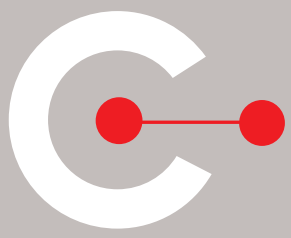
CONFI-DENTIA-LITY RULES

CONFIDENTIALITY RULES



CASD

TABLE OF CONTENTS

1 DOCUMENT DESCRIPTION

When you are working on the data you were authorised to access, you may want to export files outside of your CASD secure work environment. This process is called an "Output request". You will find a detailed description of the different procedures to undertake an output in our user guide available on our website (https://www.casd.eu/en/useful-documents/).

In the present document, you will find all the anonymization principles and confidentiality rules which must be applied to the files you want to export. The purpose of this document is to help you comply with the various confidentiality regulations in force (statistical, tax, etc.) and the different obligations relating to the legal framework of data access (GDPR, the French Data Protection Act, etc.), as well as to make sure that no published information can lead to the direct or indirect identification of either a physical person, a household or a company.

If your export does not respect the confidentiality regulations in force applying to the data, CASD will inform you of this non-compliance and will help you to either mask the values in question or aggregate the variables in a different way in order to overcome the problem (wider geographical zone, age range instead of age, etc.). **However, CASD cannot, in any way whatsoever, modify the content of the files you have requested to export.**

2 ANONYMIZATION METHODS

2.1 DATA ANONYMIZATION

In order to respect the statistical confidentiality of the results you wish to export, they must be anonymized.

The CNIL gives the following definition: "Anonymization is a process using a set of techniques which, in practice, irreversibly renders any individual identification, using any means, impossible".[1]

The CNIL recommends generalization as a method to anonymize data. This entails transforming data to make them refer to a set of persons instead of a single person.

2.2 ANONYMIZATION REQUIREMENTS

According to the CNIL, for data to be anonymized, they should meet three main requirements which are outlined below.

2.2.1 Non-individualization

Anonymized data must not allow for the individualization of published information. In other words, it must not be possible to isolate an individual in a dataset.

Example:

A database of natural persons in which surnames and first names have been replaced by a unique number is not an anonymized database. It is pseudonymized.

[1] https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles

2.2.2 Non-correlation

Anonymized data must not allow correlation between several pieces of published information. This means that it must not be possible to link multiple datasets together regarding the same individual.

Example:

One table containing the home addresses of individuals and another table containing these same addresses combined with other information about the same individuals can be linked and are therefore not anonymized.

2.2.3 Non-inference

Anonymized data must not allow information about an individual to be inferred. In other words, it must not be possible to deduce new information about an individual from the data.

Example:

A table showing the tax rates of people who responded to a questionnaire in which all men between the ages of 20 and 25 are not taxpayers is not anonymized. If we know a 24-year-old man who answered this questionnaire, we can deduce with certainty that he is not a taxpayer.

2.3 ANONYMIZATION ISSUES

The anonymization of data raises various issues. In the following sections, three such issues will be outlined: thresholds, data diversification and high contributions.

2.3.1 Threshold

Minimum thresholds have been defined for certain data sources in order to limit the risk of re-identification of individuals or companies.

Example:

The table below, which gives the distribution of companies by categories of size and their budget in R&D, contains low values.

Company Category	Number of companies	Amount R&D (K€)
Micro-enterprises	7	300
SME	2	800
Big enterprises	10	5200

This table cannot be published. Two techniques are possible in order to render it compliant with confidentiality rules: aggregation or elimination of problematic values.

2.3.1.1 Aggregation

Aggregation consists of modifying the categories of a variable to a larger scale. This technique is recommended by the CNIL because it directly limits any risk of re-identification.

Example:

If we take the previous example on the distribution of companies by categories of size and their budget in R&D, aggregation consists of using larger categories in order to be able to avoid having categories with only a small number of companies.

Company Category	Number of companies	Amount R&D (K€)
Micro-enterprises & SME	9	1100
Big enterprises	10	5200

2.3.1.2 Elimination of problematic values

The second anonymization technique is to eliminate problematic values by masking them.

Example:

We can reconsider the table on the distribution of companies by categories of size. The number of SMEs is low.

Company Category	Number of companies	Amount R&D (K€)
Micro-enterprises	7	300
SME	2	800
Big enterprises	10	5200

In order to export this in a way that complies with anonymization requirements, we can replace the problematic value by a letter, "S" for example, as in the table below.

Enterprise Category	Number of enterprises	Amount R&D (K€)
Micro-enterprises	7	300
SME	**S**	**S**
Big enterprises	10	5200

Be mindful of secondary confidentiality!

This second technique presents the risk of re-identification as other available data may allow the masked value to be deduced. Consequently, we must observe what is called **secondary confidentiality**, which aims to prevent the reconstitution of masked cells. In the example on the distribution of companies per size categories, if we found the total number of companies included elsewhere, the masked value would be easily retrieved. Therefore, in practice, to be able to export this table we must mask two values to ensure secondary confidentiality is observed.

2.3.2 Data Diversification

The purpose of data diversification is to achieve a distribution of group characteristics which is sufficiently diverse to reduce the risk of certain or near-certain deductions.

Example:

The table below presents fictitious data of diagnoses made during hospital stays for a given month and a given department.

Diagnoses made during hospital stays for a given month and a given department						
Age group	Number of patients	Hypertension	Diabetes	Asthma	Cancer	Respiratory deficiency
20-29	13	0	4	13	0	0
30-39	36	6	10	9	5	7
40-49	52	15	9	11	16	8
50-59	49	14	11	6	10	8
60-69	53	12	9	8	11	23
70-79	58	8	31	12	6	56

This table presents two problems:

- People in the "20-29" age group are all diagnosed with asthma. Therefore, if we know from elsewhere that a person, aged between 20 and 29, was hospitalized that given month, in that given department, we can deduce with certainty that this person has asthma.
- A large majority of people in the "70-79" age group are diagnosed with respiratory deficiency. Therefore, if we know from elsewhere, that a person aged between 70 and 79, was hospitalized that given month, in that given department, we will deduce with almost certainty in that this person has suffered from respiratory deficiency.

As before, to render this table compliant with confidentiality rules, it is possible to either aggregate or eliminate the problematic values.

2.3.3 High contributions

In some cases, in order to comply with confidentiality rules, data should not contain high contributions when it comes to amount variables. In other words, one unit should not contribute more than a certain percentage of an amount variable. This avoids the possibility of the whole amount or a large part of it being attributed to this unit.

Example:

The table below presents turnovers in the construction sector for the year 2019.

Business sector	Number of companies	Turnover
Building construction	467	860 745
Civil engineering	389	1 696 872
Special construction work	804	973 610

If we also look at contributions of the maximum values, we can see that one company occupies a quasi-monopolistic position in the civil engineering sector.

Business sector	Maximum turnover	Maximum percentage
Building construction	256 804	29.83%
Civil engineering	1 531 794	90.27%
Special construction work	41 947	4.30%

Consequently, the first table of turnovers by business sector presents a high risk of almost certain deduction. As in the two previous cases, there are two possible solutions: we can either aggregate or eliminate the problematic values.

3 DESCRIPTION OF OUTPUT FILES

First of all, you should indicate which **data sources** were used for the output and precisely describe the contents of your output. You can do this in the email requesting the output or in a separate text file inserted in your output. In addition to this, and in order to be able to verify your outputs, you must include specific information according to the type of files you wish to export.

3.1 SCRIPTS

You can request an export of your scripts to be able to reuse them outside of CASD. Scripts cannot contain confidential data.

3.2 REGRESSION AND ECONOMETRIC MODELS

For the results of regressions or econometric models, indicate the **number of observations** in order to allow CASD to ensure that confidentiality rules are respected.

3.3 MAPS AND GRAPHICS

Maps can concern very small populations (communal or infra-communal). In order to verify that the confidentiality rules are respected, you should provide **the population that was used to generate this map**.

For graphics, such as curves, histograms, scatterplots, diagrams, etc., you should provide **the population and the meaning of the variables used**.

Other types of graphics are more complicated to review because they can contain individual information. For example, boxplots can not only contain maximum and minimum values, but also extreme points (outliers) which should not be identified.

You should also pay particular attention to factor analysis graphics representing individuals. These can contain atypical individuals which can be identified (for example a PCA on companies where the SIRET is used in the graphic as an identifier of each point).

Pay particular attention to Stata graphics in LIVE format!

Stata Graphics in LIVE Format contain the datasets from which they are generated. It is preferable to convert these datasets into another format (AS-IS: a format which does not contain the datasets, pdf, jpg, etc.).

3.4 AGGREGATED DATA TABLES

For aggregated data tables, in the description of the output you should provide the **meaning** of all variables used, **the number of observations** in each cell, and for amount variables, the information on the **maximal contribution** in the cell in a **separate non-anonymized control file**.

Control file:

In order to verify that results regarding amount variables do not contain high contributions, you should include a control file in your export containing information on the maximum values and the percentage this maximum represents for amount variables.

The control file, **therefore containing confidential data on a single unit, will be deleted from the export before being sent to you. It is thus necessary to create an additional file for the control file which is separate from the other files you wish to export.**

Example:

The following table is based on the previous example of turnovers in the construction sector for the year 2019. It is a table containing control information: the maximum turnover and the percentage this maximum represents in each business sector.

Business sector	Number	Total turnover	Maximum turnover	Maximum percentage
Building construction	467	860 745	256 804	29.83%
Civil engineering	389	1 696 872	1 531 794	90.27%
Specialized construction work	804	973 610	41 947	4.30%

The table shows that one company has a quasi-monopolistic position in the civil engineering sector. Thus, the corresponding amount variable, in this case the turnover in the civil engineering sector, presents a high risk of almost certain deduction. Therefore, the problematic value must be either aggregated or eliminated. The second technique is used in the following table which is destined to be exported.

Business sector	Number	Total turnover
Building construction	467	860 745
Civil engineering	389	**S**
Specialized construction work	804	973 610

3.5 FINALISED ARTICLES

The finalised articles that you wish to export outside your CASD work environment cannot contain data concerning a small population of individuals.

4 GENERAL RULES

Confidentiality rules have been defined in order to respect the confidentiality regulations. There are general rules which apply to all sources which are discussed first below.

4.1 "HOUSEHOLD" DATA

Statistical confidentiality requirements specify that information may be published only if it has been processed in such a way as to make it impossible to identify natural persons. This obligation is also applicable to individual entrepreneurs.

In practice, we consider that statistical confidentiality is respected if the knowledge of one characteristic of an individual cannot lead to the knowledge of another characteristic with which it is crossed in a table.

Example:

The following table represents age distribution according to matrimonial situation. It indicates that persons aged between 50 and 59 years all share the same matrimonial status: "divorced". Statistical confidentiality is therefore not respected in this table, thus making it unpublishable. Indeed, if we know otherwise that a given individual is in the age range of 50-59 years, the table also informs us that this individual is divorced, even though the cell crossing the levels "50 to 59 years" and "divorced" contains multiple individuals.

Matrimonial situation and age range	18-25 years	26-49 years	50-59 years	60 years and plus +
Married	7	27	0	30
Divorced	0	11	9	22
Other	21	12	0	4

4.2 "COMPANY" DATA

For tables containing aggregated company data, the rules are as follows:

- No cell should contain less than **3 units** (decision dated 13 June 1980 by the general director of INSEE)
- No cell should contain information for which a single unit contributes more than **85% of the total** (ruling by the CNIS dated 7 July 1960)

The rules in force are applied to the SIREN, which pertains to units with a legal personality, and not to the SIRET, which pertains to establishments. They also concern individual entrepreneurs. If you wish to export data on establishments, you should provide us with the corresponding company registration number.

Note: No export of SIREN/SIRET is authorized.

4.3 "AGRICULTURAL" DATA

For tables containing aggregated agricultural data, the rules are as follows:

- No cell should contain less than **3 units**
- No cell should contain information for which a single unit contributes more than **85% of the total**

These rules are applied to farms, not to plots.

4.4 MIXED SOURCES

Mixed sources result either from combinations (merges) between statistical surveys and administrative data or sources containing financial and economic information (of companies) and private facts and behavioural information (households) combined.

When working on this type of source, the principle is simple: the rules to take into consideration are the accumulation of the rules applied to statistical surveys and those applied to administrative data. For example, the FARE (File approaching the results of the Elaboration of annual statistics of companies) data source contains both company data and data from tax returns. The cumulative rules apply to every variable of the mixed source.

In general terms, pay attention to the sources combined with fiscal data (part 5.2 list – DGFIP data)

4.5 SURVEY DATA

In the case of sample surveys, units are subject to one or several weighting to achieve representativeness of the studied population. **The rules are applied to weighted data**.

5 DETAILED RULES FOR EACH SOURCE

In addition to the general rules, there are specific rules for each source, which are detailed in the following. These rules are defined by the data producers according to the confidentiality level they consider as needed for their data.

5.1 INSEE DATA

Confidentiality rules related to INSEE sources are described in the statistical confidentiality guidelines issued by INSEE.

5.1.1 DADS (Annual declaration of social data) / BTS (All employees databases)

No published table can contain any information leading to the direct or indirect identification of either an employee or a company.

- Tables about place of residence ("household" vision):
  - No cell can contain less than 5 employees
  - No cell can contain a sole employee contributing for more than 80% of the workforce
- Tables about place of work ("company" vision), **in addition to the two previous rules,** the following rules concerning company data apply:
  - No cell can contain less than 3 companies or establishments
  - No cell can contain a sole company or establishment contributing for more than 85% of the total

5.1.2 CLAP (Local knowledge of the production system) and FLORES (localized files on wages and employment)

CLAP and FLORES data belong to "company" data. Indicators subject to statistical confidentiality are salaries and remuneration. The rules to apply are the following:

- No cell can contain less than 3 units (a unit is either a company or an establishment)
- No cell can contain a sole unit contributing for more than 85% of the total
- No cell can contain less than 5 employees

5.1.3 FARE (File approaching the results of the Elaboration of annual statistics of companies) / FICUS (Unified Corporate Statistical Collection System)

FARE and FICUS data are mixed sources, they are both "statistical and tax" data at the same time. The confidentiality rules to apply in this case are the sum of rules applied to statistical surveys and tax data.

- No cell can contain less than 3 units
- No cell can contain 1 unit contributing for more than 85% of the total
- As for sole proprietorship, no cell can contain less than 11 units

5.1.4 SINE (Information system for new companies)

SINE is also a "company" source. The confidentiality rules to apply are the following:

- No result can pertain to less than 3 units per cell
- No data in which a sole company contributes more than 85% of the total value

Furthermore, rates of survival should not be calculated on a population containing less than 20 companies. This minimum threshold of 20 companies is also required for zoning and particular regroupings.

5.1.5 The population census

The rules of publishing data drawn from the population census have evolved with the Bylaw of 19 July 2007 relating to the publication of population census results. This bylaw replaced that of 22 May 1998, modified on 8 April 2002, and which also related to the publication of population census results.

➔ Because of varying sampling rates between different editions of the Population Census, the following information summarizes the thresholds to respect according to the census year and the processing type.

- Until the 1999 Population Census:
  - Main processed census: at least 4 units per cell after weighting
  - Complementary processed census:

Census year	Sampling rate	Minimal thresholds to be published (after weighting)
1962	1/20	80 units
1968	1/4	16 units
1975	1/5	20 units
1982	1/4	16 units
1990	1/4	16 units
1999	1/4	16 units

- For the modernized Census (after 2006): at least 10 units per cell after weighting

The data in the RP Saphir are based on the complementary processed census, so the rules for dissemination are those of the complementary processed census.

➔ Concerning "sensitive" variables, specific geographical thresholds should be respected. Sensitive variables are the following: current nationality (or nationality at birth), place of birth, former place of residency, year of arrival (or duration) in France. Notions of immigrants and French Citizenship by acquisition are included in the range of sensitive variables up to the census of 1999.

The geographical thresholds, after weighting, for these variables are as follows:

Up to 1999	From 2006 (annual census)
Municipalities with more than 5 000 residents	Municipalities with more than 5 000 residents
Threshold of 10 000 residents for the arrondissements, employment zones, urban airs, urban units (or their regroupings) and zones of urban public policies or their regroupings	Threshold of 5 000 residents for the arrondissements, employment zones, urban airs, urban units and zones of urban policies
Infra-communal zones resulting from the combination of 3 neighbourhoods (a fixed zone resulting from the division of the municipality into geographical zones containing around 2 000 residents)	Infra-communal zones resulting from the combination of 3 neighbourhoods (a fixed zone resulting from the division of the municipality into geographical zones containing around 2 000 residents)
Department in the year (or duration) of arrival	Department in the year (or duration) of arrival

For output concerning sensitive variables of the Census, and for the purposes of verification, you should provide us with the geographical scale on which your data are based. In the case of data aggregated on a municipality scale, you should provide us with the number of inhabitants for each municipality or group of municipalities for all variables concerned.

5.1.6 SIASP (System for Information on Civil Servants) and FGE (File on central civil servants)

Publication of statistical results based on SIASP data or FGE data should conform to the rules in force pertaining to statistical confidentiality and texts on the protection of individual data. In particular, no table meant to be published should lead to the direct or the indirect identification of either an employee or an establishment.

Tables about place of residence:

- No cell can contain less than 5 employees
- No cell can contain a sole employee contributing to more than 80% of the workforce

Tables about place of work, **in addition to the two previous rules,** the following rules must be applied:

- No cell can contain less than 3 establishments
- No cell can contain a sole establishment contributing to more than 85% of the total

Exception: for data relating to the State civil service, on a national, regional or departmental level, it is not always possible to find three establishments with a Siret number. In consequence, this rule need not be applied only in this context.

5.1.7 FIDELI (Demographic files on households and individuals)

The rules for FIDELI are the following:

- For municipalities with 5 000 residents or more, the minimal geographical scale for producing tables is the municipality or the IRIS.
- For municipalities with less than 5 000 residents, the minimal geographical scale for producing tables is the *Établissement Public de Coopération Intercommunale* (EPCI: the public institution of intercommunal cooperation). EPCI where the total population of municipalities of less than 5000 residents is less than 2000 residents, cannot lead to the production of tables.
- For results concerning Urban policy priority neighbourhoods (QPV), the minimal geographical scale for producing results is the region.
- For all published results, no cell can contain less than 11 individuals.

Note: For output concerning a particular population, the rules defined above must be applied to that population. For example, in order to publish results on a population of secondary residencies, it is necessary to ensure that there are more than 5,000 **inhabitants of secondary residencies** in the municipality concerned.

For output concerning FIDELI Data, and for the purposes of verification, you should provide us with the geographical scale on which your data are based. In the case of the data being aggregated on a municipality or EPCI scale, you should provide us with the number of inhabitants in each municipality for all variables concerned.

5.2 DGFiP DATA – MINISTRY OF FINANCE

For company data:

- No cell in the table should pertain to less than three units (decision of the 13 June 1980 of the General Director of the Insee).
- No cell in the table should contain information where a single company represents more than 85% of the total (CNIS rule, 7 July 1960).

For household data: no cell should contain less than 11 individuals.

For fiscal data (rules endorsed by the CNIL in a notice of the 27 May 1997 and set out in §30 of the BOI-DJC-CADA-20):

- Units number rule: an aggregated data is not communicated when it concerns less than 11 units;
- Units weight rule: an aggregated data is not communicated when including a dominant element that represents more than 85% of the aggregated amount.

Sources combining survey results with fiscal information are subjects to these rules. More precisely, this includes:

- Statistics on Income and Living Conditions Survey (SRCV) since 2008
- Household Wealth Survey
- Life History and Wealth Survey (EHVP)
- Household Budget Survey (BDF)
- Training and Vocational Skills Survey (FQP) 2014
- Housing Survey (ENL) 2006 and 2013
- Inter-Regime Sample of Retirees (EIR)
- Individual reporting of Handicap Compensation Benefit applicants and beneficiaries (RI-PCH)
- Survey on the capacities, aids and resources of elderly people – institutions section (CARE-I)
- Survey on the capacities, aids and resources of elderly people – household section (CARE-M)

- Survey on the performance of homes, equipment, energy needs and uses - homes, occupants and energy expenses characteristics (PHEBUS CLODE)
- Survey on the performance of homes, equipment, energy needs and uses - energy performance diagnosis (PHEBUS DPE)
- Beneficiaries of Social Minima Survey (BMS)
- Unique Integration Contract (CUI)

5.3 DARES DATA – MINISTRY OF LABOUR

The rules for DARES's data are the following:

- For individual/household data: no cell can contain less than 5 individuals
- For company data: no cell can contain less than 5 individuals and no cell can contain a single enterprise contributing for more than 85% of the total

5.4 DEPP DATA – MINISTRY OF EDUCATION

The rules for the DEPP's data are the following:

- For individual data: no cell can contain less than 10 individuals
- For aggregated data at the establishment level (secondary school, high school, etc.): no cell can contain less than 10 establishments

5.5 SSP DATA – MINISTRY OF AGRICULTURE

For tables containing aggregated agriculture data, the rules are as follows:

- No cell should contain less than **3 units**
- No cell should contain data in which a sole farm is contributing to more than **85% of the total**
- No cell should contain less than **5 employees** (for wage-earning data)

Rules to follow should be applied to farms, which means to units with a legal personality and not to plots. If you wish to export data on plots, you should provide us with the corresponding farms number.

Some data from the agricultural census are exempt from the above-mentioned rules of statistical confidentiality. For more information, see the decision of the Comité du Secret Statistique on the subject.

5.5.1 PKGC (Filed crop cultural practices)

- When the total of a table is less than 30 plots, it cannot be disseminated
- No cell should contain less than 3 aggregated parcels

5.5.2 Ecopeche (Surveys on economic data outputs in the marine fishing industry)

- No information can pertain to less than 5 vessels

5.6 FRANCE AGRIMER DATA

The rules are as follows (for tables):

- No cell should contain less than 5 individuals/sole corporations
- For other companies: at least 3 units for each cell

5.7 DPMA DATA – MINISTRY OF AGRICULTURE

No information (box, graph item/map, etc.) can pertain to strictly less than 5 unique vessels (field CFR_COD).

5.8 ASP DATA – SINGLE PAYMENT AGENCY

5.8.1 Graphic Parcel Register

For tables containing aggregated agriculture data, the rules are as follows:

- No cell should contain less than **3 units**
- No cell should contain data in which a sole farm is contributing to more than **85% of the total**
- No cell should contain less than **5 employees** (for wage-earning data)

5.9 SDSE DATA– MINISTRY OF JUSTICE

5.9.1 Data sources regarding natural persons (CPH files, PACS and National Criminal Records Statistical File)

At least 5 statistical units for each cell disseminated or calculable by cross-checking information from disseminated data. This rule should be respected for all types of statistical units (for example, it applies to crimes from the Cassiopée data). This implies that, when not all modalities of a variable are presented, this minimum of five units also applies to the complements of the number of modalities presented out of the total number of the studied population. This also implies that no mention should be made of a unanimous distribution of a characteristic (other than the application of a legal criterion) within a sub-population.

5.9.2 Data sources regarding companies (Compulsory Liquidation data and Business Safeguarding)

Two conditions:

- At least 3 statistical units for each cell disseminated or calculable by cross-checking information from disseminated data
- In each disseminated or calculable cell, no statistical unit can contribute to more than 85% of the total presented

These two rules are also applied, when not all modalities of a variable are presented, to frequencies and contributions of complement variables to those presented.

5.10 SDES DATA – MINISTRY OF THE ENVIRONMENT

5.10.1 EPTB (Building Land Price Survey)

A minimum of 11 permits in each cell of the table. Unweighted frequency should be referenced.

5.10.2 ECLN (Survey on the marketing of new housing)

For data regarding sale or reservation price, no cell should pertain to less than 3 property developers and no property developer should contribute for more than 85% of the total value.

For data about reservations and ongoing housing, no restrictions are to be applied to geographical zones of 50 000 and more inhabitants for which at least 5 property developers have implemented a marketing program during the previous year. For geographical zones that do not respect these conditions, no cell should pertain to less than 3 property developers and no property developer should contribute for more than 85% of a total amount.

For other commercial data allowing the property developer to promote themself, i.e. the petitioner reference, program features (address, type of housing, number of housings, start and end of the commercialization quarter), no restrictions are be applied regardless of the geographical zone.

5.10.3 FILOCOM (Housing record by municipality)

In order to respect complete and localized data confidentiality, the following rules must be applied :

- For housing, owners and occupants or their occupation variables, aggregated results should concern at least 11 units.
- For amount variables, no single housing should represent more than 85% of a total aggregated amount.

5.10.4 SITADEL (Information System and Automated Processing of Elementary Data on housing and buildings)

The SITADEL file rendered available by the CASD is not covered by statistical confidentiality. It is aimed to be open data.

5.10.5 Rép-Entreprôts (Warehouses and logistic platforms directory)

- No aggregation should concern less than 3 establishments
- No aggregation should contain data for which a single establishment represents more than 85% of the total value

5.10.6 PHEBUS CLODE and DPE (Survey on the performance of homes, equipment, energy needs and uses)

Tables should respect the following rules :

- No information can concern less than 11 homes
- The smallest geographical scale permitted for producing tables is the region
- Only income variables by aggregated scales can be published (revenu_dispo_2012, ermrevtra and ermaidetr). Other income variables cannot be published (ermrev and ermaide).
- Ermerparg, elamr, elcm, elchauf, eleaucho, elchoeach, elcmpro, elth, epmdr2, efamr, efcm, efchauf, efeaucho, efchoeach, efcmpro, efth, eftfpb, eqmont and elld variables, as well as ecomcar_X, montant_XXX and reduc_XXX can be published

5.10.7 CEE (Energy saving certificates)

- An aggregated data is not communicated when it concerns less than 11 units;
- An aggregated data is not communicated when including a dominant element that represents more than 85% of the aggregated amount.

5.10.8 EMP (Individual Mobility Survey)

For data on individuals or households, it is forbidden to publish data that would allow direct or indirect identification of a person or even, without being able to identify this person, to obtain information about him/her. In practice, no cell or row in any table published should contain less than 10 individuals/households (before weighting).

5.11 ANAH DATA

The rules for the MPR data are as follows:

- An aggregated data is not communicated when it concerns less than 11 units;
- An aggregated data is not communicated when including a dominant element that represents more than 85% of the aggregated amount.

5.12 SIES DATA – MINISTRY OF RESEARCH

The rules of SIES data are as follows:

- For data on educational establishment :
    - o Following the favorable opinion of the Statistical Confidentiality Committee on March 19, 2024, the Ministry of Higher Education and Research authorizes the dissemination of total enrolment and graduation figures by gender, by program and by institution, in view of their interest for the public.
    - For other indicators, no table, by establishment, should contain data from private establishments that objected the release (or were unable to express their objection) of their headcount during the survey; these establishments can be circumscribed by the oppos= O variable.
- For student data, direct or indirect identification of individuals should be impossible. In practice, we consider that statistical confidentiality is respected when there are no cell that contains less than 5 individuals

5.13 DSED DATA – MINISTRY OF IMMIGRATION

No individual or territorial analysis will be allowed (except for typologies of territories)

Any published table must not under any circumstances allow the direct or indirect identification of an individual:

- No cell should contain less than 10 units (before weighting)
- No cell should contain a single unit contributing for more than 85% of the total analyzed

Furthermore, geographical analyses aimed at comparing geographical areas (departments, regions, municipalities, IRIS, QPV…) will not be authorized. Only analyses comparing **typologies of territories** (for example: rural versus urban…) will be authorized.

5.14 CIFRE DATA OF THE MESRI – MINISTRY OF RESEARCH

The rules for CIFRE data of the MESRI are as follows:

- In the case of individual data, it is forbidden to publish information which can lead to the direct or indirect identification of a person. In addition to this, even if a person cannot be identified, it is forbidden to publish any information about them. These rules limit the preciseness of published information available. Strict rules are defined specifically for the population census. No output can contain 10 observations concerning individuals.
- In the case of company data, no published result can pertain to less than 3 companies and no data can pertain to a single company contributing more than 80% of the total value. However, the dissemination of lists extracted from the directory of companies or establishments mentioning economic activity, a range of frequencies and range of turnovers is allowed.
- No edition of lists providing the name, address or any other individual characteristics of beneficiary students or enterprises is allowed.

5.15 DGE DATA

No cell of the table should:

- Pertain to less than three units
- Contain data for which one company represents more than 85% of the total

Statistics on hub or project can only be exported if they cannot be identified.

5.16 INJEP DATA

For aggregated tables about members of sports federations:

- No cell can pertain to less than 3 units;
- Data from the *Fédération Française Maccabi*, *Fédération Sportive de la Police Nationale* and *Fédération des Clubs de la Défense* should be aggregated with the data from another federation and should not be presented separately.

For aggregated tables from the *Enquête Nationale sur les Pratiques Physiques et Sportives*:

- The smallest geographical scale permitted for producing tables is the region;
- No cell can pertain to less than 10 units unweighted.

5.17 **CEREQ DATA**

A minimum of 5 observations in each cell of the table.

5.18 **BPIFRANCE DATA**

A minimum of 10 observations in each cell of the table.

5.19 **ANIL DATA**

The rules for ANIL data are as follows:

- No dissemination of data concerning housing
- A minimum of 50 observations in each cell of the table (cell of an aggregated table, data for model calibration).

5.20 **ODR DATA**

Information based on a geographical scale which is smaller than the canton cannot be published. A minimum of 3 observations per cell should also be respected.

For output concerning ODR Data, and for the purposes of verification, you should provide us with the geographical scale on which your data are based.

5.21 **MSA DATA**

The rules for MSA data are as follows:

- No cell can contain less than 5 units
- No cell can contain data for which a single non-employee contributor represents more than 85% of the total
- Having knowledge of an individual characteristic cannot lead to gaining knowledge of another with which it is crossed in the same table.

5.22 **ACOSS DATA**

A minimum of 10 observations in each cell of the table.

5.23 **CNAF DATA**

A minimum of 5 observations in each cell of the table.

5.24 BMO DATA – POLE EMPLOI

No cell should contain less than 60 observations (unweighted observations).

5.25 PMSI DATA – ATIH

No cell can contain information on less than 11 units regarding the number of patients, the number of hospital stays and the number of doses.

5.26 DREES DATA

The rules for DREES data are as follows:

- For individual/household data, excluding fiscal data, a minimum threshold of 5 units per cell should be observed
- For fiscal data, a minimum threshold of 11 individuals per cell should be observed
- For establishment data, a minimum threshold of 3 units per cell should be observed

For data from the survey of training schools for the health and social professions, the statistical secret only apply for surveys from 2021. Also, an exemption allows the following indicators to be disseminated below the thresholds:

- Number of enrolees: total number of enrolees by type of training, at establishment and regional level
- Number of graduates and trainees: total number of graduates and trainees by type of training, including partial VAE and reduced tuition, at establishment and regional level
- Number of places financed: total number of places financed, by type of financing organization and by type of training, at regional level

5.27 IRDES DATA

For HYGIE and ESPS, no results should concern less than 15 individuals.

5.28 BANQUE DE FRANCE DATA

Same rules as INSEE company data rules:

- No cell can contain less than 3 companies or establishments
- No cell can contain a sole company or establishment contributing for more than 85% of the total

5.29 FRENCH CUSTOMS AND INDIRECT TAXATION AUTHORITIES (DGDDI) DATA

Same rules as INSEE company data rules:

- No cell can contain less than 3 companies or establishments
- No cell can contain a sole company or establishment contributing for more than 85% of the total

5.30 INSTITUTE OF YOUTH AND POPULAR EDUCATION (INJEP-MEDES)

Regarding ENPPS data, the files are representatives at the regional level but not at a sub-regional level. More detailed geographical data are included in the individuals details file (X and Y coordinates, Insee code commune and département of residence) but it cannot be used to produce representative results at commune or département level.

5.31 CONSTANCES AND GAZEL COHORTS DATA (INSERM)

No cell can contain information on less than 11 units regarding the number of patients, the number of hospital stays and the number of doses.


C/\SD