

DOSSIER DONNÉES PERSONNELLES : QUI A ACCÈS À QUOI ? #CHERCHEURS

ENTREPRISES, CHERCHEURS, ÉTAT... DE NOMBREUX ACTEURS RÉCUPÈRENT ET UTILISENT NOS DONNÉES PERSONNELLES. MAIS QUELLES SONT LES RÈGLES EN LA MATIÈRE ? LE CITOYEN-CONSOMMATEUR PEUT-IL GARDER LA MAIN SUR SES PROPRES DONNÉES ?

CHERCHEURS : FACILITER L'ACCÈS POUR BOOSTER LA SCIENCE

Grâce à des avancées législatives et technologiques, l'accès des scientifiques aux données que l'État collecte sur les citoyens commence à se déverrouiller. À la clef : la promesse d'une recherche plus performante. Décryptage.

Revenus et patrimoine, état et dépenses de santé, parcours scolaire et professionnel, indemnités chômage, taxe d'habitation, allocations familiales... Les services de l'État possèdent une foule d'informations très précises sur chaque citoyen. Une véritable mine d'or lorgnée par les chercheurs pour booster leurs études dans de nombreux domaines : épidémiologie, sociologie, économie, santé publique, etc.

“Mais pour garantir la confidentialité et le respect des libertés individuelles, cette mine d'or est protégée par des dispositifs de loi rendant l'accès à ces données extrêmement restrictif”, estime Kamel Gadouche, qui dirige le Centre d'accès sécurisé aux données (CASD). Les chercheurs doivent notamment soumettre un dossier au comité du secret statistique pour obtenir une autorisation de l'administration des archives, et réaliser des démarches préalables auprès de la Commission nationale de l'informatique et des libertés (CNIL). Au total, obtenir une telle autorisation peut prendre jusqu'à un an !

DES AVANCÉES SIGNIFICATIVES

Toutefois, les choses commencent à bouger. Venant tout juste d'entrer en vigueur, la Loi pour une République Numérique élargit par exemple l'accès aux données très détaillées à potentiellement toutes les sources de données de l'administration. Concrètement, ce sont des données dont on n'a retiré aucun détail, hormis le nom, le prénom et l'adresse exacte remplacée par une indication géographique assez précise.

“Ce nouveau texte offre aussi aux chercheurs de nouvelles possibilités de croisement de données, à partir d'un identifiant unique et fiable : le numéro de sécurité sociale”, ajoute Kamel Gadouche.

En 2016, la Loi de modernisation du système santé a aussi ouvert aux chercheurs l'accès au Système national des données de santé (SNDS) à des fins de recherche, d'étude ou d'évaluation d'intérêt public. Issu de la base de données de la Caisse Nationale d'Assurance Maladie (CNAM), l'une des plus grandes au monde en

santé, ce SNDS regroupe les données anonymisées issues de 1,2 milliard de feuilles de soins, de 500 millions d'actes médicaux et de 11 millions d'hospitalisations annuelles !

L'ANONYMISATION FREINE LES RECHERCHES

Mais avant de pouvoir en profiter, ces données doivent être anonymisées. Or, cela s'avère problématique pour les chercheurs. Car anonymiser ce n'est pas juste enlever le nom de la personne, mais toutes les données qui, croisées entre elles, pourraient permettre de l'identifier. *“À l'ère du numérique, la donnée personnelle a changé de nature : des informations non directement identifiantes peuvent le devenir si on les agrège à d'autres, confirme Francesca Musiani, chercheuse à l'Institut des Sciences de la Communication du CNRS. D'où l'importance de considérer le contexte d'accès et non pas seulement la donnée en elle-même.”*

Et ce ne sont pas là que des suppositions... Un exemple ? Juste en croisant diverses sources de données publiques, des chercheurs sont déjà parvenus à retrouver l'identité de donneurs anonymes de prélèvements génétiques. La mise en évidence de cette faille avait même fait l'objet d'une publication dans la revue *Science*¹. *“Le problème c'est qu'en anonymisant trop les données, on gomme beaucoup d'informations, précise toutefois Emmanuel Bacry qui coordonne l'Initiative “Data Science” à l'École Polytechnique. Or c'est précisément en ouvrant les données dans leur granularité la plus fine qu'on peut obtenir les meilleurs résultats.”* Pas facile donc de trouver le bon équilibre entre recherche de qualité et protection de la vie privée...

ACCÈS SÉCURISÉ

Autre danger du traitement de données personnelles par des chercheurs : le risque de piratage. Voilà pourquoi certains scientifiques souhaitant travailler sur des données sensibles ont des locaux dédiés au sein même de l'administration. C'est d'ailleurs le cas de l'équipe d'Emmanuel Bacry qui possède ses propres bureaux à la CNAM.

Mais tous les chercheurs n'ont pas cette possibilité... Heureusement pour eux, les technologies progressent.

“Aujourd'hui, plus de 1500 chercheurs et datascientists en Europe utilisent le Centre d'accès sécurisé aux données (CASD) pour traiter des données individuelles très détaillées de l'administration directement ou indirectement identifiantes”, indique Kamel Gadouche.

Le concept ? Ces données restent confinées dans des locaux techniques du CASD baptisés “la bulle”; pour y pénétrer à distance et accéder à des outils pour les traiter, tout chercheur doit utiliser un boîtier d'authentification par empreintes digitales nommé “SD-Box”. Résultat : ce système de bulle sécurisée hermétique crée une isolation totale du boîtier qui fonctionne donc en circuit fermé, sans aucun contact avec l'extérieur. Des procédures qui peuvent paraître contraignantes, mais c'est sans doute le prix à payer pour garantir le respect de notre vie privée.

NOTES

¹ M. Gymrek et al., Science, 339, 321, 2013

Retrouvez la suite dossier sur les Données personnelles :

[Citoyens : la voie du “self data”](#)

Termes liés :

- [Données](#)
- [Data science](#)