection method is that users will receive complete tables with a controllable global accuracy (absolute noise variance added to cell values). Depending on the required protection level and the data source the noise will well remain within the intrinsic uncertainty of the original data (measurement, processing, …).

Noise protection methods are broadly accepted and they are being integrated in the standard statistical disclosure control tools of many statistical authorities across the world. The methods using additive noise function well in the domains of social statistics with predominantly qualitative (categorical) variables. In business statistics often featuring quantitative (magnitude) variables and skewed distributions these methods need to be carefully adapted. For instance, a variation using multiplicative noise (i.e. multiplying a noise factor to the original data value) is under investigation, but feasible trade-offs between data protection and utility still need to be re-assessed on a more fundamental level. Therefore some other ideas were listed to deal with confidentiality in business statistics: passive confidentiality, use of waivers and expiration data. These are not methods for treating confidential data, but rather ideas that question where protection is needed and proportional.

**Annex – Illustration of the Cell Key Method Algorithm**

1) Assign each record a random number (***record key*** "Rkey"), for example here between 1 and 200

|  |  |
| --- | --- |
| Record | Rkey |
| r1 | 54 |
| r2 | 104 |
| r3 | 93 |
| … | … |
| rN | 26 |

2) Create the frequency table. For each cell, sum Rkeys and take the modulo to get the ***cell key***

|  |  |
| --- | --- |
| Record | Rkey |
| r2 | 104 |
| r4 | 61 |
| r56 | 7 |
| r72 | 90 |
| Sum Rkey = 262 |
| Cell key = 262 mod 200 = 62 |

|  |  |  |
| --- | --- | --- |
| **Age by Sex** | Male | Female |
| 0-15 | . | . |
| 16-24 | . | 4 |
| 25-34 | . | . |
| … |  |  |

3) Use perturbation table (***p-table***) to get perturbation value from cell value and cell key

|  |  |
| --- | --- |
| **p-table** | Cell key (1-200) |
| 1 | 2 | 3 | … | 62 | … | 200 |
| Cell Value | 1 |  | +1 |  |  |  |  |  |
| 2 | -1 |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  | -1 |
| 4 | +1 |  |  |  | +1 |  |  |
| 5 |  |  | -1 |  |  |  |  |
| … |  |  |  |  |  |  |  |

4) Apply the chosen perturbation to the cell

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Age by Sex** | Male | Female |  | **Age by Sex** | Male | Female |
| 0-15 | . | . | → | 0-15 | . | . |
| 16-24 | . | 4 | 16-24 | . | 5 |
| 25-34 | . | . | 25-34 | . | . |
| … |  |  |  | … |  |  |

**Note:** This example provided by the UK's ONS is based on unweighted frequencies (person counts) as appearing e.g. in census data. However, ABS proposes in its original paper (footnote 3) a straightforward extension to weighted population counts: apply the noise to the *unweighted* number of observations in each cell and then multiply the result with the average weight per observation of this cell.

1. The ESS is the joint body of Eurostat and the NSIs of all EU countries and Iceland, Liechtenstein, Norway and Switzerland. It is responsible for the development and quality assurance of official European statistics. [↑](#footnote-ref-1)
2. Note for instance the Centre of Excellence on SDC: <https://ec.europa.eu/eurostat/cros/content/centre-excellence-statistical-disclosure-control-0_en>; ongoing open-source software developments are available at: <https://github.com/sdcTools> [↑](#footnote-ref-2)
3. <http://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder>;
see <http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2013/Topic_1_ABS.pdf> for a more detailed explanation. [↑](#footnote-ref-3)
4. <http://www.stats.govt.nz/browse_for_stats/businesses/business_characteristics/new-method-for-confidentialising-tables.aspx> [↑](#footnote-ref-4)
5. E.g. Gießing S (2016) Computational issues in the design of transition probabilities and disclosure risk estimation for additive noise. Proceedings of the 2016 International Conference on Privacy in Statistical Databases, <https://link.springer.com/book/10.1007/978-3-319-45381-1> [↑](#footnote-ref-5)
6. The definition of a "risky" household is based on a given geographic breakdown (e.g. NUTS, LAU), where matched pairs for swapping need to be located in different regions of this breakdown. [↑](#footnote-ref-6)
7. <https://ec.europa.eu/eurostat/cros/content/harmonised-protection-census-data_en> [↑](#footnote-ref-7)
8. See also the Q2018 contribution A Wroński, D Thorogood, F Bach: "Making census statistics more relevant – towards geo-enabled statistics" [↑](#footnote-ref-8)
9. For details see F Bach: "Statistical Disclosure Control in Geospatial Data: The 2021 EU Census Example", to be published in J Döllner, M Jobst, P Schmitz (Eds.): "Service Oriented Mapping" (chapter 18), <https://www.springer.com/gp/book/9783319724331> [↑](#footnote-ref-9)
10. <https://statswiki.unece.org/download/attachments/129174390/UNECE%20Sept%202017%20Spicer%20and%20Dove%20ONS.docx?version=1&modificationDate=1503501685961&api=v2> [↑](#footnote-ref-10)
11. This principle is recommended by the United Nations in its IMTS 2010 publication and set out in the EU legislation for detailed statistics on intra- and extra-EU trade. [↑](#footnote-ref-11)
12. This subparagraph is largely based on [*Use of waivers to reduce confidentiality suppressions in business statistics*](https://www.conference-service.com/NTTS2017/documents/agenda/data/x_abstracts/x_abstract_298.docx) (ALA-KIHNIA, BUJNOWSKA and KLOEK, 2017) [↑](#footnote-ref-12)
13. <https://en.wikipedia.org/wiki/Census_in_the_United_Kingdom> [↑](#footnote-ref-13)
14. Guide du secret statistique: <https://www.insee.fr/fr/information/1300624> [↑](#footnote-ref-14)