

GUIDE DU SECRET STATISTIQUE

Les agents du [service statistique public](#), comme tout agent public, sont tenus au secret professionnel à l'égard des données confidentielles dont ils peuvent être dépositaires au titre de leurs missions ou de leurs fonctions ([article L121-6](#) du Code général de la fonction publique). Ils s'exposent, en cas de violation de cette obligation, à des sanctions pénales qui peuvent aller jusqu'à un an d'emprisonnement et 15 000 euros d'amende ([article 226-13](#) du Code pénal).

Le **secret statistique** renforce l'obligation de secret professionnel lorsqu'il s'applique **aux données confidentielles que le service statistique public détient, collecte ou produit à des fins statistiques**. Le secret statistique garantit aux personnes concernées que leurs données ne peuvent être utilisées ou réutilisées qu'à des fins statistiques, ou de recherche scientifique ou historique, et exclut que l'usage de ces données puissent venir à l'appui de toute mesure ou décision prise à l'égard d'une personne en particulier. Le secret statistique est ainsi opposable à toute demande de réquisition, qu'elle émane d'une autorité judiciaire ou d'une autorité administrative (services fiscaux ou des Douanes par exemple).

Le secret statistique garantit ainsi, notamment, le respect :

- De la **confidentialité due à la vie privée, personnelle et familiale**, pour les personnes physiques ;
- Du **secret commercial et des affaires**, pour les entreprises.

Les obligations relatives au secret statistique relèvent de textes de niveaux national et européen.

Le secret statistique découle à l'échelon national des obligations définies par la [loi n° 51-711 du 7 juin 1951](#) modifiée sur l'obligation, la coordination et le secret en matière de statistiques définit le secret statistique.

À l'échelon européen, la confidentialité des informations statistiques est affirmée par [l'article 338](#) du traité de l'Union.

Le secret statistique fait l'objet du chapitre V du règlement [n° 223/2009](#) modifié sur les statistiques européennes, et du règlement d'application [n° 557/2013](#) en ce qui concerne l'accès aux données



confidentielles à des fins statistiques. Le règlement 223 interdit toute communication de données confidentielles obtenues pour la production de statistiques européennes, sauf motif statistique ou de recherche. Relevant de règles de communication définies par des normes européennes, de rang supérieur, ces données échappent ainsi par ailleurs à toute règle de communication ou de diffusion applicable strictement en droit national.

Le secret statistique dépend aussi des dispositions introduites par le [règlement général sur la protection des données](#) (RGPD) concernant le principe de finalité. Celui-ci interdit toute réutilisation de données à caractère personnel issues d'un traitement à finalité statistique à d'autres fins que statistiques ou de recherche¹, ce qui en pratique revient à imposer le secret statistique à la diffusion ou la communication de ces données.

¹ Voir le considérant 162 du RGPD : « Par « fins statistiques », on entend toute opération de collecte et de traitement de données à caractère personnel nécessaires pour des enquêtes statistiques ou la production de résultats statistiques. Ces résultats statistiques peuvent en outre être utilisés à différentes fins, notamment des fins de recherche scientifique. Les fins statistiques impliquent que le résultat du traitement à des fins statistiques ne constitue pas des données à caractère personnel mais des données agrégées, et que ce résultat ou ces données à caractère personnel ne sont pas utilisés à l'appui de mesures ou de décisions concernant une personne physique en particulier. »



Sommaire

Table des matières

Règles pratiques pour le respect du secret statistique	4
A. Informations tirées d'enquêtes statistiques	4
1. Tableaux agrégés	4
2. Fichiers de données individuelles.....	5
B. Informations tirées de sources administratives	7
C. Le cas des sources mixtes	8
1. Fichiers combinant des données statistiques et des données administratives	8
2. Fichiers combinant des données sur les entreprises et des données sur les ménages	9
Annexe.....	10
Définitions	12

Règles pratiques pour le respect du secret statistique

Il faut distinguer les informations collectées au moyen d'enquêtes statistiques et celles qui ont été transmises par des tiers au service statistique public à des fins d'établissement des statistiques (données administratives).

Les enquêtes statistiques sont celles qui ont reçu le visa prévu par l'article 2 de la loi [n° 51-711 du 7 juin 1951](#) sur l'obligation, la coordination et le secret en matière de statistiques et qui figurent dans la liste des enquêtes de la statistique publique publiée chaque année au *Journal Officiel*.

A. Informations tirées d'enquêtes statistiques

Le secret doit être respecté dans les tableaux agrégés ([voir annexe](#)) et les fichiers de données individuelles.

1. Tableaux agrégés

- a) Pour les tableaux fournissant des données agrégées sur les **entreprises**, la règle est la suivante :
 - 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 de données pour lesquelles une entreprise représente plus de 85% du total (règle de diffusion définie le 7 juillet 1960 par le Comité de coordination des enquêtes statistiques, prédécesseur du CNIS, Conseil National de l'Information Statistique).

- b) Pour les tableaux fournissant des données agrégées sur les **personnes physiques**, il ne doit pas être possible d'identifier, dans le principe, les personnes ([article L.312-1-2](#) du code des relations entre le public et l'administration). Des critères applicables en cas de diffusion ont été définis par la [Cnil](#) (« [L'anonymisation des données, un traitement clé pour l'open data](#) »). La Cnil souligne en particulier les risques liés, dans le cas de données agrégées, à la réidentification **par inférence**, correspondant aux cas où il est possible de déduire de façon quasi certaine de nouvelles informations sur un individu².

Néanmoins, conformément à l'[article 116 du décret n° 2019-536 du 29 mai 2019](#) pris pour l'application de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux

² Exemple (repris de la Cnil) : « si un jeu de données supposément anonyme contient des informations sur le montant des impôts de personnes ayant répondu à un questionnaire, que tous les hommes ayant entre 20 et 25 ans ayant répondu sont non imposables, il sera possible de déduire, si on sait que M. X, homme âgé de 24 ans, a répondu au questionnaire, qu'il est non imposable. »

fichiers et aux libertés, les données issues de traitements à des fins statistiques « *ne peuvent pas être diffusées sans avoir été préalablement anonymisées **sauf si l'intérêt des tiers à cette diffusion prévaut sur les intérêts ou les libertés et droits fondamentaux de la personne concernée.*** » Il faut considérer par ce biais que la balance entre l'intérêt du public d'une part et celui des personnes concernées d'autre part peut permettre, selon les cas, en fonction notamment de la sensibilité des données diffusées, d'aménager des modalités de diffusion, ne respectant pas strictement le principe d'anonymisation préalable, tout en garantissant un niveau suffisant de protection pour la vie privée des personnes concernées.

L'article [D.312-1-3](#) du Code des relations entre le public et l'administration, pris pour application de l'article L.312-1-2 précédemment cité, qui autorise la publication de certaines catégories de données sans anonymisation préalable **lorsqu'elles sont nécessaires à l'information du public**, octroient également des marges d'appréciation pour la définition et l'application de règles de confidentialité.

Des règles spécifiques de diffusion ont ainsi été édictées pour le recensement de la population, compte tenu de son exhaustivité dans les petites communes. Ces règles de diffusion sont regroupées dans [l'arrêté du 19 juillet 2007](#) relatif à la diffusion des résultats du recensement de la population.

2. Fichiers de données individuelles

- a) Il n'est pas possible de fournir des **données individuelles sur les personnes morales** en respectant strictement l'anonymat, sauf à supprimer du fichier des variables considérées comme essentielles pour l'analyse de tels fichiers : activité économique, localisation, indicateur de la taille.

Les données d'ordre économique et financier collectées par voie d'enquêtes statistiques sont librement accessibles à l'issue d'un délai de vingt-cinq ans suivant la date de réalisation de l'enquête. Pour répondre aux besoins d'informations ponctuelles sur les entreprises avant l'expiration de ce délai, la loi a créé en 1984 un [Comité du secret statistique](#)³. Celui-ci peut proposer des dérogations d'accès aux données individuelles sur les entreprises, en réponse à des demandes effectuées à des fins d'établissement de statistiques ou de recherche scientifique ou historique.

³ Les compétences du Comité du secret statistique, depuis sa création, se sont progressivement étendu à l'ensemble des sources statistiques sur les entreprises et les ménages, ainsi qu'aux sources fiscales, et depuis [la loi pour une République numérique](#) (2016), à toutes les bases de données détenues par des administrations, sous réserve que celles-ci en fassent la demande auprès du Comité.

- b) Les **enquêtes auprès des ménages**⁴ réalisées par l'Insee donnent en général naissance à deux types de fichiers de données individuelles :
- Des fichiers « grand public » mis à disposition sur le site internet de l'Insee. Ces fichiers sont anonymisés, autrement dit construits de telle sorte qu'il n'est pas possible d'identifier, directement ou indirectement, un individu. Pour cela, un certain nombre de variables sont retirées du fichier d'origine : commune de résidence, profession détaillée, voire certaines autres variables propres à l'enquête qui permettraient de reconnaître une personne.
 - Des fichiers dits de « production et de recherche » (FPR). Ces fichiers ne comportent pas d'éléments d'identification et sont traités par des méthodes de regroupement ou de reclassification, de manière à limiter au maximum les risques de réidentification par recombinaison de variables. Ils conservent des informations plus fines que celles qui figurent dans le fichier « grand public » pour répondre aux besoins spécifiques des chercheurs et d'organismes publics comme les services statistiques ministériels. Pour éviter tout risque de rupture de confidentialité par croisement avec d'autres sources de données, la mise à disposition de ces fichiers n'est possible que par l'intermédiaire du comité du secret statistique, selon une procédure similaire, quoique simplifiée, à celle mise en œuvre pour les données individuelles complètes (voir ci-dessous).

Les données relatives à la vie personnelle et familiale, et plus généralement aux faits et comportements d'ordre privé sont librement accessibles à l'issue d'un délai de soixante-quinze ans, ou de vingt-cinq ans après le décès de la personne concernée, si ce délai est plus bref.

Avant l'expiration de ces délais, il est possible, pour des motifs d'établissement de statistiques ou de recherche scientifique ou historique d'accéder aux données individuelles non anonymisées obtenues à partir d'enquêtes auprès des ménages, après avis du [Comité du secret statistique](#) et sur décision de l'administration des archives. La procédure est la même que pour l'accès aux informations individuelles sur les entreprises (voir plus haut).

⁴ Hormis les recensements et les enquêtes annuelles de recensement, qui font l'objet de dispositions spéciales.

L'Insee a fait connaître au comité du secret statistique, lors de sa réunion du 6 octobre 2009, les conditions mises pour son accord⁵ à la communication de données individuelles directement ou indirectement nominatives sur les ménages :

- Pour les services statistiques ministériels (SSM), l'accord est donné sous réserve que le SSM s'engage sur des mesures qui garantissent, au sein de leur ministère, le contrôle de l'accès à ces fichiers et leur utilisation à des fins exclusivement statistiques.
- Pour les autres demandeurs, l'accès aux informations indirectement nominatives s'effectue par le biais du Centre d'accès sécurisé à distance ([CASD](#)). Le CASD, via une infrastructure dédiée hautement sécurisée, permet de mettre à disposition, via Internet, des fichiers de données individuelles en empêchant toute extraction ou édition des données. L'accès à ces postes n'est possible que pour les personnes à qui une autorisation nominative a été accordée. Ces personnes signent toutes obligatoirement un engagement de confidentialité, qui les contraint au respect des règles de sécurité et du secret et les expose à de lourdes sanctions en cas de manquement.
- Le Centre d'accès sécurisé permet aux chercheurs d'accéder aussi bien à des données d'enquêtes ménages que d'enquêtes entreprises, ainsi qu'à des informations d'origine administrative, indirectement nominatives, cédées à l'Insee ou à un SSM au titre de l'article 7 bis de la [loi n° 51-711 du 7 juin 1951](#), et dans les conditions prévues à son article 7 ter.
- L'accès à des informations directement identifiantes est examiné au cas par cas par le comité de direction de l'Insee. L'accès n'est en principe accordé que pour répondre à des demandes de tirage d'échantillons en vue de la réalisation d'enquêtes ayant reçu le visa du ministre prévu par l'article 2 de la loi de 1951.

B. Informations tirées de sources administratives

Les informations transmises à l'Insee ou aux services statistiques ministériels à des fins d'établissement des statistiques, en dehors des enquêtes statistiques, sont également couvertes par le secret statistique. Ceci découle des dispositions de l'article 7 bis de la [loi n° 51-711 du 7 juin 1951](#) sur l'obligation, la coordination et le secret en matière de statistiques.

Les règles du secret qui s'attachent à la diffusion de résultats issus de l'exploitation de ces fichiers peuvent être différentes d'une source à l'autre. Certains producteurs de données définissent des règles qui vont au-delà de ce que demandent les règles du secret statistique (voir plus haut). Ainsi, pour la diffusion de tableaux tirés d'informations fiscales, la règle est qu'aucune case ne doit comporter moins

⁵ Accord requis pour que l'administration des archives puisse délivrer l'autorisation de communication ([article L.213-3 du code du patrimoine](#)).

de onze individus. Pour les tableaux tirés des déclarations annuelles de données sociales (DADS), aucune case ne doit concerner moins de cinq individus.

Il convient donc se renseigner, avant toute publication, sur les règles de diffusion fixées par l'organisme qui a communiqué les informations. En général, ces règles sont inscrites dans la convention qui a permis la transmission des données.

L'accès aux données aux données individuelles issues de sources administratives est interdit sauf dérogations accordées pour des motifs d'établissement de statistiques ou de recherche scientifique ou historique selon des procédures comparables à celles qui permettent l'accès aux données individuelles issues d'enquêtes statistiques. L'avis de l'administration ou de la personne morale ayant procédé à la collecte des données concernées peut être recueilli avant cette communication.

Suite à la [loi n° 2016-1321 du 7 octobre 2016](#) pour une République numérique en son article 36, l'article L.311-8 du code des relations publiques entre le public et l'administration ouvre désormais, en réponse à une demande d'accès à une base de données, à l'administration détenant la base ou à l'administration des Archives, la possibilité de solliciter l'avis du comité du secret. Le demandeur s'engage à respecter la confidentialité des données communiquées et les règles de secret applicables à la source. Le comité du secret peut recommander la mise en place d'un dispositif d'accès sécurisé, à distance ou sur place.

C. Le cas des sources mixtes

Par sources mixtes, on entend :

- Les sources provenant de combinaisons d'enquêtes statistiques et de données administratives ;
- Les 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)

1. Fichiers combinant des données statistiques et des données administratives

Les règles à prendre en considération s'obtiennent par la combinaison des règles applicables à chacun des deux types de sources.

On peut citer parmi ces sources l'enquête Revenus fiscaux et sociaux, qui apparie les résultats de l'enquête emploi, des données fiscales et des données fournies par les Caisses d'allocations familiales, ou bien le dispositif Ésane (Élaboration des statistiques annuelles d'entreprises), qui repose sur la combinaison d'enquêtes statistiques et de données fiscales.

Il est à noter que la diffusion de données individuelles d'origine fiscale est interdite par le livre des procédures fiscales ([article L.103](#)). Des exceptions néanmoins sont prévues. L'[article L.135 D](#) du livre des procédures fiscales prévoit ainsi la possibilité pour les agents du fisc et des douanes de communiquer :

- Toute information aux agents de l'Insee et des SSM, à des fins exclusives d'établissement de statistiques ;
- Les données issues des procédures de contrôle et de recouvrement à des fins de recherche scientifique, dans les conditions de la loi de 1951, c'est-à-dire après avis du comité du secret statistique ;
- Les données d'entreprises, à des fins exclusives de réalisation d'études économiques, aux agents de services de l'État chargés de la réalisation d'études économiques.

2. Fichiers combinant des données sur les entreprises et des données sur les ménages

Ces fichiers peuvent provenir d'enquêtes statistiques (ECMOSS, enquête sur le coût de la main d'œuvre et la structure des salaires) ou de sources administratives (Déclarations annuelles de données sociales).

Les conditions d'accès aux données associent les conditions décrites d'une part pour les données entreprises, d'autre part pour les données ménages.

Annexe

Pourquoi le secret statistique s'applique-t-il aussi aux résultats statistiques agrégés et pas seulement aux données individuelles dont ces résultats sont issus ?

Le secret statistique a essentiellement pour fonction de garantir la confidentialité d'informations individuelles, relatives à des personnes physiques ou morales. L'article 6 de la loi de 1951 interdit ainsi la communication des « renseignements individuels figurant dans les questionnaires » des enquêtes statistiques⁶. Prenant le texte à la lettre, on pourrait penser que le secret statistique n'est alors opposable qu'en cas de communication ou diffusion de données individuelles (quelle que soit leur origine, données d'enquêtes ou de fichiers).

Cette interprétation néanmoins méconnaît le fait que les résultats statistiques, qui, sans constituer des données individuelles, en sont issues, peuvent indirectement, par recoupement avec d'autres informations connues par ailleurs du public, révéler les caractéristiques de personnes.

Une statistique est une valeur numérique ou une mesure qui est utilisée pour résumer, analyser ou interpréter des données dans une population. Une statistique peut se présenter sous de très nombreuses formes : somme, moyenne, pourcentage, taux de croissance, classe de population, etc. Il est bien compris qu'une statistique, comme telle et par elle-même, occulte les informations individuelles dont elle est issue. Dès lors que l'on a calculé le revenu moyen des habitants pour un territoire donné, il n'est plus possible de retrouver, sur la seule base de cette information, le revenu de chacun des habitants de ce territoire. Une statistique, par elle-même, est anonyme.

Cet anonymat peut cependant ne plus être vérifié si l'on prend en considération les informations qui peuvent par ailleurs être accessibles ou disponibles concernant le lien entre des personnes et une statistique.

Prenons trois exemples :

- Le chiffre d'affaires total réalisé par des entreprises d'un secteur donné dans un territoire donné :

Grâce au répertoire des entreprises et des établissements (Sirene), qui est public, on peut connaître quelles sont les entreprises concernées. Si deux entreprises seulement composent la population décrite par cette statistique, chaque entreprise, à partir de la connaissance de son propre chiffre d'affaires, est capable de déduire celui de sa concurrente.

⁶ Sauf décision de l'administration des archives, prise après avis du comité du secret statistique.

- Le taux de croissance annuel du chiffre d'affaires total pour des entreprises d'un secteur donné dans un territoire donné :

Toujours à partir de Sirene, qui diffuse les tranches d'effectifs salariés, il est possible d'estimer la contribution de chaque entreprise concernée au calcul de cette statistique. Pour peu que, sur la base de son effectif salarié, on puisse juger qu'une entreprise prédomine par rapport à ses concurrentes, il sera possible de déduire, avec un bon degré de certitude, que l'évolution mesurée pour le chiffre d'affaires total représente celui de cette seule entreprise.

- La situation au regard du marché du travail (chômage/en emploi/inactif) en fonction de la classe d'âges sur un territoire donné :

Si toutes les personnes de 20 à 24 ans apparaissent comme étant au chômage, et si quelqu'un, parce qu'il la connaît à titre personnel ou en fonction d'informations rendues publiques sur les réseaux sociaux, sait qu'une personne réside dans ce territoire et appartient à cette tranche d'âges, il lui sera possible de déduire que cette personne est au chômage.

En creux, à travers ces trois exemples, on voit que le risque de réidentification de données individuelles est d'autant plus important que la population concernée est peu élevée, raison pour laquelle, et de longue date, le service statistique public a défini des « règles de secret statistique » selon des critères de taille d'effectifs. Dans le cas de la statistique d'entreprises, les deux principales règles appliquées – pas de diffusion d'une statistique si elle concerne moins de trois unités ou si une des unités contribue à plus de 85% de sa valeur – sont ainsi issues de décisions prises dès les années 60. Dans le cas des statistiques relatives aux ménages, l'Insee a mis en place dès les années 80, pour les besoins de la diffusion des résultats du recensement, des règles de diffusion définies en fonction de la population des territoires concernées, règles qui ont été définies après avis de la Cnil.

Ainsi, contrairement à la position exprimée par la Cada (commission d'accès aux documents administratifs) dans [son avis n° 20191797 du 16 janvier 2020](#), des résultats statistiques agrégés peuvent être soumis à une obligation de secret. Cette analyse est corroborée par le Conseil d'État : comme l'indique le rapporteur public dans ses conclusions sur la décision du 7 octobre 1998 (n° 186073), « *il est essentiel de relever [...] que la notion de donnée statistique et celle de donnée nominative ne s'excluent pas nécessairement. Une donnée statistique présentée sous une forme générale et impersonnelle peut devenir une donnée nominative si elle est ramenée à un trop petit groupe de personnes dans des conditions permettant [...] « sous quelque forme que ce soit, directement ou non, l'identification des personnes physiques auxquelles elle (...) s'applique (...) » ».*

Définitions

Renseignement individuel

Est considéré comme renseignement individuel toute information par laquelle il est possible d'identifier, directement ou indirectement, l'unité à laquelle l'information se rapporte (voir plus bas). Cette définition vaut pour toutes les unités statistiques, qu'il s'agisse de personnes morales ou de personnes physiques.

Données à caractère personnel

Ces termes sont définis à l'article 4 du règlement général sur la protection des données ([règlement 2016/679 du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données](#)). On entend par « données à caractère personnel » toute information se rapportant à une personne physique identifiée ou identifiable. Est réputée être une « personne physique identifiable » une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale.

Identification directe ou indirecte

Ces termes sont notamment définis à l'article 3 du règlement [n° 223/2009](#) modifié :

- Identification directe : identification d'une unité statistique à partir de son nom ou de son adresse, ou d'un numéro d'identification accessible au public ;
- Identification indirecte : identification d'une unité statistique par tout autre moyen que l'identification directe.

« Données sensibles »

Les termes de « données sensibles » réfèrent, selon l'article 9 du [règlement général sur la protection des données](#) :

- Aux données à caractère personnel qui révèlent l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale ;
- Aux données génétiques ou biométriques aux fins d'identification unique ;
- Aux données concernant la santé, la vie sexuelle ou l'orientation sexuelle d'une personne physique.



Par principe, le traitement de telles données est interdit. Des dérogations sont néanmoins possibles, notamment lorsque ce traitement est nécessaire à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques.

Les données à caractère personnel relatives aux condamnations pénales, aux infractions ou aux mesures de sûreté connexes relèvent également de dispositions spécifiques, définies à [l'article 46](#) de la loi Informatique et libertés.