

# RÈGLES

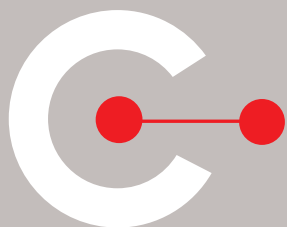
# DE

RÈGLES DE CONFIDENTIALITÉ

# CONFI-

# DENTIA-

# LITE



CASD

12/2021

# TABLE DES MATIERES

---

<b>1</b>	<b>DESCRIPTION DU DOCUMENT</b>	<b>3</b>
<b>2</b>	<b>METHODES D'ANONYMISATION</b>	<b>3</b>
2.1	L'ANONYMISATION DES DONNEES	3
2.2	LES CRITERES DE DONNEES ANONYMISEES	3
2.2.1	LA NON-INDIVIDUALISATION	3
2.2.2	LA NON-CORRELATION	4
2.2.3	LA NON-INFERENCE	4
2.3	LES PROBLEMATIQUES D'ANONYMISATION	4
2.3.1	LES SEUILS	4
2.3.2	LA DIVERSIFICATION DES DONNEES	5
2.3.3	LES FORTES CONTRIBUTIONS	6
<b>3</b>	<b>DESCRIPTION DES FICHIERS POUR LES SORTIES</b>	<b>7</b>
3.1	LES PROGRAMMES	7
3.2	LES REGRESSIONS, MODELES ECONOMETRIQUES	7
3.3	LES CARTES ET LES GRAPHIQUES	7
3.4	LES TABLEAUX DE DONNEES AGREGES	8
3.5	LES ARTICLES « FINALISES »	9
<b>4</b>	<b>REGLES GENERALES</b>	<b>9</b>
4.1	LES DONNEES « MENAGES »	9
4.2	LES DONNEES « ENTREPRISES »	10
4.3	LES DONNEES « AGRICOLES »	10
4.4	LES DONNEES « MIXTES »	10
4.5	LES DONNEES ISSUES D'ENQUETES	10
<b>5</b>	<b>REGLES DETAILLEES POUR CHAQUE SOURCE</b>	<b>11</b>
5.1	DONNEES DE L'INSEE	11
5.1.1	DADS (DECLARATION ANNUELLE DE DONNEES SOCIALES)	11
5.1.2	CLAP (CONNAISSANCE LOCALE DE L'APPAREIL PRODUCTIF) ET FLORES (FICHIER LOCALISE DES REMUNERATIONS ET DE L'EMPLOI SALARIE)	11
5.1.3	FARE (FICHIER APPROCHE DES RESULTATS D'ESANE) / FICUS (FICHIER DE COMPTABILITE UNIFIE DANS SUSE)	11
5.1.4	SINE (SYSTEME D'INFORMATION SUR LES NOUVELLES ENTREPRISES)	11
5.1.5	RECENSEMENT DE LA POPULATION	12
5.1.6	SIASP (SYSTEME D'INFORMATION SUR LES AGENTS DES SERVICES PUBLICS) ET FGE (FICHIER GENERAL DE L'ETAT)	13
5.1.7	FIDELI (FICHIERS DÉMOGRAPHIQUES SUR LES LOGEMENTS ET LES INDIVIDUS)	13
5.2	DONNEES DE LA DGFIP – MINISTERE DES FINANCES	13
5.3	DONNEES DE LA DARES – MINISTERE DU TRAVAIL	14
5.4	DONNEES DE LA DEPP – MINISTERE DE L'ÉDUCATION	14
5.5	DONNEES DU SSP – MINISTERE DE L'AGRICULTURE	14

5.5.1	PKGC (PRATIQUES CULTURALES SUR LES GRANDES CULTURES)	15
5.5.2	ECOPECHE (ENQUETE SUR LA PRODUCTION DES DONNEES ECONOMIQUES DANS LE SECTEUR DES PECHES MARITIMES)	15
<b>5.6</b>	<b>DONNEES DE FRANCE AGRIMER</b>	<b>15</b>
<b>5.7</b>	<b>DONNEES DE LA DPMA – MINISTERE DE L’AGRICULTURE</b>	<b>15</b>
<b>5.8</b>	<b>DONNEES DE LA SDSE – MINISTERE DE LA JUSTICE</b>	<b>15</b>
5.8.1	FICHIERS CPH, PACS ET FICHER STATISTIQUE DU CASIER JUDICIAIRE NATIONAL	15
5.8.2	FICHER STATISTIQUE DES TRIBUNAUX DE COMMERCE	15
<b>5.9</b>	<b>DONNEES DU SDES – MINISTERE DE L’ENVIRONNEMENT</b>	<b>16</b>
5.9.1	EPTB (ENQUETE SUR LES PRIX DES TERRAINS A BATIR)	16
5.9.2	ECLN (ENQUETE SUR LA COMMERCIALISATION DES LOGEMENTS NEUFS)	16
5.9.3	FILOCOM (FICHER DES LOGEMENTS PAR COMMUNES)	16
5.9.4	SITADEL (SYSTEME D’INFORMATION ET DE TRAITEMENT AUTOMATISE DES DONNEES ÉLEMENTAIRES SUR LES LOGEMENTS ET LES LOCAUX)	16
5.9.5	REP-ENTREPOTS (REPERTOIRE DES ENTREPOTS ET DES PLATEFORMES LOGISTIQUES)	16
<b>5.10</b>	<b>DONNEES DU SIES – MINISTERE DE LA RECHERCHE</b>	<b>16</b>
<b>5.11</b>	<b>DONNEES DU DSED – MINISTERE DE L’IMMIGRATION</b>	<b>17</b>
<b>5.12</b>	<b>DONNEES CIFRE DU MESRI – MINISTERE DE LA RECHERCHE</b>	<b>17</b>
<b>5.13</b>	<b>DONNEES DE LA DGE</b>	<b>17</b>
<b>5.14</b>	<b>DONNEES DE L’INJEP</b>	<b>17</b>
<b>5.15</b>	<b>DONNEES DU CEREQ</b>	<b>17</b>
<b>5.16</b>	<b>DONNEES DE BPIFRANCE</b>	<b>18</b>
<b>5.17</b>	<b>DONNEES DE L’ANIL</b>	<b>18</b>
<b>5.18</b>	<b>DONNEES DE L’ODR</b>	<b>18</b>
<b>5.19</b>	<b>DONNEES DE LA MSA</b>	<b>18</b>
<b>5.20</b>	<b>DONNEES DE L’ACOSS</b>	<b>18</b>
<b>5.21</b>	<b>DONNEES DE LA CNAF</b>	<b>18</b>
<b>5.22</b>	<b>DONNEES BMO DE POLE EMPLOI</b>	<b>18</b>
<b>5.23</b>	<b>DONNEES DU PMSI DE L’ATIH</b>	<b>18</b>
<b>5.24</b>	<b>DONNEES DE LA DREES</b>	<b>18</b>
<b>5.25</b>	<b>DONNEES DE L’IRDES</b>	<b>19</b>

## 1 DESCRIPTION DU DOCUMENT

---

Suite aux traitements des données auxquelles vous avez été autorisé à accéder, vous serez amené à opérer des exports de fichiers depuis votre environnement de travail sécurisé CASD. Ce processus est appelé « demande de sortie ». Vous trouverez une description des différentes procédures pour effectuer une sortie dans le [guide utilisateur disponible sur le site web du CASD](#).

Vous trouverez dans ce document l'ensemble des principes d'anonymisation et des règles de confidentialité à appliquer aux résultats que vous souhaitez exporter. Il s'agit de vous aider à respecter les différents types de secret applicables (statistique, fiscal, etc.) et les différentes obligations liées aux cadres juridiques de l'accès aux données (RGPD, Loi Informatique et Liberté...) et de s'assurer qu'aucune information permettant d'identifier une personne physique, un ménage ou une entreprise ne sera divulguée.

Si votre sortie est de nature à générer une rupture du secret applicable aux données concernées, le CASD vous en informera. Il vous indiquera les cases à masquer concernées ou les variables à agréger différemment pour que le problème soit résolu (zone géographique plus grande, tranche d'âge regroupé au lieu de l'âge fin...). **Cependant, le CASD ne pourra en aucun cas modifier le contenu des fichiers que vous avez demandé de sortir.**

## 2 METHODES D'ANONYMISATION

---

### 2.1 L'ANONYMISATION DES DONNEES

Les résultats que vous exportez de l'environnement sécurisé doivent être anonymisés.

La CNIL donne la définition suivante : « L'anonymisation est un traitement qui consiste à utiliser un ensemble de techniques de manière à rendre impossible, en pratique, toute identification de la personne par quelque moyen que ce soit et de manière irréversible ». <sup>1</sup>

La CNIL recommande la méthode de **la généralisation** : c'est-à-dire transformer les données pour qu'elles ne se réfèrent plus à une seule personne mais qu'elles soient communes à un ensemble de personnes.

### 2.2 LES CRITERES DE DONNEES ANONYMISEES

Selon la CNIL, pour que des données soient anonymisées, elles doivent respecter trois critères présentés ci-dessous.

#### 2.2.1 La non-individualisation

Des données anonymisées ne doivent pas permettre l'individualisation de l'information diffusée. Autrement dit, il ne doit pas être possible d'isoler un individu dans un jeu de données.

#### **Exemple :**

Une base de données sur des personnes physiques dans laquelle les noms et les prénoms ont été remplacés par un numéro unique n'est pas une base de données anonymisée. C'est une base pseudonymisée.

---

<sup>1</sup> <https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles>

### 2.2.2 La non-corrélation

Des données anonymisées ne doivent pas permettre la corrélation entre plusieurs informations diffusées, c'est-à-dire qu'il ne doit pas être possible de relier des ensembles de données sur un même individu.

**Exemple :**

Un tableau de données cartographiques qui contient les adresses des domiciles de personnes physiques et un autre tableau qui contient ces mêmes adresses associées à d'autres informations sur ces mêmes individus peuvent être liés et par conséquent ne sont pas anonymisés.

### 2.2.3 La non-inférence

Des données anonymisées ne doivent pas permettre d'inférer une information sur un individu. Autrement dit, il ne doit pas être possible de déduire d'une façon quasi certaine de nouvelles informations sur un individu.

**Exemple :**

Un tableau qui indique le taux d'imposition de personnes ayant répondu à un questionnaire dans lequel tous les hommes ayant entre 20 et 25 ans ne sont pas imposables n'est pas un tableau anonymisé. Si on connaît un homme de 24 ans ayant répondu à ce questionnaire, on peut déduire d'une manière certaine que cet homme n'est pas imposable.

## 2.3 LES PROBLÉMATIQUES D'ANONYMISATION

L'anonymisation des données pose diverses problématiques. Dans ce qui suit, trois seront exposées : les seuils, la diversification des données et les fortes contributions.

### 2.3.1 Les seuils

Des seuils minimums d'effectifs sont définis pour certaines sources de données dans l'objectif de limiter les risques de ré-identification d'individus ou d'entreprises.

**Exemple :**

Le tableau ci-dessous, qui donne la répartition des entreprises par catégories de taille, contient des effectifs faibles.

Catégorie d'entreprise	Nombre d'entreprises
<b>TPE</b>	7
<b>PME</b>	2
<b>Grandes entreprises</b>	10

Ce tableau n'est pas diffusable. Deux techniques peuvent être utilisées pour le rendre conforme aux règles de confidentialité : l'agrégation et la suppression des valeurs posant problème.

#### 2.3.1.1 L'agrégation

L'agrégation consiste à modifier l'échelle de grandeur des catégories d'une variable pour prendre une échelle plus grande. C'est une technique recommandée par la CNIL car elle permet de limiter directement tout risque de ré-identification.

**Exemple :**

En reprenant l'exemple sur la répartition des entreprises par catégories de taille, l'agrégation consiste à prendre des catégories plus grandes pour ne plus avoir d'effectifs faibles.

Catégorie d'entreprise	Nombre d'entreprises
<b>PME &amp; TPE</b>	9
<b>Grandes entreprises</b>	10

### 2.3.1.2 Suppression de valeurs posant problème

La deuxième technique d'anonymisation est la suppression d'information.

#### **Exemple :**

Le tableau ci-dessous donne la répartition des entreprises selon leurs catégories de taille. Il contient un effectif faible pour les PME.

Catégorie d'entreprise	Nombre d'entreprises
<b>TPE</b>	7
<b>PME</b>	2
<b>Grandes entreprises</b>	10

Pour pouvoir l'exporter d'une manière à ce qu'il respecte les critères d'anonymisation, on peut masquer l'effectif posant problème en le remplaçant par un « S », comme indiqué ci-dessous.

Catégorie d'entreprise	Nombre d'entreprises
<b>TPE</b>	7
<b>PME</b>	<b>S</b>
<b>Grandes entreprises</b>	10

#### **Attention au secret secondaire !**

Cette technique est plus risquée car il faut faire attention à ce que d'autres informations disponibles par ailleurs ne puissent pas permettre de recalculer l'information masquée. Pour contrôler ce risque, il faut appliquer ce que l'on appelle le **secret secondaire** qui vise à empêcher la reconstitution, par somme ou par différence, des cases masquées. Dans l'exemple donné, si on a par ailleurs un tableau qui donne le total des nombres d'entreprises, la valeur masquée est facilement calculable. On doit alors masquer deux valeurs pour pouvoir l'exporter tout en respectant le secret secondaire.

### 2.3.2 La diversification des données

L'objectif de la diversification des données est d'obtenir des caractéristiques d'un groupe suffisamment diversifiées pour réduire les risques de déduction certaine ou quasi certaine.

#### **Exemple :**

Le tableau ci-dessous présente des données fictives des diagnostics pris en charge lors des séjours hospitaliers pour un mois et un département donnés.

Diagnostics pris en charge lors des séjours hospitaliers pour un mois et un département donnés						
Catégorie d'âge	Effectif total	Hypertension	Diabète	Asthme	Cancer	Insuffisance respiratoire
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

Ce tableau présente deux problèmes :

- Les personnes appartenant à la catégorie d'âge « 20-29 » sont toutes diagnostiquées avec de l'asthme. Par conséquent, si on connaît par ailleurs qu'une personne, ayant entre 20 et 29 ans, a été hospitalisée ce mois donné, dans ce département donné, on en déduira d'une façon certaine que cette personne a de l'asthme.
- Une grande majorité des personnes appartenant à la catégorie d'âge « 70-79 » sont diagnostiquées avec une insuffisance respiratoire. Par conséquent, si on connaît par ailleurs qu'une personne, ayant entre 70 et 79 ans, a été hospitalisée ce mois donné, dans ce département donné, on en déduira d'une façon quasi-certaine que cette personne a souffert d'insuffisance respiratoire.

Comme précédemment, pour rendre ce tableau conforme aux règles de confidentialité il est possible soit d'agréger, soit de supprimer les valeurs posant problème.

### 2.3.3 Les fortes contributions

Dans certains cas, pour respecter les règles de confidentialité, les données ne doivent pas contenir de fortes contributions quand s'il s'agit de variables de montant. Autrement dit, une unité ne doit pas contribuer pour plus d'un certain pourcentage dans une variable de montant. Cela permet d'éviter que ce montant puisse lui être, en tout ou en trop grande partie, attribué.

#### **Exemple :**

Le tableau suivant représente les chiffres d'affaires dans la construction pour l'année 2019.

Secteur d'activité	Effectif	Chiffre d'affaires total
<b>Construction de bâtiment</b>	467	860 745
<b>Génie civil</b>	389	1 696 872
<b>Travaux de construction spécialisés</b>	804	973 610

En regardant par ailleurs la contribution des valeurs maximales, il apparaît qu'une entreprise occupe une position quasi-monopolistique dans le secteur du génie civil.

Secteur d'activité	Chiffre d'affaires maximum	Pourcentage du maximum
Construction de bâtiment	256 804	29,83%
Génie civil	1 531 794	90,27%
Travaux de construction spécialisés	41 947	4,30%

Par conséquent le premier tableau des chiffres d'affaires par secteur d'activité présente un risque élevé de déduction quasi-certaine. Comme dans les deux cas précédents, deux solutions sont possibles : agréger ou supprimer les valeurs posant problème.

### 3 DESCRIPTION DES FICHIERS POUR LES SORTIES

---

Tout d'abord, vous devez indiquer **les sources de données utilisées** et décrire de manière précise le contenu de votre sortie, soit dans votre mail de demande de sortie, soit dans un fichier texte inclus dans la sortie. Par ailleurs, et afin de pouvoir vérifier vos sorties, vous devez inclure des informations spécifiques aux types des fichiers sortis.

#### 3.1 LES PROGRAMMES

Vous pouvez demander à sortir vos programmes pour les réutiliser en dehors du CASD. Tout programme sorti ne doit pas contenir de données confidentielles.

#### 3.2 LES REGRESSIONS, MODELES ECONOMETRIQUES

Pour les résultats de régressions ou modèles économétriques, vous devez indiquer le **nombre d'observations** afin de permettre au CASD de s'assurer du respect des règles de confidentialité.

#### 3.3 LES CARTES ET LES GRAPHIQUES

Les cartes peuvent concerner des populations très fines (communales ou infra-communales). Afin que le CASD puisse vérifier que les règles de confidentialité sont bien respectées sur la carte, vous devez fournir **la population qui a servi de base à la construction de cette carte**.

Pour les graphiques comme les courbes, les histogrammes, les nuages de points, les diagrammes, etc., vous devez fournir **la population et la signification des variables utilisées**.

D'autres types de graphiques sont plus complexes à traiter car ils peuvent contenir des informations individuelles. Par exemple, les box plot peuvent contenir, non seulement le maximum et le minimum, mais aussi des points extrêmes (outliers) qui ne doivent pas être identifiés.

Attention aux graphiques d'analyse factorielle représentant les individus s'ils contiennent des individus atypiques qui peuvent être identifiés (par exemple une ACP sur des entreprises où le SIRET est utilisé sur le graphique comme identifiant de chaque point).



**Attention aux graphiques STATA au format LIVE !**

Les graphiques **STATA** au format **LIVE** contiennent la base de données qui a servi à les constituer, il est préférable de les convertir dans un autre format (AS-IS : format ne contenant pas les bases de données, pdf, jpg...).

**3.4 LES TABLEAUX DE DONNEES AGREGES**

Pour les tableaux de données agrégées, vous devez fournir dans la description de votre sortie **la signification** de toutes les variables utilisées, **les effectifs** correspondant à chaque case et pour les variables de montant, l’information portant sur la **plus forte contribution** dans la case dans **un fichier de contrôle non secretisé**.

**Fichier de contrôle :**

Afin de vérifier que les résultats portant sur les variables de montant ne contiennent pas de fortes contributions, vous devez inclure dans votre sortie un fichier de contrôle contenant l’information du maximum et le pourcentage que représente ce maximum pour les variables de montant.

Ce fichier de contrôle, **contenant donc des données confidentielles sur une seule unité, sera retiré de votre sortie avant son envoi. Il est ainsi nécessaire de le mettre à part des autres fichiers que vous souhaitez sortir.**

**Exemple :**

Le tableau suivant reprend l’exemple sur les sommes des chiffres d’affaires dans la construction pour l’année 2019. Il s’agit d’un tableau qui contient des informations de contrôle : le chiffre d’affaires maximum et le pourcentage que ce maximum représente dans les différents secteurs d’activité de la construction.

Secteur d’activité	Effectif	Chiffre d’affaires total	Chiffre d’affaires maximum	Pourcentage du maximum
<b>Construction de bâtiment</b>	467	860 745	256 804	29,83%
<b>Génie civil</b>	389	1 696 872	1 531 794	90,27%
<b>Travaux de construction spécialisés</b>	804	973 610	41 947	4,30%

Le tableau de contrôle montre qu’une entreprise est en situation de quasi-monopole dans le secteur du Génie Civil. Ainsi la variable de montant correspondante, en l’occurrence le chiffre d’affaires dans le secteur du Génie Civil présente un risque élevé de déduction quasi certaine. La valeur posant problème doit être soit agrégée soit supprimée. Cette dernière technique est choisie dans le tableau ci-dessous qui est destiné à être exporté.

Secteur d'activité	Effectif	Chiffre d'affaires total
<b>Construction de bâtiment</b>	467	860 745
<b>Génie civil</b>	389	<b>S</b>
<b>Travaux de construction spécialisés</b>	804	973 610

### 3.5 LES ARTICLES « FINALISES »

Les articles finalisés que vous souhaitez sortir de votre espace projet CASD ne doivent pas contenir de données sur une population ne respectant pas les règles indiquées ci-après.

## 4 REGLES GENERALES

Des règles de confidentialité ont été définies afin de respecter l'obligation de confidentialité. Il existe d'abord des règles générales, qui s'appliquent à toutes les sources.

### 4.1 LES DONNEES « MENAGES »

Le secret statistique impose qu'une information n'est publiable que si elle a fait l'objet d'un traitement rendant impossible l'identification des personnes physiques. C'est une obligation qui est également applicable aux entrepreneurs individuels.

En pratique, la règle de confidentialité imposée est que la connaissance d'une caractéristique pour un individu ne doit pas entraîner la connaissance d'une autre caractéristique avec laquelle elle est croisée dans un tableau. Cette règle évoque le principe de « non-corrélation » décrit ci-dessus.

#### *Exemple :*

Le tableau ci-dessous donne la répartition par âge et la situation matrimoniale et indique que les personnes entre 50 et 59 ans ont toutes le même état matrimonial « divorcé ». Le secret statistique n'est plus respecté dans ce tableau, et ce dernier n'est donc pas diffusable. En effet, si l'on sait par ailleurs qu'un individu donné se situe dans la tranche des 50-59 ans, le tableau nous informe que cette personne est divorcée, et ceci même si la case qui croise les modalités « 50 à 59 ans » et « divorcé » comporte plusieurs individus.

Situation matrimoniale et classe d'âge	18-25 ans	26-49 ans	50-59 ans	60 ans et +
<b>Marié</b>	7	27	<b>0</b>	30
<b>Divorcé</b>	0	11	<b>9</b>	22
<b>Autre</b>	21	12	<b>0</b>	4

## 4.2 LES DONNEES « ENTREPRISES »

Pour les tableaux fournissant des données agrégées sur les entreprises, les règles sont les suivantes :

- Aucune case ne doit concerner moins de **trois unités** (décision du 13 juin 1980 du directeur général de l'Insee)
- Aucune case ne doit contenir des données pour lesquelles une entreprise représente plus de **85% du total d'un montant** (règle du CNIS, 7 juillet 1960)

Les règles à suivre s'appliquent aux SIREN, c'est-à-dire aux unités dotées de la personnalité juridique et non aux SIRET, c'est-à-dire les établissements. Elles concernent également les entrepreneurs individuels. Si vous souhaitez sortir des données sur des établissements, vous devez indiquer le nombre d'entreprises correspondant.

**Note** : La sortie de liste de SIREN/SIRET associée à d'autres informations n'est pas autorisée.

## 4.3 LES DONNEES « AGRICOLES »

Pour les tableaux fournissant des données agrégées sur les exploitations, les règles sont les suivantes :

- Aucune case ne doit concerner moins de **trois unités**
- Aucune case ne doit contenir des données pour lesquelles une entreprise représente plus de **85% du total d'un montant**

Les règles à suivre s'appliquent aux exploitations, et non aux parcelles.

## 4.4 LES DONNEES « MIXTES »

Les sources mixtes sont des sources provenant de combinaisons (appariements) d'enquêtes statistiques et de données administratives ou bien des sources comportant à la fois des informations d'ordre économique et financier (entreprises) et des informations relatives à des faits et comportements d'ordre privé (ménages).

La démarche à adopter face à de telles sources est, dans son principe, très simple : les règles à prendre en considération s'obtiennent par le cumul des règles applicables d'une part aux enquêtes statistiques, d'autre part aux fichiers administratifs. Par exemple, les données FARE (Fichier approché des résultats d'Esane) contiennent d'une part des données sur les entreprises, et d'autre part des données issues des liasses fiscales. La règle de vérification déduite (cumul des règles) s'applique à toutes les variables de la source mixte.

**De façon générale, prêtez un soin particulier aux sources appariées aux données fiscales (liste référencée dans la partie 5.2 – Données de la DGFIP).**

## 4.5 LES DONNEES ISSUES D'ENQUETES

Dans le cas d'enquêtes statistiques portant sur des échantillons, les unités font l'objet d'une ou plusieurs pondérations afin de rendre l'ensemble représentatif de la population étudiée. Les règles de confidentialité s'appliquent **par vérification des données pondérées**.

## 5 REGLES DETAILLEES POUR CHAQUE SOURCE

---

S'ajoutent aux règles générales, des règles spécifiques par sources de données qui sont détaillées dans ce qui suit. Ces règles ont été définies par les producteurs de données selon la nécessité de confidentialité qu'ils estiment par rapport à leurs données.

### 5.1 DONNEES DE L'INSEE

Pour les sources produites par l'Insee, les règles du secret sont décrites dans le guide du secret statistique. Ce guide est disponible à l'adresse :

<https://www.insee.fr/fr/statistiques/fichier/1300624/guide-secret.pdf>

#### 5.1.1 DADS (Déclaration annuelle de données sociales)

Tout tableau diffusé ne doit en aucun cas permettre l'identification directe ou indirecte d'un salarié ou d'un établissement.

- Tableaux au lieu de résidence (logique « ménages »)
  - Aucune case ne doit comporter moins de 5 salariés
  - Aucune case ne doit comporter un salarié représentant plus que 80% de la masse salariale
- Tableaux au lieu de travail (logique « entreprises » en plus), **en plus des deux règles précédentes qui s'appliquent**, s'ajoutent les règles concernant les données sur les entreprises qui sont les suivantes :
  - Aucune case ne doit comporter moins que 3 entreprises ou établissements
  - Aucune case ne doit comporter une entreprise ou un établissement représentant plus de 85% du total

#### 5.1.2 CLAP (Connaissance locale de l'appareil productif) et FLORES (Fichier localisé des rémunérations et de l'emploi salarié)

CLAP et FLORES font partie des données « entreprises ». Les indicateurs qui sont soumis au secret statistique sont les effectifs et les rémunérations. Les règles à suivre sont les suivantes :

- Aucune case ne doit concerner moins de 3 unités (l'unité est un établissement ou une entreprise)
- Aucune case ne doit comporter une unité représentant plus que 85% du total
- Aucune case ne doit concerner moins de 5 salariés

#### 5.1.3 FARE (Fichier approché des résultats d'Esane) / FICUS (Fichier de comptabilité unifié dans SUSE)

Les données FARE et FICUS sont des sources mixtes « fiscale et statistique ». Les règles de confidentialité applicables sont donc le cumul des règles applicables d'une part aux enquêtes statistiques, d'autre part aux données fiscales. Elles sont alors les suivantes :

- Aucune case ne doit concerner moins de 3 unités
- Aucune case ne doit comporter une unité représentant plus que 85% du total
- Pour les entreprises individuelles, aucune case ne doit concerner moins de 11 unités

#### 5.1.4 SINE (Système d'information sur les nouvelles entreprises)

SINE est également une source « entreprises », les règles de confidentialité à appliquer sont les suivantes :

- Aucun résultat qui concerne moins de trois entreprises

- Aucune donnée pour laquelle une seule entreprise représente 85 % ou plus de la valeur obtenue

Par ailleurs, les taux de survie ne doivent pas être calculés pour des populations de moins de 20 entreprises. Ce seuil minimum de 20 entreprises est également requis pour les zonages et regroupements particuliers.

### 5.1.5 Recensement de la Population

Les principes de diffusion des données du RP ont évolués à partir de 2006 avec l'arrêté du 19 juillet 2007 relatif à la diffusion des résultats du recensement de la population. Celui-ci est venu en remplacement de l'arrêté du 22 mai 1998, modifié le 8 avril 2002, relatif à la diffusion des résultats du recensement de la population.

→ En raison des taux de sondage des différentes exploitations des recensements de la population, on retiendra les seuils minimums de diffusion suivants pour les tableaux de résultats :

- Pour le RP jusqu'en 1999 :
  - RP exploitation principale : au moins 4 unités
  - RP exploitations complémentaires :

Année du RP	Taux de sondage	Seuil minimal de diffusion (après pondération)
1962	1/20	80 unités
1968	1/4	16 unités
1975	1/5	20 unités
1982	1/4	16 unités
1990	1/4	16 unités
1999	1/4	16 unités

- Pour le RP rénové (après 2006) : au moins 10 unités après pondération

→ Respecter les seuils suivants de diffusion des variables dites « sensibles » : nationalité actuelle (ou à la naissance), lieu de naissance, lieu de résidence antérieur, année (ou ancienneté) d'arrivée en France. Les notions d'immigrés et de français par acquisition sont dans le champ des variables sensibles jusqu'au recensement de 1999.

Zones minimales de diffusion de ces variables, seuil après pondération :

1999 et avant	A partir de 2006 (RP annuel)
<b>Communes de plus de 5 000 habitants</b>	Communes de plus de 5 000 habitants
<b>Seuil de 10 000 habitants pour les arrondissements, zones d'emploi, aires urbaines, unités urbaines (ou leurs regroupements) et zones de la politique de la ville ou leurs regroupements</b>	Seuil de 5.000 habitants pour les arrondissements, zones d'emploi, aires urbaines, unités urbaines et zones de la politique de la ville
<b>Zones infracommunales résultant du regroupement de 3 quartiers (zone fixe résultant du découpage de la commune en zones géographiques d'un seul tenant d'environ 2 000 habitants)</b>	Zones infracommunales résultant du regroupement de 3 quartiers (zone fixe résultant du découpage de la commune en zones géographiques d'un seul tenant d'environ 2 000 habitants)
<b>Département pour l'année (ou ancienneté) d'arrivée</b>	Département pour l'année (ou ancienneté) d'arrivée

**Pour les sorties qui concernent les variables sensibles du recensement, afin de vérifier les fichiers à sortir, vous devez préciser les échelles géographiques utilisées pour construire les données. Dans le cas où ces dernières ont été agrégées au niveau des communes, vous devez fournir le nombre d'habitants dans chaque commune ou groupe de communes pour toutes les variables concernées.**

**5.1.6 SIASP (Système d'information sur les agents des services publics) et FGE (Fichier Général de l'Etat)**  
 La diffusion de résultats statistiques tirés des données issues de SIASP ou de FGE doit être conforme aux dispositions prévues par les textes relatifs au secret en matière de statistique et à la protection des données individuelles. En particulier, aucun tableau destiné à la diffusion ne doit permettre l'identification directe ou indirecte d'un salarié ou d'un établissement.

Pour les tableaux au lieu de résidence les critères sont les suivants :

- Aucune case ne doit comporter moins de 5 salariés
- Aucune case ne doit comporter un salarié représentant plus de 80% de la masse salariale de la case

Pour les tableaux au lieu de travail, **en plus des deux critères précédents**, les critères sont les suivants :

- Aucune case ne doit se rapporter à moins de 3 établissements
- Aucun établissement ne doit représenter plus de 85 % de la grandeur étudiée dans la case

Exception : pour le champ de la fonction publique d'État, aux niveaux national, régional ou départemental, il n'est pas toujours possible de trouver 3 établissements sirétisés. Par conséquent, cette règle ne sera pas appliquée, uniquement dans ce contexte.

### **5.1.7 FIDELI (Fichiers DÉmographiques sur les Logements et les Individus)**

Les règles pour FIDELI sont les suivantes :

- Pour les communes de 5000 habitants ou plus, la maille géographique la plus fine pour produire des tableaux est la commune ou l'IRIS ;
- Pour les communes de moins de 5000 habitants, la maille géographique la plus fine pour produire des tableaux est l'Établissement public de coopération intercommunale (EPCI). Les EPCI, pour lesquels la population totale des communes de moins de 5000 habitants dans l'EPCI est inférieure à 2000 habitants, ne peuvent toutefois donner lieu à la production de tableaux ;
- Pour les résultats portant sur les quartiers prioritaires de la politique de la ville (QPV), la maille géographique la plus fine pour produire des tableaux est la région ;
- Pour tout tableau produit, chaque case doit comporter plus de 11 individus.

**Pour les sorties qui concernent des données issues de FIDELI, afin de vérifier les fichiers à sortir, vous devez préciser les échelles géographiques utilisées pour construire les données de votre sortie. Dans le cas où les données ont été agrégées au niveau des communes ou des EPCI, vous devez fournir le nombre d'habitants dans chaque commune pour toutes les variables concernées.**

## **5.2 DONNEES DE LA DGFIP – MINISTERE DES FINANCES**

Pour les données sur les entreprises :

- Aucune case du tableau ne doit concerner moins de trois unités (décision du 13 juin 1980 du directeur général de l'Insee)
- Aucune case du tableau ne doit contenir des données pour lesquelles une entreprise représente plus de 85% du total (règle du CNIS, 7 juillet 1960)

Pour les données sur les ménages : aucune case ne doit comporter moins de 11 individus.

Pour les données de nature fiscale (règles avalisées par la CNIL dans un avis du 27 mai 1997 et énoncées au §30 du BOI-DJC-CADA-20 :

- Règle du nombre d'unités : une donnée agrégée n'est pas communiquée lorsqu'elle concerne moins de 11 unités
- Règle du poids des unités : une donnée agrégée n'est pas communiquée lorsqu'elle comprend un élément dominant qui représente plus de 85 % du montant agrégé

Les sources de données comprenant des résultats d'enquêtes appariées à des informations fiscales entrent dans le champ de ces règles de confidentialité. Il s'agit plus précisément de :

- Statistiques sur les Ressources et les Conditions de Vie (SRCV) depuis 2008
- Enquête Patrimoine
- Enquête Histoire de Vie et Patrimoine (EHVP)
- Enquête Budget de Famille (BDF)
- Enquête Formation et Qualification Professionnelle (FQP) 2014
- Enquête logement (ENL) 2006 et 2013
- Echantillon interrégimes de retraités (EIR)
- Remontées individuelles des demandeurs et bénéficiaires de la PCH (RI-PCH)
- Enquête Capacités, Aides et Ressources des seniors – volet institutions (CARE-I)
- Enquête Capacités, Aides et Ressources des seniors – volet ménages (CARE-M)
- Enquête Performance de l'Habitat, Equipements, Besoins et Usages de l'énergie – volet CLODE (PHEBUS CLODE)
- Enquête Performance de l'Habitat, Equipements, Besoins et Usages de l'énergie – volet DPE (PHEBUS DPE)
- Enquête Bénéficiaires de Minima Sociaux (BMS)
- Contrat Uniques d'Insertion (CUI)

### 5.3 DONNEES DE LA DARES – MINISTERE DU TRAVAIL

Les règles pour les données de la DARES sont les suivantes :

- Pour les données individus/ménages : respecter un effectif minimum de 5 individus par case
- Pour les données entreprises : respecter la règle de 3 entreprises au minimum par case et aucune case ne doit contenir des données pour lesquelles une entreprise représente plus de 85% du total

### 5.4 DONNEES DE LA DEPP – MINISTERE DE L'ÉDUCATION

Les règles pour les données de la DEPP sont les suivantes :

- Pour les données sur les individus : aucune case avec moins de 10 individus
- Pour les données agrégées au niveau établissement (collèges, lycée, etc.) : les cases ne doivent pas contenir moins de 10 établissements

### 5.5 DONNEES DU SSP – MINISTERE DE L'AGRICULTURE

Pour les tableaux fournissant des données agrégées agricoles, les règles sont les suivantes :

- Aucune case ne doit concerner moins de **trois unités**

- Aucune case ne doit contenir des données pour lesquelles une exploitation représente plus de **85% du total d'un montant**

Les règles à suivre s'appliquent aux exploitations, c'est-à-dire aux unités dotées de la personnalité juridique et non aux parcelles. Si vous souhaitez sortir des données sur des parcelles, vous devez indiquer le nombre d'exploitations correspondant.

#### 5.5.1 PKGC (Pratiques culturelles sur les grandes cultures)

- Lorsque l'effectif total d'un tableau porte sur moins de 30 parcelles au total, celui-ci ne peut pas être diffusé
- Aucune case d'un tableau ne doit concerner moins de trois parcelles agrégées

#### 5.5.2 Ecopeche (Enquête sur la production des données économiques dans le secteur des pêches maritimes)

- Aucune information ne peut concerner moins de 5 navires

### 5.6 DONNEES DE FRANCE AGRIMER

Les règles sont les suivantes (applicables aux tableaux) :

- avoir dans chaque case complétée du tableau au moins 5 individus/entreprises unipersonnelles
- pour les autres entreprises : avoir au moins 3 entreprises dans chaque case du tableau.

### 5.7 DONNEES DE LA DPMA – MINISTERE DE L'AGRICULTURE

Aucune information (case, item de graphique/carte, etc.) ne doit concerner strictement moins de 5 navires uniques (champ CFR\_COD).

### 5.8 DONNEES DE LA SDSE – MINISTERE DE LA JUSTICE

#### 5.8.1 Fichiers CPH, PACS et fichier statistique du casier judiciaire national

Au moins 5 unités statistiques pour chaque case diffusée ou calculable par recoupement d'informations à partir des données diffusées. Cela implique que, dans le cas où toutes les modalités d'une variable ne sont pas présentées, ce minimum de 5 unités s'applique également au complémentaire à l'effectif des modalités présentées sur l'effectif total de la population étudiée. Cela implique également qu'il ne doit pas être fait mention d'une distribution unanime d'une caractéristique (autre que l'application d'un critère légal) au sein d'une sous-population.

#### 5.8.2 Fichier statistique des tribunaux de commerce

Deux conditions :

- Au moins 3 unités statistiques pour chaque case diffusée ou calculable par recoupement d'informations à partir des données diffusées
- Au sein de chaque case diffusée ou calculable, aucune unité statistique ne peut avoir de contribution supérieure à 85 % à l'agrégat présenté

Ces deux règles s'appliquent également, dans le cas où toutes les modalités d'une variable ne sont pas présentées, aux effectifs et contributions des modalités complémentaires à celles présentées.



## **5.9 DONNEES DU SDES – MINISTERE DE L'ENVIRONNEMENT**

### **5.9.1 EPTB (Enquête sur les prix des terrains à bâtir)**

A minima un effectif de 11 permis dans chaque case du tableau. Il faut se référer à l'effectif non pondéré.

### **5.9.2 ECLN (Enquête sur la commercialisation des logements neufs)**

Pour les données de prix de vente ou de réservation, aucune case ne doit concerner moins de 3 promoteurs dont aucun ne peut représenter à lui seul plus de 85% de la donnée agrégée.

Pour les données de réservation et en cours de logements, aucune restriction n'est à appliquer pour les zones géographiques comprenant au moins 50 000 habitants et pour lesquelles au moins 5 promoteurs ont eu un programme de commercialisation au cours de l'année civile précédente. Pour les zones géographiques ne vérifiant pas ces conditions, un effectif d'au moins 3 promoteurs doit être respecté dont aucun ne représente à lui seul 85 % du total.

Pour les autres données de mise en vente, sur lesquelles le promoteur fait sa promotion, à savoir : la référence du pétitionnaire, les caractéristiques du programme (adresse, type de logements, nombre de logements, trimestre de début et de fin de commercialisation), il n'y a aucune restriction à appliquer quelle que soit la zone géographique.

### **5.9.3 FILOCOM (Fichier des logements par communes)**

Pour les variables liées au logement, les résultats doivent concerner au moins 11 logements.

Pour les variables liées aux occupants et à leur occupation, un effectif minimal de 11 foyers fiscaux doit être respecté, et une unité ne peut représenter plus de 85 % du total d'une case. Cela comprend les entreprises individuelles.

### **5.9.4 SITADEL (Système d'Information et de Traitement Automatisé des Données Élémentaires sur les Logements et les locaux)**

Le fichier SITADEL mis à disposition par le CASD n'est pas couvert par la confidentialité statistique. Il a vocation à être en open data.

### **5.9.5 Rép-Entrepôts (Répertoire des entrepôts et des plateformes logistiques)**

- Aucune agrégation ne doit concerner moins de trois établissements
- Aucune agrégation ne doit contenir des données pour lesquelles un établissement représente plus de 85% de la surface totale d'entreposage de l'agrégation

## **5.10 DONNEES DU SIES – MINISTERE DE LA RECHERCHE**

Les règles pour les données du SIES sont les suivantes :

- Pour les données sur les établissements éducatifs ou les entreprises :
  - Aucune case du tableau ne doit concerner moins de trois unités
  - Aucune case du tableau ne doit contenir des données pour lesquelles un établissement représente plus de 85% du total
- Pour les données sur les étudiants, l'identification directe ou indirecte des individus doit être impossible. Dans la pratique, on considèrera que le secret statistique est respecté si aucune case du tableau ne concerne moins de cinq individus.

### 5.11 DONNEES DU DSED – MINISTERE DE L’IMMIGRATION

Aucune analyse individuelle ou territoriale ne sera autorisée (en dehors des typologies de territoires). Tout tableau diffusé ne doit en aucun cas permettre l’identification directe ou indirecte d’un individu :

- Aucune case ne doit comporter moins de 10 unités (avant pondération)
- Aucune case ne doit comporter une unité représentant plus que 85% du total analysé.

Par ailleurs, les analyses géographiques visant à comparer des zones géographiques (départements, régions, communes, IRIS, QPV...) ne seront pas autorisées. Seules les analyses comparant des **typologies de territoires** (ex : rural versus urbain...) seront autorisées.

### 5.12 DONNEES CIFRE DU MESRI – MINISTERE DE LA RECHERCHE

Les règles pour les données CIFRE du MESRI sont les suivantes :

- S’agissant des particuliers, il est interdit de publier des données qui permettraient une identification directe ou indirecte d’une personne ou même, sans pouvoir l’identifier, d’obtenir une information à son sujet. Ces règles limitent la finesse des informations disponibles en diffusion. Des règles très strictes sont fixées spécifiquement pour les recensements de la population. Pas de sortie de données sur moins de 10 observations en ce qui concerne les individus.
- S’agissant des entreprises, aucun résultat n’est publié s’il concerne moins de trois entreprises, ni aucune donnée pour laquelle une seule entreprise représente 80% ou plus de la valeur obtenue. Cependant, il est admis que la diffusion de listes extraites du répertoire des entreprises ou des établissements peut mentionner l’activité économique, une classe d’effectifs et une tranche de chiffres d’affaires.
- Pas d’édition de listes donnant le nom, l’adresse ou autre caractéristique individuelle des étudiants bénéficiaires ou des entreprises.

### 5.13 DONNEES DE LA DGE

Aucune case du tableau ne doit :

- Concerner moins de trois unités
- Contenir de données pour lesquelles une entreprise représente plus de 85% du total

### 5.14 DONNEES DE L’INJEP

Pour les tableaux de données agrégées sur les licenciés :

- Aucune case ne doit concerner moins de 3 unités ;
- Les données de la Fédération française Maccabi, de la Fédération sportive de la police nationale et de la Fédération des clubs de la défense devront être agrégées avec celles d’une autre fédération et ne devront être traitées isolément.

Pour les tableaux de données agrégées de l’Enquête nationale sur les pratiques physiques et sportives :

- La maille géographique la plus fine pour produire des tableaux est la région ;
- Aucune case ne doit concerner moins de 10 unités non pondérées.

### 5.15 DONNEES DU CEREQ

A minima un effectif de 5 observations dans chaque case du tableau.

### **5.16 DONNEES DE BPIFRANCE**

A minima un effectif de 10 observations dans chaque case du tableau.

### **5.17 DONNEES DE L'ANIL**

Les règles pour les données de l'ANIL sont les suivantes :

- Pas de diffusion de données au logement
- A minima un effectif de 50 observations par résultat (case d'un tableau agrégé, données pour calibration de modèles)

### **5.18 DONNEES DE L'ODR**

Ne pas sortir des informations des données à une échelle géographique inférieure au canton en respectant la règle de 3 observations minimum par case.

**Pour les sorties qui concernent les données de l'ODR, afin de vérifier les fichiers à sortir, vous devez préciser les échelles géographiques utilisées pour construire les données de votre sortie.**

### **5.19 DONNEES DE LA MSA**

Les règles pour les données de la MSA sont les suivantes :

- Aucune case du tableau ne doit concerner moins de 5 unités
- Aucune case du tableau ne doit contenir des données pour lesquelles un cotisant non-salarié représente plus de 85% du total
- La caractéristique pour un individu ne peut pas entraîner la connaissance d'une autre caractéristique avec laquelle elle est croisée dans un tableau

### **5.20 DONNEES DE L'ACOSS**

A minima un effectif de 10 observations dans chaque case du tableau.

### **5.21 DONNEES DE LA CNAF**

A minima un effectif de 5 observations dans chaque case du tableau.

### **5.22 DONNEES BMO DE POLE EMPLOI**

Aucune case ne doit comporter moins de 60 observations (observations non pondérées).

### **5.23 DONNEES DU PMSI DE L'ATIH**

Aucune case ne doit concerner moins de 11 unités relatives à des nombres de patients, à des nombres de séjours et à des nombres de doses.

### **5.24 DONNEES DE LA DREES**

Les règles pour les données de la DREES sont les suivantes :

- Pour les données individus/ménages, hors données fiscales, l'effectif minimum est de 5 individus par case
- Pour les données fiscales, l'effectif minimum est de 11 individus par case

## **5.25 DONNEES DE L'IRDES**

Pour les données HYGIE et ESPS, aucun résultat sur moins de 15 individus.

CASD 