

## Des modifications juridiques pour accompagner les évolutions du numérique

L'ouverture des données publiques pour les différents publics et leur utilisation par la statistique publique progresse grâce aux modifications des lois qui encadrent l'accès à ces données. De même l'utilisation des données privées par la statistique publique nécessite une évolution législative. L'actualité récente est riche. Le projet de loi pour une république numérique, la loi portant modernisation du système de santé et le règlement européen sur les données à caractère personnel sont des preuves que l'ouverture des données est en marche. Dans ce contexte d'utilisation croissante des données, les règles qui assurent la protection de l'individu sont cruciales et font l'objet de recherches permanentes.

*Michel Isnard, responsable de l'unité affaires juridiques et contentieuses*

### Loi pour une république numérique

Pour accompagner le développement de l'économie numérique, le gouvernement a proposé un projet de loi pour une république numérique qui couvre un vaste champ organisé en trois volets : la circulation des données et du savoir, la protection des droits dans la société numérique et l'accès au numérique.

Le texte du projet de loi a fait l'objet d'une consultation le 26 septembre 2015 puis a été déposé le 9 décembre 2015 à l'assemblée. Il devrait être adopté par le parlement d'ici fin septembre.

Le service statistique public est concerné par plusieurs sujets :

- L'**article 1** instaure la gratuité des échanges de données entre les administrations de l'État et les établissements publics. Cela devrait concerner les données que nous recevons de l'**IGN, Acoss, Pôle Emploi, Cnav**.
- L'**article 7bis** indique que la réutilisation des informations publiques produites par le service statistique public ne peut donner lieu au versement d'une redevance. **Les données Sirene** seront en conséquence **diffusées gratuitement** à partir du 1<sup>er</sup> janvier 2017 avec une compensation budgétaire.
- L'**article 12** modifie la **loi du 7 juin 1951** pour introduire un accès aux données privées. Il permet en effet de demander une transmission électronique des données détenues par des personnes morales privées sur décision du ministre chargé de l'économie, après une étude de faisabilité et d'opportunité, une concertation avec les personnes morales privées et un avis du Cnis. Dans le cas d'une non transmission, une amende est prévue après avis du comité du contentieux, d'un montant plus élevé que les montants en vigueur actuellement pour les enquêtes (25 000 € pour une première amende et 50 000 € pour une récidive). Cet article permettra d'asseoir la **production de l'indice des prix à la consommation** à partir des **données de caisse**, et potentiellement à l'avenir d'autres productions ([voir groupe Insee-Cnis](#)).
- L'**article 18** assouplit les conditions dans lesquelles des appariements entre bases de données sur la base d'un **identifiant dérivé du NIR**, mais non signifiant (obtenu par une technique de hachage), pourront être réalisés : les appariements réalisés pour la statistique publique feront l'objet d'une déclaration à la Cnil, et ceux réalisés pour la recherche publique relèveront du régime d'autorisation, alors que les appariements sur la base du NIR requièrent jusqu'à présent un décret en Conseil d'État. Ces

dispositions nécessitent des précautions pour la mise en œuvre pratique.

**Pour en savoir plus :**

<http://www.senat.fr/dossier-legislatif/pjl15-325.html> et [Appariements sécurisés et statistique \(2000-2011\) : une décennie d'expériences, Courrier des statistiques n° 131, septembre 2011.](#)

### Loi santé

Les données de santé ont toujours fait l'objet de dispositions spécifiques en raison de leur sensibilité. La loi du 26/01/2016 encadre l'ouverture des données de santé à travers une modification du **Code de la Santé publique** et de la **Loi informatique et libertés du 6 janvier 1978**.

Dans le code de la santé publique, la loi crée un Système national des données de santé (SNDS) qui rassemble et met à disposition : les données de l'assurance maladie (le SNIIRAM), les données hospitalières (le PMSI) et les causes médicales de décès, auxquelles devraient s'ajouter par la suite les données sur le handicap et un échantillon de données des organismes d'assurance maladie complémentaire. La CNAMTS réunit et organise l'ensemble de ces composantes pour les mettre à disposition dans le cadre de grandes orientations définies par l'État.

Le texte de loi clarifie et homogénéise les règles d'accès à ces données. Il instaure une gouvernance unifiée, en créant l'Institut National des Données de Santé (INDS), ainsi que le Comité d'Expertise pour les Recherches, les Études et les Évaluations (CEREES) dans le domaine de la santé. L'INDS est un groupement d'intérêt public composé de représentants de l'État, des utilisateurs et des producteurs de données, des acteurs et des usagers du système de santé. Son rôle est de veiller à la qualité des données et de promouvoir leur utilisation. Il assure un secrétariat unique des demandes d'accès au SNDS, et émet si cela est nécessaire un avis sur le caractère d'intérêt public que présente un projet. Quant au CEREES, il est composé d'experts chargés de se prononcer sur la pertinence et la qualité du protocole d'étude et la nécessité d'avoir recours aux données SNDS demandées pour réaliser le projet.

En dehors des données anonymes qui peuvent être mises à disposition librement et gratuitement, l'accès au SNDS est restreint à des projets de recherche, d'étude ou d'évaluation présentant un caractère d'intérêt public. Pour l'accès et l'utilisation des données à caractère personnel présentant un risque fort d'atteinte à la vie privée, l'avis de l'INDS et du CEREES intervient avant que la CNIL décide *in fine* des autorisations de traitement.

Pour les données présentant un risque plus faible il est prévu une procédure d'accès simplifiée, préalablement homologuée par la CNIL.

La loi santé fait également évoluer la loi Informatique et libertés pour simplifier les démarches à engager vis-à-vis de la CNIL pour tout projet mené à des fins de recherche, d'étude ou d'évaluation ; ainsi pour tout appariement avec le NIR, il n'est plus exigé un décret en Conseil d'État mais une autorisation de traitement.

Les textes d'application devraient paraître avant la fin du 1<sup>er</sup> trimestre 2017.

#### **Pour en savoir plus :**

<https://casd.eu/event/2016/presentations>

[Rapport de la Commission open data en santé, juillet 2014](#)

[Autorité de la statistique publique, séance du 22 juin 2016](#)

### **Règlement européen sur la protection des données individuelles**

Le Parlement européen a adopté le 14 avril 2016 le règlement européen sur la protection des données qui met à jour et modernise les principes énoncés dans la directive de 1995 sur la protection des données. Il définit les droits des personnes physiques et fixe les obligations des personnes qui effectuent le traitement des données et de celles qui sont responsables de ce traitement. Il définit également les méthodes visant à assurer le respect des dispositions prévues ainsi que l'étendue des sanctions imposées à ceux qui enfreignent les règles.

Il sera applicable en 2018 dans tous les pays membres de l'Union européenne pour les données détenues par les pays membres. Il modifiera les dispositions en vigueur en France comme la loi de 1978. Les discussions sont en cours sur les modalités d'application. La Cnil a lancé le 16 juin une [consultation pour construire le mode d'emploi](#).

### **Confidentialité et Big Data : la quadrature du cercle**

La multiplication des données à disposition des instituts de statistique publique se conjugue avec une volonté accrue des institutions nationales et européennes de diffuser plus de résultats, plus fréquemment et avec plus de détails. En parallèle, il est important de s'assurer du respect de la confidentialité pour les individus concernés par ces données.

L'anonymisation directe des données est une condition nécessaire mais non suffisante pour le respect de la confidentialité. AOL l'a appris à ses dépens en diffusant en 2006 des données de connexion Internet non nominatives mais ayant permis au [New York Times de ré-identifier l'internaute 4417749](#) en combinant plusieurs recherches effectuées par celle-ci.

L'une des difficultés de la gestion de la confidentialité des données est de déterminer un compromis entre utilité des informations diffusées et protection de la vie privée des répondants. Ce compromis n'est pas sans lien avec la sensibilité des données diffusées. Une personne souhaitant ré-identifier un individu au sein d'une base de données profite de la multiplication des données à sa disposition. L'augmentation exponentielle des données disponibles rend d'autant plus complexe l'assurance du respect de la confidentialité. Par exemple, [de Montjoye et al. \(2013\)](#) ont montré qu'en utilisant des données de téléphonie mobile, il est possible d'identifier 95 % des individus de la base testée dès lors qu'on connaît 4 points localisation/horaire par utilisateur, avec une précision temporelle horaire et une localisation donnée au niveau de l'antenne téléphonique.

La protection de la confidentialité des données dans le contexte d'un institut de statistique se gère différemment selon le type de diffusion envisagé - résultats agrégés, données cartographiées,

données individuelles, etc. - et le canal de diffusion : disponible pour tout public ou données réservées à des utilisateurs accrédités. Elle repose sur des techniques différentes pour mettre en place la diffusion : anonymisation des jeux de données, restrictions des accès à des utilisateurs.

Le sujet de la protection de la vie privée est un sujet de recherche vaste et très actif. La Cnil et l'Inria ont ainsi lancé un prix européen pour encourager la recherche scientifique sur la protection de la vie privée. Dans le système statistique européen, un [centre d'excellence sur la confidentialité des données](#) auquel participe l'Insee travaille sur l'harmonisation des pratiques pour la gestion de la confidentialité, et notamment à l'amélioration des outils utilisés par les instituts statistiques.

#### **Pour en savoir plus :**

[de Montjoye : Computational privacy » ou comment le comportement humain limite les possibilités d'anonymisation conférence Casd 6 avril Dossiers Solidarité et Santé n° 64, Juillet 2015.](#)

### **Actualités / Brèves**

Le centre d'accès sécurisé aux données (CASD) a organisé le 6 avril dernier une [conférence](#) dédiée aux enjeux de l'accès sécurisé aux données pour la recherche scientifique et la science des données, ouverte par Axelle Lemaire, Secrétaire d'État chargée du numérique.

Pauline Givord (DMCSI) a présenté l'état des réflexions aux [48e journées de la Statistique de la SFdS](#) qui ont eu lieu du 30 mai au 3 juin (voir [site intranet division MAEE](#) ou [groupe Yammer BigData et datascience](#)).

Une session sur « Big Data et nowcasting macroeconomic indicators » était organisée par Eurostat à la 1<sup>ère</sup> [« conference on Advanced Research Methods and Analytics »](#). Stéphanie Combes (DMCSI) a présenté une étude sur l'intégration de séries issues de Google Trends pour améliorer la prévision économique, dans le prolongement de [Bortoli et Combes \(2015\)](#).

Le [rapport du groupe Insee-Cnis](#) sur « [La réutilisation par le système statistique public des informations des entreprises](#) » présidé par M. Bon a été diffusé officiellement fin juin.

Un colloque « [Big Data en santé : quels usages, quelles solutions](#) » organisé par la Drees a eu lieu le 4 juillet 2016.

Une rencontre d'échanges avec les opérateurs de la téléphonie mobile est organisée au niveau européen les 22 et 23 septembre prochain dans le cadre de l'ESSnet Big Data.

Un workshop sur les avancées sur les sujets juridiques, éthiques, les aspects de communication et sur les compétences requises est organisé par [Eurostat à Ljubljana les 13 et 14 octobre](#) dans le cadre de l'ESSnet qui travaille spécifiquement sur ces sujets.

Sur la base de l'article 12 du projet de loi « numérique » (voir plus haut), l'étude de faisabilité et d'opportunité pour l'utilisation des données de caisse est en cours. Elle devrait être présentée pour approbation au bureau du Cnis d'ici la fin de l'année.

[Ont participé à ce numéro Maxime Bergeat, Françoise Dupont, Michel Isnard, Annie Henrion, Stéphanie Combes et Pauline Givord. Cette lettre est une occasion d'informer largement et d'échanger. N'hésitez pas à transmettre vos réactions et suggestions d'articles à Stéphanie Combes : \[stephanie.combes@insee.fr\]\(mailto:stephanie.combes@insee.fr\)](#)

Les archives de cette lettre sont disponibles sur :

<http://www.agora.insee.fr/jahia/Jahia/site/dmcsi/SiteDMCSI/DMSaccueil/DMAEEaccueil/BigData>

[Demande d'inscription individuelle à la lettre : \[dg75-1101@insee.fr\]\(mailto:dg75-1101@insee.fr\)](#)