

GUIDE TO STATISTICAL CONFIDENTIALITY

Statistical confidentiality is a particular form of **professional confidentiality** that applies to bodies that form part of the [official statistical system](#).

Statistical confidentiality gives people who provide information used to produce statistics the assurance that this information will not be used in a way that might harm them.

Statistical confidentiality prohibits any depositary of information:

- from communicating to any third parties individual information collected through a statistical survey;
- from divulging any information that has been transmitted to them by third parties for the sole purpose of compiling statistics.

Statistical confidentiality thereby guarantees that:

- the **confidentiality of the private, personal and family life** of natural persons will be respected,;
- the **trade secrets and commercial confidentiality** of businesses will be respected.

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

At national level, [Law no. 51-711 of 7 June 1951](#) (amended) on legal obligation, coordination and confidentiality in the field of statistics defines statistical confidentiality, its limits and conditions of application. The Law deals specifically with *legal obligation and statistical confidentiality*. The official statistical system carries out a number of surveys which are not compulsory, but nevertheless the rules of statistical confidentiality apply to these surveys as well as to compulsory surveys.

At European level, the confidentiality of statistical information is set out 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.

Contents

Index

- Practical rules for ensuring compliance with statistical confidentiality 3
 - A. Information drawn from statistical surveys 3
 - 1. Aggregated data tables 3
 - 2. Individual data files 3
 - B. Information drawn from administrative sources 6
 - C. Mixed sources 6
 - 1. Files combining statistical data and administrative data 6
 - 2. Files combining business data and household data 7
- Definitions 8

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 [Law no. 51-711 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 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 aggregated data on **natural persons**, statistical confidentiality requires that it is not possible to obtain information or deduce information about individuals. For example, if a table provides a breakdown for a municipality by age and marital status, and everyone aged 50 to 59 has the marital status “divorced”, then statistical confidentiality is not respected in this table. If it is known that a person is aged between 50 and 59, then the table informs us that this person is divorced.

Specific dissemination rules have been introduced 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 businesses** while still 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 law created a [Statistical Confidentiality Committee](#), which can allow derogations for accessing individual company data.

b) **Household surveys**¹ 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 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 are also anonymised, but nevertheless 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. Thus the 1931 population census returns held by archive departments are freely accessible (except for people who were minors at that time).

Non-anonymised individual data obtained from household surveys can be accessed before these time periods have elapsed after receiving a favourable opinion from the [Statistical Confidentiality Committee](#) and based on a decision from the archives administration. The procedure is the same as for access to individual information on

¹ With the exception of censuses and the annual census surveys, which are subject to special provisions.

businesses (see above). In the case of information concerning facts and behaviour of a private nature, derogation from the statistical confidentiality rules can only be granted for the purpose of official statistics or scientific or historic research.

At its meeting on 6 October 2009, INSEE informed the Statistical Confidentiality Committee of the conditions set for its agreement² 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 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 7a of [Law no. 51-711 of 7 June 1951](#), in accordance with the conditions set out in its Article 7 ter.
- Access to data which is directly nominative is examined on a case-by-case basis by the INSEE Executive Committee. Access is normally only granted for requests to constitute a sample, in order to conduct a survey that has been approved by the Minister in accordance with Article 2 of the Law of 1951.

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 is explicitly provided for in Article 7a of [Law no. 51-711 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

² Agreement required by the archives administration to issue the communication authorisation (Article L.213-3 of the French Heritage Code).

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 the purposes of research or for carrying out studies, 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 [Law no. 2016-1321 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.

C. Mixed sources

Mixed sources means:

- 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 source.

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 Law, 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).

A request for derogation from statistical confidentiality is made to the committee in a single application. 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.

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.