GUIDE TO STATISTICAL CONFIDENTIALITY

Like any public official, the agents of the official statistic service are bound by professional confidentiality with regard to confidential data which they may hold in the course of their duties or functions (Article L121-6 of the General Code of the Civil Service). In the event of a breach of this obligation, they are liable to criminal penalties of up to one year's imprisonment and a fine of 15,000 euros (Article 226-13 of the Penal Code).

Statistical confidentiality reinforces the obligation of professional confidentiality when it applies to **confidential data that the official statistical service holds, collects or produces for statistical purposes**. Statistical confidentiality guarantees data subjects that their data may only be used or re-used for statistical, scientific or historical research purposes, and excludes the use of such data in support of any measure or decision taken in respect of a particular person.. Statistical confidentiality may accordingly be invoked against any request for information, whether from a judicial authority or an administrative authority (e.g. tax or customs authorities).

Statistical confidentiality thus guarantees, in particular, the respect for:

- **Confidentiality due to private, personal and family life**, for natural persons;
- **Commercial and business confidentiality**, for companies.

The obligations relating to statistical confidentiality are based on legal texts at national and European level.

By law, it applies both to information from surveys (article 6) and to data from administrative sources (article 7 bis).

Statistical confidentiality at the national level derives from the obligations defined by the no. 51-711 Act of 7 June 1951 (amended) on legal obligation, coordination and confidentiality in the field of statistics. By law, it applies both to information from surveys (article 6) and to data from administrative sources (article 7 bis)[1].

---

[1] See recitals 5 and 6 of Council of State decision no. 472883 of 31 May 2024.

Updated: March 2025

At European level, the confidentiality of statistical information has been established in Article 338 of the Treaty on European Union.

Statistical confidentiality is covered in Chapter V of Regulation 223/2009 (amended) and implementing Regulation 557/2013 regarding access to confidential data for statistical purposes. Regulation 223 prohibits any communication of confidential data obtained for the production of European statistics, except for statistical or research purposes. As these data are subject to disclosure rules defined by higher-ranking European standards, they are not subject to any disclosure or dissemination rules strictly applicable under national law.

Statistical confidentiality also depends on the provisions defined by the General Data Protection Regulation (GDPR) concerning the purpose principle. This prohibits any re-use of personal data from statistical processing for purposes other than statistics or research[2], which in practice means that statistical confidentiality is imposed on the dissemination or communication of these data.

Statistical confidentiality applies both to individual data and to the aggregated results obtained from them, insofar as these aggregated results could allow them to be re-identified[3]. The obligation to verify the application of statistical confidentiality to published results leads to the implementation of various anonymisation or secretisation techniques, in particular the application of 'statistical confidentiality rules'.

---

[2] See Recital 162 of the GDPR: "*Statistical purposes mean any operation of collection and the processing of personal data necessary for statistical surveys or for the production of statistical results. Those statistical results may further be used for different purposes, including a scientific research purpose. The statistical purpose implies that the result of processing for statistical purposes is not personal data, but aggregate data, and that this result or the personal data are not used in support of measures or decisions regarding any particular natural person.*"

[3] See recital 6 of the aforementioned Council of State decision: "*It follows that the online publication of statistical data derived from individual information protected by statistical confidentiality pursuant to Article 6 of the Law of 7 June 1951 is only possible, before the expiry of the time limits mentioned in that Law and* **whatever the level of aggregation of that data**, *provided that the natural or legal persons from whom the individual information has been collected, who constitute the statistical observation units of the survey in question, cannot be identified, directly or indirectly, taking into account all the means which could reasonably be used by a third party having access to the data thus disseminated.* "

Contents

Index

Updated: March 2025

Practical rules for ensuring compliance with statistical confidentiality

A distinction must first be made between information collected through statistical surveys and information transmitted to the official statistical system by third parties for the purpose of producing statistics (administrative data).

Statistical surveys are those which have received the approval provided for in Article 2 of the no. 51-711 Act of 7 June 1951 on legal obligation, coordination and confidentiality in the field of statistics and which appear in the list of official statistics service surveys published every year in the Official Journal.

A. Information drawn from statistical surveys

Confidentiality must be respected in tables of aggregated data (see appendix) and files containing individual data.

1. Aggregated data tables

a) For tables providing aggregated data on **businesses**, the rule is as follows:

- No cell in the table should refer to fewer than three units (decision of 13 June 1980 by the Director General of INSEE);

- No cell in the table should contain data of which over 85% of the total relates to a single company (dissemination rule defined on 7 July 1960 by the Statistical Survey Coordination Committee, predecessor of the CNIS, the French National Council for Statistical Information).

b) For tables providing aggregate data on natural persons, it must not be possible to identify, in principle, the persons (Article L.312-1-2 of the Code on Relations between the Public and the Administration). Criteria applicable in the case of dissemination have been defined by the Cnil ("Anonymization of data, a key treatment for open data"). In particular, the Cnil underlines the risks linked, in the case of aggregated data, to re-identification **by inference**, corresponding to cases where it is possible to deduce with near certainty new information about an individual[4].

Nonetheless, in accordance with Article 116 of Decree No. 2019-536 of 29 May 2019 for the application of Act No. 78-17 of 6 January 1978 on data processing, files and

---

[4] Example (taken from the Cnil): "If a supposedly anonymous dataset contains information on the amount of taxes of people who have answered a questionnaire, that all the men between 20 and 25 years old who have answered are non-taxable, it will be possible to deduce, if we know that Mr. X, a 24-year-old man, has answered the questionnaire, that he is non-taxable. "

freedoms, data from processing for statistical purposes "*may not be disseminated without first being anonymised* **unless the interests of third parties in such dissemination prevail over the interests or fundamental rights and freedoms of the person concerned.**" This means that the balance between the interest of the public on the one hand and that of the persons concerned on the other hand may allow, depending on the case, and in particular on the sensitivity of the data being disseminated, to adjust dissemination methods that do not strictly respect the principle of prior anonymisation, while guaranteeing a sufficient level of protection for the privacy of the data subjects.

Article D.312-1-3 of the Code on Relations between the Public and the Administration, taken for application of the aforementioned article L.312-1-2, which authorizes the publication of certain categories of data without prior anonymization **when they are necessary for public information**, also grants margins of appreciation for the definition and application of confidentiality rules.

Specific rules of dissemination were thus enacted for the population census, given its exhaustive nature in small municipalities. These rules are contained in the Decree of 19 July 2007 concerning the dissemination of the results of the population census.

2. Individual data files

a) It is not possible to provide **individual data on legal entities** while preserving strict anonymity, unless some variables which are considered as essential for analysing these files are deleted: economic activity, location, size indicator.

Economic and financial data collected through statistical surveys are freely accessible after twenty-five years have elapsed following the date the survey was completed. To meet the occasional need for information on companies before this period has elapsed, a 1984 Act created a **Statistical Confidentiality Committee[5]**. It may propose exemptions from access to individual business data in response to requests made for statistical, or scientific or historical research purposes.

b) **Household surveys[6]** carried out by INSEE usually produce two types of individual data files:

- Files for the "general public", made available on the INSEE website. These files are anonymised, which means they are designed in such a way as to make it impossible to identify an individual, either directly or indirectly. To ensure this, a

---

[5] The competence of the Statistical Confidentiality Committee, since its creation, has gradually been extended to all statistical sources on companies and households, as well as to tax sources, and since the **Digital Republic Act** (2016), to all databases held by administrations, provided that the latter make a request to the Committee.

[6] With the exception of censuses and the annual census surveys, which are subject to special provisions.

Updated: March 2025

certain number of variables are removed from the original file: municipality of residence, details of profession, and perhaps some other variables specific to the survey which would make it possible to identify a person.

- So-called "production and research" files. These files do not contain identification elements and are processed by grouping or reclassification methods so as to minimise the risk of re-identification by recombining variables. They retain more detailed information than that in the "general public" file to cater for the specific needs of researchers and public bodies like the Ministerial Statistical Offices. To avoid any risk of a breach of confidentiality by cross-referencing with other data sources, these files are only made available through the Statistical Confidentiality Committee, using a similar, though simplified, procedure to that used for the non-anonymised individual data (see below).

Data pertaining to private and family life, and more generally to facts and behaviour of a private nature, become freely accessible after a period of seventy-five years, or twenty-five years after the death of the person concerned, whichever comes first.

Non-anonymised individual data obtained from household surveys can be accessed, before these time periods have elapsed, for statistical, or scientific or historical research purposes, after receiving a favourable opinion from the **Statistical Confidentiality Committee** and based on a decision from the administration of the Archives. The procedure is the same as for access to individual information on businesses (see above).

At its meeting on 6 October 2009, INSEE informed the Statistical Confidentiality Committee of the conditions set for its agreement[7] to the communication of individual data on households, directly or indirectly nominative:

- For the Ministerial Statistical Offices (MSO), agreement will be given on condition that the MSO makes a commitment that measures will be in place within their Ministry to guarantee that access to these files will be controlled and to ensure that they are used for exclusively statistical purposes.

- For other applicants, access to indirectly nominative data is through the Secure Remote Access Centre (CASD). Via highly secure dedicated infrastructure, the CASD uses Internet to make files of individual data available but prevents any data extraction or modification. Access to these stations is only possible for people who have been granted personalised authorisation. They must all sign a confidentiality agreement, which obliges them to comply with security and

---

[7] Agreement required for the administration of the archives to be able to issue authorisation for communication (Article L.213-3 of the Heritage Code).

Updated: March 2025

confidentiality rules and exposes them to severe penalties in the event of non-compliance.

- The secure access centre also provides researchers with access not only to household survey data but also to business survey data, and information from indirectly nominative administrative sources, transferred to INSEE or an MSO under Article 7 bis of the no. 51-711 Act of 7 June 1951, in accordance with the conditions set out in its Article 7 ter.

- In principle, access to directly identifying data is examined on a case-by-case basis by the INSEE Executive Committee. Access is normally only granted for requests to draw samples for surveys that has been approved by the Minister in accordance with Article 2 of the 1951 Act.

B. Information drawn from administrative sources

Information transmitted to INSEE or to the Ministerial Statistical Offices for the purpose of producing statistics outside the scope of the statistical surveys, is also covered by statistical confidentiality. This follows from the provisions of Article 7 bis of the no. 51-711 Act of 7 June 1951 on legal obligation, coordination and confidentiality in the field of statistics.

The rules of confidentiality that apply to the dissemination of results derived from using these files may differ from one source to another. Some data providers define rules that go beyond the requirements of statistical confidentiality (see above). For instance, for the dissemination of tables based on tax information, the rule states that no cell should refer to fewer than eleven individuals. For tables based on annual declarations of social data (DADS), no cell should refer to fewer than five individuals.

It is therefore necessary, before anything is published, to find out about the dissemination rules set by the body that has provided the information. In general, these rules are laid down in the data transmission agreement.

Access to individual data derived from administrative sources is forbidden unless derogations have been made for statistical, or scientific or historical research purposes, according to similar procedures to those that allow access to individual data from statistical surveys. The opinion of the administration or the corporation that collected the data concerned may be sought before the data is communicated.

As a result of the no. 2016-1321 Act of 7 October 2016 for a Digital Republic, Article 36, when a request is made to access a database, Article L.311-8 of the French Code of Public Relations between the public and general government now opens the way for the administration holding the database or the archives administration to request the opinion of the confidentiality committee. The applicant agrees to respect the confidentiality of the data communicated and the rules of confidentiality applicable to the source. The confidentiality committee may recommend that a remote or on-site secure access system be put in place.

Updated: March 2025

C. Mixed sources

Mixed sources mean:

- Sources derived from combinations of statistical surveys and administrative data;

- Sources containing both information of an economic and financial nature (businesses) and information relating to facts and behaviour of a private nature (households)

1. Files combining statistical data and administrative data

The rules to be taken into consideration are obtained by combining the rules applicable to each of the two types of sources.

These sources include the Tax and Social Incomes survey, which matches up results from the Employment survey, tax data and data supplied by the Family Benefits Fund or the ESANE business statistics system (*Élaboration des statistiques annuelles d'entreprises*), which combines statistical surveys and tax data.

It should be noted that the dissemination of individual tax data is forbidden according to the handbook of tax procedures (Article L.103). There are exceptions to this principle, however. Article L.135 D of this handbook states that tax and customs officers may communicate:

- Any information to INSEE and MSO employees, for the sole purpose of producing statistics;

- Data from monitoring and recovery procedures for purposes of scientific research, according to the conditions of the 1951 Act, i.e., after a favourable opinion from the Statistical Confidentiality Committee;

- Business data, for the sole purpose of carrying out economic studies, to employees of government departments responsible for producing economic studies.

2. Files combining business data and household data

These files may come from statistical surveys (ECMOSS, Labour cost and structure of earnings annual survey) or administrative sources (DADS, Annual declarations of social data).

The conditions under which data can be accessed combine the conditions already mentioned, for business data on the one hand, and for household data on the other.

Appendix

Why does statistical confidentiality also apply to aggregate statistical results and not only to the individual data from which these results are derived?

The main purpose of statistical confidentiality is to safeguard the confidentiality of individual information relating to natural or legal persons. Article 6 of the 1951 Act thus prohibits the communication of "individual information contained in the questionnaires" of statistical surveys[8]. Reading the text literally, one might think that statistical confidentiality is only enforceable in the case of communication or dissemination of individual data (whatever their origin, data from surveys or databases).

This interpretation, however, ignores the fact that the statistical results, which do not constitute individual data, may indirectly, by cross-checking with other information otherwise known to the public, disclose the characteristics of individuals.

A statistic is a numerical value or measurement that is used to summarise, analyse or interpret data in a population. A statistic can take many forms: sum, average, percentage, growth rate, population class, etc. It is well known that statistics, as such and by themselves, hide the individual information from which they are derived. Once the average income of the inhabitants of a given territory has been calculated, it is no longer possible to find, on the basis of this information alone, the income of each of the inhabitants of this territory. A statistic, by itself, is anonymous.

However, this anonymity may no longer be verified if one takes into account the information that may otherwise be accessible or available about the link between individuals and a statistic.

Consider the following three examples:

- The total turnover of companies in a given industry in a given territory:

  Through the public register of companies and establishments (Sirene), it is possible to find out which companies are concerned. If only two companies make up the population described by this statistic, each company, from the knowledge of its own turnover, is able to deduce that of its competitor.

- The annual growth rate of total turnover for companies in a given industry in a given territory:

  Still using Sirene, which disseminates the number of employees, it is possible to estimate the contribution of each company involved in the calculation of this statistic. If, on the basis of its number of employees, it can be judged that a company predominates over its competitors, it will be possible to deduce, with a good degree of certainty, that the change measured for total turnover represents that of that company alone.

---

[8] Except by decision of the archive administration, taken after consulting the Statistical Confidentiality Committee.

Updated: March 2025

- The labour market situation ("unemployed/employed/inactive") according to age group in a given territory:

  If all people between 20 and 24 years of age appear to be unemployed, and if someone, because they know them personally or from information made public on social networks, knows that someone resides in this territory and belongs to this age group, it will be possible to deduce that this person is unemployed.

These three examples show that the risk of re-identification of individual data is all the greater when the population concerned is small, which is why, for a long time, the official statistical service has defined "statistical confidentiality rules" according to the size of the population involved. In the case of business statistics, the two main rules that are applied - no dissemination of a statistic if it concerns less than three units or if one of the units contributes more than 85% of its value - are the result of decisions taken as early as the 1960s. In the case of household statistics, INSEE introduced dissemination rules in the 1980s, for the purposes of disseminating census results, which were defined according to the population of the territories concerned, and which were drawn up after consultation of the CNIL.

Thus, contrary to the position expressed by the Cada (Commission for access to administrative documents) in [its opinion no. 20191797 of 16 January 2020](#), aggregate statistical results may be subject to an obligation of statistical confidentiality. This analysis is corroborated by the Council of State in its decision no. 472883 of 31 May 2024: "*the online publication of statistical data derived from individual information protected by statistical confidentiality pursuant to article 6 of the law of 7 June 1951 is only possible, before the expiry of the deadlines mentioned by this law and **whatever the level of aggregation of these data, provided that the natural or legal persons from whom the individual information has been collected, who constitute the statistical observation units of the survey in question, cannot be identified, directly or indirectly**, taking into account all the means which could reasonably be used by a third party having access to the data thus disseminated.*"[9]

---

[9] See also the conclusions of the public rapporteur (CE no. 186073 - 7 October 1998): '*it is essential to note [...] that the concept of statistical data and that of nominative data are not necessarily mutually exclusive. **Statistical data presented in a general and impersonal form may become nominative data if it is reduced to too small a group of persons** under conditions which allow [...] 'in whatever form, directly or indirectly, the identification of the natural persons to whom it (...) applies (...)*'.

Updated: March 2025

Definitions

Individual information

Individual information is any information which makes it possible to identify, either directly or indirectly, the unit to which the information relates (see below). This definition applies to all statistical units, whether they are corporations or natural persons.

Personal data

These terms are defined in Article 4 of the General Data Protection Regulation (Regulation 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data). "Personal data" means any information relating to a natural person who is identified or identifiable. An "identifiable natural person" is said to be a natural person who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online username, or one or more factors specific to his physical, physiological, genetic, mental, economic, cultural or social identity.

Direct or indirect identification

These terms are defined in Article 3 of Regulation no. 223/2009 (amended):
- Direct identification: identification of a statistical unit from its name or address, or from a publicly available identification number;
- Indirect identification: identification of a statistical unit by any means other than direct identification.

Sensitive data

According to Article 9 of the General Data Protection Regulation, the term "sensitive data" refers to:
- Data of a personal nature which reveal racial or ethnic origins, political opinions, religious or philosophical beliefs or trade union membership;
- Genetic or biometric data for the purpose of unique identification;
- Data about a natural person's health, sex life or sexual orientation.

In principle, it is forbidden to process such data. Derogations are nevertheless possible, especially when such processing is needed for archival purposes in the public interest, for scientific or historic research or for statistical purposes.

Personal data relating to criminal convictions, offences or related security measures are also subject to specific provisions, defined in Article 46 of the Data Protection Act.

Updated: March 2025