

LE MONDE DES DONNÉES



SFJS

LE SECRET STATISTIQUE

Jean-Pierre Le Gléau

 edp sciences

Le secret statistique

Le secret statistique

JEAN-PIERRE LE GLÉAU

Préface de Jean Gaeremynck



17, avenue du Hoggar – P.A. de Courtabœuf
BP 112, 91944 Les Ulis Cedex A

Publications de la SFdS

Collection *Le Monde des données* (EDP Sciences)

Droesbeke J.-J. et Vermandele C. (2018), *Histoire(s) de(s) données numériques*.

Collection des *Journées d'étude en Statistique* (Éditions Technip)

Droesbeke J.-J., Maumy-Bertrand M., Saporta G. et Thomas-Agnan C. Éd. (2014), *Approches statistiques du risque*.

Droesbeke J.-J., Saporta G. et Thomas-Agnan C. Éd. (2015), *Méthodes robustes en statistique*.

Bertrand F., Droesbeke J.-J., Saporta G. et Thomas-Agnan C. Éd. (2017), *Model choice and model aggregation*.

Maumy-Bertrand M., Saporta G. et Thomas-Agnan C. Éd. (2018), *Apprentissage statistique et données massives*.

Collection *La statistique autrement* (Éditions Technip)

Droesbeke J.-J. et Vermandele C. (2016), *Les nombres au quotidien. Leur histoire, leurs usages*.

Ardilly P. et Lavallée P. (2017), *Les sondages pas à pas*.

Collection *Pratique de la statistique* (Presses Universitaires de Rennes)

Husson F. Éd. (2018), *R pour la statistique et la science des données*.

Bécue-Bertaut M. (2018), *Analyse textuelle avec R*.

Genuer R. et Poggi J.-M. (2019), *Forêts aléatoires avec R*.

Composition et mise en pages : Patrick Leleux PAO
Couverture : Conception graphique de B. Defretin, Lisieux

Imprimé en France
ISBN (papier) : 978-2-7598-2332-1
ISBN (ebook) : 978-2-7598-2342-0

Tous droits de traduction, d'adaptation et de reproduction par tous procédés, réservés pour tous pays. La loi du 11 mars 1957 n'autorisant, aux termes des alinéas 2 et 3 de l'article 41, d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective », et d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (alinéa 1^{er} de l'article 40). Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles 425 et suivants du code pénal.

© EDP Sciences, 2019

SOMMAIRE

<i>Préface</i>	7
<i>Introduction</i>	19
1. Qu'est-ce que le secret statistique ?	25
1.1 La loi du 7 juin 1951	25
1.2 Les évolutions de la loi depuis 1951	41
2. Le secret statistique et la diffusion	65
2.1 Diffusion de données agrégées	75
2.2 Diffusion de données individuelles	81
2.3 Diffusion de données géographiques fines	101
2.4 Diffusion des données sur l'environnement.....	105
3. L'accès aux données confidentielles	107
3.1 L'accès aux données couvertes par le secret statistique	108
3.2 L'accès aux données couvertes par le secret fiscal.....	135
3.3 L'accès aux données de santé	146
3.4 L'accès aux données détenues par la Banque de France	152
3.5 L'accès aux autres données	153
3.6 En conclusion	158
4. Comment ça se passe ailleurs ?	161
4.1 Le secret statistique au niveau international	162
4.2 Le secret statistique dans quelques pays.....	166
5. Chronologie	179

6. Principaux textes autour du secret statistique.....	183
6.1 Textes législatifs ou réglementaires – France	183
6.2 Autres documents	189
6.3 Europe et international.....	192
6.4 Sites.....	195
<i>Index</i>	197

PRÉFACE

Le livre de Jean-Pierre Le Gléau a bien des mérites. D'abord celui d'exister, puisque c'est le premier qui soit spécialement consacré au secret statistique, du moins dans la période récente. Ensuite parce qu'il constitue une mine d'informations de toute nature sur les arcanes de la statistique publique. Qu'elles soient historiques, juridiques, administratives, qu'elles décrivent des pratiques ou qu'elles portent des jugements, et l'auteur ne s'en prive pas, les pages qui suivent nous apprennent toujours quelque chose. Et par-dessus le marché, mais en fait pour cette raison même, ce livre donne matière à réfléchir.

Par exemple, il y a un paradoxe apparent à parler de *secret statistique*. En effet, qu'y a-t-il *a priori* de plus étranger au secret que les statistiques ? Leur raison d'être n'est-elle pas d'éclairer, de révéler une vérité cachée sous la surface bouillonnante des faits comme dans la forêt des chiffres ? N'ont-elles pas pour objet, et pour mérite, de dévoiler la réalité de *l'être collectif* que constitue un ensemble d'individus, au-delà de la simple addition des éléments qui le composent ? En ce sens, les statistiques procèdent du besoin de savoir et de la mise en œuvre d'un savoir scientifique, appuyé sur des techniques spécifiques, qui leur confèrent une valeur unique. Que l'on se réfère aux

remarquables productions de l'Insee, à commencer par le recensement de la population, mais aussi à l'enquête logement, au « portrait social » de la France, aux études sur la pauvreté. On pourrait citer également nombre de productions des services statistiques ministériels (SSM), par exemple celles de la Drees sur les dépenses de santé ou de la Dares sur les conditions de travail. On trouve dans tous ces travaux non pas des opinions, ou des appréciations, mais des constats, c'est-à-dire des vérités, qu'on ne trouvera nulle part ailleurs. Les statistiques, c'est l'alliance du miroir, puisqu'elles donnent à voir *ce qui est*, et du rayon laser, car elles franchissent le mur des apparences. À première vue, il y a donc antinomie entre la catégorie du secret et la notion de statistiques.

Mais évidemment, le paradoxe n'est qu'apparent. Car il y a partie liée, manifestement, entre la garantie du secret et la qualité des statistiques. Une vérité solide ne pouvant s'appuyer sur le mensonge ou les approximations, des statistiques de qualité ne peuvent procéder que de matériaux de qualité. Quand ces matériaux sont des réponses à des enquêtes, ce qui est le schéma de base de la loi statistique du 7 juin 1951, il faut que les réponses d'abord existent et ensuite qu'elles soient exactes et sincères. C'est pourquoi ladite loi avait initialement rendu *obligatoires* les réponses aux enquêtes statistiques et qu'elle a par ailleurs garanti le secret à ces réponses. Les enquêtés ne pouvant craindre l'utilisation de leurs réponses à d'autres fins que celles – statistiques – affichées par l'enquête, n'ont ainsi pas de raison de manquer à la sincérité. Ils peuvent avoir confiance, cela leur est garanti par la loi. Il y a donc ici, au cœur de la loi statistique, un cercle vertueux entre la qualité des statistiques, le secret des enquêtes et la confiance des répondants.

Là-dessus, la loi n'a fait que prendre acte, et traduire en droit une réalité structurelle. On voit donc que parler de secret statistique est un raccourci pour désigner le secret qui s'attache aux données utilisées dans le traitement statistique, mais qui ne sont pas encore le résultat statistique, c'est-à-dire *les statistiques*.

Il y a une autre raison d'être, plus formelle, à cette expression de secret statistique. C'est qu'il s'agit d'une modalité particulière du secret, dont le régime est fixé par la loi statistique du 7 juin 1951 : sur ce point, l'intention du législateur, et l'importance qu'il attache au secret, ressortent du titre même de cette loi, « sur l'obligation, la coordination *et le secret* en matière de statistiques ». On observera au passage que l'expression de secret statistique ne se retrouve nulle part dans la loi : ce secret n'est donc pas spécialement défini, il résulte simplement des prescriptions de la loi.

Mais procédant d'une législation spéciale, ce secret se distingue des autres formes particulières de secret protégées par d'autres lois. En matière de secret en effet, la législation française est d'une grande richesse : indépendamment du secret professionnel en général, imposé par le code pénal (article 226-13), on peut citer le secret fiscal, le secret des affaires, le secret médical. On n'ose citer le secret de l'instruction, qui mérite bien peu son nom, ni le secret de la confession, qui évidemment est étranger à la loi républicaine. On observera au passage que la structure (qualité/secret/confiance) mentionnée ci-dessus à propos du secret statistique fonctionne également pour les autres modalités du secret. Une qualité reconnue du système fiscal français, par exemple, est son excellent taux de recouvrement : à quoi l'attribuer, sinon à la confiance des contribuables dans la capacité de l'administration fiscale à garder le secret sur leurs déclarations ?

Si la loi de 1951, on l'a dit, ne définit pas le secret statistique, mais l'organise, elle le fait en distinguant les *finalités*. Ainsi, depuis 1951, la loi interdit expressément l'utilisation des renseignements individuels d'ordre économique et financier collectés (auprès des entreprises) à l'occasion des enquêtes statistiques « à des fins de contrôle fiscal ou de répression économique ». La protection de ce secret résulte donc de l'*indépendance des finalités*. La finalité statistique ne saurait céder devant la finalité fiscale.

C'est donc un édifice extrêmement solide que cette loi de 1951, et le système de protection des données individuelles d'enquête qu'elle a instauré a bien duré plus de trente ans.

Mais il a connu une mutation importante en 1984, comme le montre Jean-Pierre Le Gléau.

La raison en réside, au fond, dans les tensions, voire les contradictions, de la demande sociale. Comme on l'a dit, le secret d'une manière générale est au cœur d'une structure qui relie la qualité de certaines productions, en l'occurrence les productions statistiques, à la préservation de la confiance. Cette équation n'a rien perdu de sa force puisqu'elle repose largement sur le besoin de protection des données individuelles, dit autrement de protection de la vie privée (ou du secret industriel et commercial s'agissant de la vie des entreprises). Mais au cœur des sociétés contemporaines, ce besoin de protection entre aujourd'hui en tension, voire en conflit, avec d'autres exigences, celles de la *transparence*, du refus de l'opacité administrative, du partage des informations. Il y a le besoin d'en savoir toujours plus sur le fonctionnement de la société elle-même, un besoin qui devient nécessité s'agissant du réglage des politiques publiques. La montée de ces exigences est elle-même concomitante avec l'extraordinaire développement de techniques quantitatives appuyé sur l'essor des instruments de calcul et des technologies numériques. On retrouve ici l'attrait du *dévoilement* de la réalité, attribué ci-dessus à la raison d'être des travaux statistiques. Ainsi, le développement des outils vient-il répondre à l'affirmation des nouvelles exigences sociales.

Ce sont ces exigences, combinées au développement de ces outils, qui sont à l'origine du mouvement en faveur des « open data », de la mise à disposition d'informations et de données en ligne, et d'une manière générale de toute l'effervescence autour des « big data ». Tout se passe comme si on ne cessait d'ouvrir des mines d'informations en vue de leur exploitation par les chercheurs plus ou moins *à ciel ouvert*.

On n'en était pas encore là en 1984, et pourtant il devenait indispensable, comme le montre très bien Jean-Pierre Le Gléau, de répondre

dès ce moment à l'appétit des chercheurs, et de concevoir à leur intention des voies d'accès à ces gisements de données que constituent les réponses aux enquêtes statistiques. Il n'était pas question, à cause précisément du secret, de permettre l'exploitation de ces gisements à ciel ouvert. Aussi, et sans qu'il soit besoin à ce stade de modifier la loi sur le secret statistique, ces nouvelles possibilités ont pris la forme, notamment, des « fichiers de production et de recherche » conçus spécialement pour faciliter les travaux statistiques tout en respectant la confidentialité des données. Il s'agit de fichiers constitués à l'aide de procédures de brouillage faisant appel à des modalités d'agrégation. Mais de ce fait, ils n'offrent pas le même intérêt que les réponses individuelles aux enquêtes statistiques. Ces dernières en effet comprennent une grande richesse d'informations, lesquelles peuvent, croisées avec d'autres données concernant les mêmes personnes par le biais des *appariements*, donner lieu à des productions statistiques plus intéressantes.

Il a paru nécessaire alors de prévoir par la loi l'accès des chercheurs à tout le corpus des données individuelles issues des grandes enquêtes statistiques. C'est ce qui a été fait, en deux temps, en 1984 pour l'accès aux données concernant les entreprises, en 2008 pour l'accès aux données concernant les ménages. Dans le même temps, en 1984, était créé le Comité du secret statistique, dont la mission est de se prononcer sur les demandes, présentées le plus souvent par des universitaires, d'accès à ces données.

Ainsi complétée par ces dispositions nouvelles, la loi de 1951, à l'origine intransigeante sur la protection du secret, s'est trouvée singulièrement enrichie, et son équilibre un peu modifié, en ce qu'elle organise cette fois non plus seulement le secret, mais les exceptions au secret que constituent les voies d'accès aux données individuelles d'enquête.

Plus complète, la loi est à l'évidence aussi plus complexe, mais son objet même la place au carrefour d'un ensemble de législations aux adhérences fortes qui produisent ensemble une autre complexité

redoutable. Cela provient du fait que ces législations répondent aux exigences différentes, voire contradictoires, de la demande sociale que nous avons déjà mentionnée. D'un côté, le besoin de protection de la vie privée inspire toute la législation sur la protection des données personnelles, c'est-à-dire la loi « informatique et libertés » de 1978 modifiée en 2004 suite à la directive européenne de 1995, et par une loi récente qui a pour but de la mettre en conformité avec la réglementation européenne qui s'applique désormais directement, à travers le fameux RGPD. Le respect de tout ce corpus est assuré, on le sait, par la Cnil.

D'un autre côté, le besoin d'ouverture a inspiré la loi du 17 juillet 1978 sur l'accès aux documents administratifs, dont les dispositions ont été reversées aujourd'hui dans le code des relations entre l'administration et le public. Dans la période récente, et après les modifications ci-dessus rappelées, de la loi statistique de 1951, il a conduit à modifier en 2013 (loi du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche) la loi fiscale (livre des procédures fiscales, article L.135D) pour prévoir explicitement, et pour la première fois, l'accès des chercheurs aux bases de données de l'administration fiscale. De la même manière, l'approche sectorielle a-t-elle conduit le législateur à prévoir dans la loi santé de 2016 les conditions d'accès à ces immenses bases que constituent les données de santé. Enfin, le même mouvement d'ouverture a généré plusieurs dispositions de la loi « pour une République numérique » du 7 octobre 2016, en particulier celle qui permet d'une manière générale d'accéder aux bases de gestion des organismes administratifs pour les besoins de l'exploitation statistique. Ces dispositions de la loi « pour une République numérique » font elles-mêmes écho à un élargissement considérable, intervenu en 2008 (loi pour la modernisation de l'économie du 4 août 2008), de la définition, incluse dans la loi de 1951, de la notion de statistiques publiques. Il ne s'agit plus seulement des travaux conduits par le service statistique public (Insee et services statistiques ministériels), mais de l'exploitation « à des fins d'information

générale » des données collectées *pour les besoins de leur gestion* par les administrations en général (y compris les organismes privés chargés d'une mission de service public). Cette exploitation est évidemment le fait des détenteurs de ces fichiers eux-mêmes et le législateur, en 2008, en a pris acte, mais la nouveauté provient de ce qu'il a qualifié pour la première fois ces travaux de statistiques publiques. De la même manière qu'en 1984, les détenteurs des réponses aux enquêtes ont perdu le monopole de leur exploitation, le législateur de 2016 demande aux détenteurs des fichiers administratifs d'autoriser, dans certaines conditions, des chercheurs extérieurs à travailler sur ces fichiers. Potentiellement, c'est une petite révolution, et on voit bien comment elle pose dans un contexte nouveau la question traditionnelle du secret.

Faut-il déduire de ce mouvement législatif, accéléré dans la période récente, que le secret statistique se trouve progressivement *démantelé* ? En aucune manière, car les avancées législatives se combinent entre elles pour produire un nouvel équilibre, dans lequel personne n'a envie de sacrifier la protection des données personnelles. Ainsi, les chercheurs qui accèdent aux bases de données restent-ils évidemment soumis à toutes les exigences et formalités de la loi Cnil.

Par ailleurs, les législations « d'ouverture » et leurs textes d'application encadrent sévèrement les conditions d'accès aux bases de données, en réaffirmant les finalités de leur exploitation – selon les cas, la recherche scientifique ou historique, la réalisation de travaux présentant un caractère d'intérêt public, ou toutes finalités équivalentes –, ce qui conduit à réserver l'accès aux bases de données à certains acteurs, ceux qui présentent les meilleures garanties de sérieux. Pour l'essentiel, ce sont des chercheurs, et notamment ceux des laboratoires universitaires.

Il faut dire que depuis le début, les bénéficiaires de l'ouverture, notamment celle opérée en 1984 et en 2008 dans la loi statistique, ont toujours « joué le jeu » et n'ont jamais manqué à l'engagement qui leur était formellement demandé de respecter le secret en s'abstenant

de toute diffusion de données individuelles. Je n'ai pas le souvenir, au cours de près de dix années de présidence du Comité du secret statistique, d'avoir été jamais saisi de faits de manquement à cette obligation de secret, et il semble que l'Insee non plus n'ait pas eu à déplorer de cas de violation du secret.

Cette exigence de respecter le secret est parfaitement intégrée dans les pratiques et cette situation jamais démentie crée à l'évidence des conditions favorables au processus d'élargissement retracé ci-dessus. C'est d'autant plus précieux que les possibilités techniques de traitement des données sont d'une telle puissance qu'elles ne cessent d'accroître les risques d'identification de données même *a priori* anonymisées. La situation de départ de 1951 est celle dans laquelle les réponses aux enquêtes sont toutes par nature des données individuelles, car elles sont produites sous une forme qui permet de les rattacher immédiatement à des individus. Comme le montre Jean-Pierre Le Gléau, les conditions de recueil des réponses aux enquêtes ont changé depuis cette époque et ces réponses ne sont plus collectées sous une forme de face-à-face avec l'enquêté, permettant une identification immédiate. Mais cela ne garantit pas pour autant la protection du secret, du fait des capacités sans cesse accrues des instruments de traitement informatique, qui élargissent considérablement les possibilités techniques d'identification. D'une certaine manière, on peut dire que le secret ne résiste que par le coût des moyens à mettre en œuvre pour le surmonter.

De ce fait, le périmètre des données couvertes par le secret s'en trouve, potentiellement, non pas rétréci, mais élargi.

Mais dans un tel contexte, la technique n'est pas seulement la menace, elle vient elle-même au secours du secret, comme le montre Jean-Pierre Le Gléau dans sa présentation du Centre sécurisé d'accès aux données (CASD). L'apparition, il y a environ huit ans, de ce nouvel outil, conçu au sein du Genes (Groupement des écoles nationales de statistiques), a permis de sécuriser très fortement le système. En permettant la conduite des travaux à distance et en interdisant

matériellement toute captation de données individuelles ou susceptibles de l'être, il impose de lui-même par la technique ce que leur engagement faisait reposer jusqu'ici sur la vertu des chercheurs.

Il résulte de toutes ces évolutions un nouvel équilibre dans lequel le secret n'a rien perdu de sa raison d'être, et dans lequel il ne court pas plus le risque d'être méconnu que par le passé.

Ce nouvel équilibre lui-même permet de focaliser l'attention sur *les enjeux* des évolutions législatives qui viennent d'être retracées et des nouvelles et immenses possibilités qu'elles ont ouvertes aux chercheurs. Ces enjeux tiennent à la multiplication et aux orientations de leurs travaux.

La multiplication est inscrite dans les chiffres (pour ne pas dire les statistiques) relatifs à l'activité du Comité du secret statistique : ces dernières années, le nombre des demandes dont il a été saisi est passé de 228 en 2012 à 300 en 2017. Il n'a pu faire face à cet accroissement qu'en faisant évoluer ses méthodes, en développant le traitement électronique d'un grand nombre de ces demandes et en réservant les séances plénières, devant lesquelles les chercheurs sont invités à présenter leurs projets, aux cas nouveaux ou difficiles.

Mais c'est surtout l'orientation et l'objet des travaux pour lesquels l'intervention du Comité conditionne l'accès aux bases de données de l'Insee (par exemple Esane pour les statistiques d'entreprises, l'échantillon démographique permanent ou encore les déclarations annuelles de données sociales, DADS), des SSM (travail, agriculture, développements durables) et des administrations représentées au Comité (douanes, impôts), qui sont remarquables. C'est évidemment à mettre en relation avec la richesse et la qualité de ces bases, qui se rapportent naturellement aux activités économiques et à la réalité sociale du pays, mais aussi avec la qualité de l'accueil réservée aux porteurs de projets de recherche par les services « producteurs » de l'Insee et des ministères. Ainsi, il n'est pas de session du Comité sans que de jeunes doctorants ne viennent présenter des travaux dont l'objet n'a rien de purement spéculatif, mais qui *a priori* seront

intéressants à prendre en compte dans les domaines du travail, du logement, du commerce extérieur, etc., qui recouvrent largement les politiques publiques.

Il n'est pas de séance non plus que le Comité, qui a d'ailleurs le droit de se les voir communiquer, ne presse ces chercheurs de lui faire connaître les résultats de leurs travaux et les invite à leur donner l'audience la plus large possible. Il y a là tout un potentiel qui soulève à nouveau la question récurrente du lien entre les travaux scientifiques et l'éclairage des circuits de décision politiques.

Si la question est récurrente, on observera qu'elle est renouvelée aujourd'hui par l'importance prise dans le discours des autorités politiques par la thématique de *l'évaluation*. Cette importance n'a d'ailleurs d'égale que le flottement conceptuel qui l'entoure et la faiblesse de l'organisation administrative. De quelle évaluation parle-t-on ? Qui en est le responsable ? Y a-t-il un socle méthodologique clairement défini ? Ces questions n'ont pas à ce jour reçu de réponse claire, mais elles n'en cessent pas moins de se poser.

Si vraiment il existe une volonté forte de progresser sur ce thème, on n'échappera pas à une réflexion sur le rôle de la statistique dans le dispositif à mettre en place. Mais s'il existe, comme nous le pensons, de nouvelles possibilités de faire participer les chercheurs à cette entreprise d'évaluation, et par là même, à la définition et au réglage des politiques publiques, cela suppose qu'un dialogue utile s'engage avec eux, pour définir les conditions de leur participation au débat public qui préservent leur indépendance et prennent en considération leurs besoins propres. Les institutions de la statistique publique – Conseil national de l'information statistique, Autorité de la statistique publique, Comité du secret statistique – sont bien placées pour avoir une vue d'ensemble du jeu des acteurs : détenteurs des bases de données, producteurs de travaux statistiques, universités et organismes de recherche, responsables de la définition et de la conduite des politiques publiques. Au-delà de leur activité quotidienne, elles devraient se saisir de ce sujet comme d'une priorité majeure.

Tout progrès en ce sens ne fera que prendre acte d'une réalité que des décennies de pratique du secret statistique ont bien mise en évidence : au-delà d'une prescription en forme d'interdiction, la règle du secret est la condition de production d'éclairages décisifs sur la vie économique et sociale, qui comptent autant pour la pertinence des politiques publiques que pour la qualité du débat démocratique.

Jean Gaeremynck
Président de section au Conseil d'État
Ancien président du Comité du secret statistique (2009-2018)

INTRODUCTION

Secret statistique ? Mais les statistiques ne sont-elles pas par définition des agrégats anonymes dans lesquels on ne peut pas isoler la valeur se rattachant à un individu ? Alors, de quel secret s'agit-il ?

Il est bien vrai que l'immense majorité des statistiques ne donnent que des résultats agrégés, où il est impossible d'identifier un individu. Mais, il est aussi possible de produire des résultats très détaillés, donc sur une population peu nombreuse. Avec un risque de pouvoir identifier une personne. De leur côté, les utilisateurs de la statistique publique demandent justement des informations de plus en plus précises, allant parfois jusqu'au niveau individuel, non pas pour savoir quelle valeur prend la variable pour une personne donnée, mais pour mettre en œuvre des techniques sophistiquées, nécessitant une information pour chaque individu : analyse des données, suivi temporel des personnes, appariements de fichiers, etc.

La réponse à ces besoins ne saurait se faire au détriment de la protection légitime des données individuelles, qu'il s'agisse de protéger la vie privée des personnes ou, pour les données sur les entreprises, de sauvegarder le secret des affaires. La problématique du secret statistique répond justement à ce besoin de trouver un équilibre entre d'une part le service rendu par la production et la diffusion de

statistiques et d'autre part la nécessaire protection de ceux qui ont fourni les informations servant à les construire.

La protection des données individuelles constitue un véritable enjeu en ce début de XXI^e siècle. La France avait fait figure de pionnière en adoptant dès 1978 la « loi informatique et libertés ». Mais les digues ainsi construites se sont avérées parfois fragiles. Certains s'affranchissent allègrement des contraintes. Ou bien profitent de failles dans le réseau législatif, puisque la protection de la vie privée, pour être efficace, doit se déployer au niveau de la planète. Des utilisations peu scrupuleuses de données, comme celles qui ont mis récemment en scène Facebook et Cambridge Analytica, en constituent une partie émergée, bien loin, hélas, d'être des cas uniques en leur genre !

L'Union européenne préparait déjà depuis plusieurs années des mesures permettant de contrecarrer ce type de contournement de la loi : cela s'est traduit par le règlement général sur la protection des données (RGPD), adopté par le Parlement européen et le Conseil le 27 avril 2016 et qui est applicable depuis le 25 mai 2018. Ce règlement élève un certain nombre de barrières contre l'usage indélicat de données à caractère personnel, tout en augmentant très fortement les pénalités en cas d'infraction. Mais, s'adressant à l'ensemble des données personnelles pouvant faire l'objet d'un traitement, il ne peut entrer dans les détails pour chaque type de données. De plus, chaque pays ayant une législation et une culture propres, une loi a été nécessaire en France pour adapter ce règlement aux spécificités de la situation française. Cette loi a été adoptée le 20 juin 2018.

L'avenir dira si ces mesures auront été efficaces, sans être excessives, pour protéger les individus d'un mauvais usage des innombrables données collectées par des milliers d'opérateurs, à l'occasion de nombreux actes de la vie quotidienne.

Mais, parmi toutes ces données, il en existe qui sont collectées spécifiquement par la puissance publique, pour ses besoins propres : ouvrir des droits sociaux, collecter des impôts, rembourser des prestations ou... établir des statistiques !

PMSI (programme de médicalisation des systèmes d'information) 84, 85, 147, 148

profilage 77, 79

R

recensement 28, 36, 41 et sq., 54, 59, 61, 62, 78, 79, 81, 93, 94, 100, 101, 103, 117, 118, 119, 167, 176, 177

regroupements 77, 80, 81, 87, 88, 90 et sq., 104, 105, 113, 114

remote access 122, 171, 173

remote execution 121, 171

RGPD (règlement général sur la protection des données) 20, 84, 146

Royaume-Uni 23, 37, 93, 120, 166, 172, 173

Russie 23, 166, 173, 175, 177

S

sanctions 25, 26, 27, 39, 42, 112, 145, 155, 157, 165

santé 23, 57, 97, 108, 117, 118, 146 et sq., 156 et sq., 167

SD-Box 127

secret des affaires 19, 144, 164

serment 63, 169

Sirene (Système informatique pour le répertoire des entreprises et des établissements) 29, 45, 49, 50

SIREN (Système d'identification du répertoire des entreprises) 27, 96, 128, 145

Sniiram (Système national d'information interrégimes de l'assurance maladie) 95, 147

sous-traitance 134

SSM (services statistiques ministériels) 22, 25, 51, 57, 58, 62, 98, 112, 115, 117, 130, 135 et sq., 159, 164, 165

T

t-proximité 90

Triris 80, 81

Tunisie 23, 166, 177, 178

U

Union européenne 20, 37, 66, 105, 113, 162, 163

V

vie privée 19, 20, 21, 30, 78, 79, 81, 111, 130, 135, 144, 155, 164, 166

LE SECRET STATISTIQUE

Pour établir leurs statistiques, l'Insee et les autres services statistiques publics récoltent de nombreuses informations auprès de nous. Ces informations sont-elles bien protégées ? Finalement, qui peut y avoir accès ? Le secret statistique garantit que nos données ne seront pas divulguées à des tiers. Les règles de mise en œuvre du secret statistique se sont adaptées aux mouvements de la société et à l'évolution des techniques. Il est désormais possible de donner accès à des informations statistiques très détaillées pour les chercheurs, sans rompre la confidentialité. Ces techniques, mises au point pour la statistique, sont maintenant utilisées pour mettre à la disposition des chercheurs des données fiscales, sur la santé, etc. tout en respectant la protection des données personnelles.

Cet ouvrage est le premier à faire un bilan complet des règles et pratiques qui permettent de garantir le secret statistique. Une première partie définit l'évolution du secret statistique depuis la guerre. Un chapitre pratique détaille les règles à observer dans la diffusion des statistiques pour préserver la confidentialité des données. Un développement particulier est ensuite fait sur la question difficile de la diffusion de données individuelles. Enfin quelques exemples pris dans d'autres pays donnent un éclairage sur les techniques retenues ailleurs.

Jean-Pierre Le Gléau est inspecteur général honoraire de l'Insee et membre de l'Institut international de statistique. Il a participé à l'élaboration des textes législatifs visant à adapter le secret statistique aux évolutions de la société contemporaine. Il a accompagné la création du centre d'accès sécurisé aux données (CASD). Il participe aux activités du groupe « Statistique et enjeux publics » de la Société Française de Statistique.

ISBN : 978-2-7598-2332-1



9 782759 823321

La Société Française de Statistique (SFDS), association reconnue d'utilité publique, a pour objectif de favoriser les développements de la statistique et d'assurer la représentation de l'ensemble des utilisateurs, enseignants et chercheurs dans ce domaine. La collection « Le monde des données » a pour ambition de proposer des ouvrages accessibles à un très large public, témoignant ainsi de sa volonté de s'ouvrir davantage vers la société civile.

SFDS

edpsciences
www.edpsciences.org