

L'IMPOSSIBLE ANONYMAT

« DES DONNÉES PARTAGEABLES PLUTÔT QU'OUVERTES »

CONVAINCU DES BIENFAITS POTENTIELS DE L'OPEN DATA, CHRISTIAN LOVIS, PROFESSEUR À LA FACULTÉ DE MÉDECINE, DEMEURE NÉANMOINS PRUDENT. LE MONDE DANS LEQUEL LES DONNÉES MÉDICALES SERONT UN JOUR GRATUITEMENT DISPONIBLES N'EST PAS PEUPLÉ QUE DE SCIENTIFIQUES HUMANISTES.

« *Il ne faut pas être naïf: les données sont le pétrole et l'intelligence artificielle est le moteur à combustion de notre époque.* » Une combinaison qui sent fort l'Eldorado et qui pousse Christian Lovis,

professeur au Département de radiologie et informatique (Faculté de médecine), à prôner la prudence vis-à-vis de l'Open Data, c'est-à-dire de la mise à disposition de tous des données de la recherche, telle que le demande la nouvelle directive du Fonds national pour la recherche scientifique (FNS, lire en page 20).

« *J'ai longtemps été un fervent partisan, voire même un évangéliste, de l'Open Data, poursuit-il. Avec les années et l'évolution en cours, toutefois, ma position s'est nuancée. Je ne mets pas en cause les bénéfices potentiels de l'ouverture des données pour la société. Au contraire, je trouve cela très bien. Mais il faut être conscient que le monde n'est pas peuplé que de scientifiques bienveillants désireux d'utiliser les travaux d'autres scientifiques afin d'accroître les connaissances de l'humanité. Il y a quantité d'autres acteurs dans ce jeu, certains très puissants, et les règles dominantes qui gèrent l'utilisation des données ne sont pas celles de la recherche scientifique, désintéressées et respectueuses de la vie privée. Ce sont celles du marché, compétitives et basées sur le profit.* »

Pour le chercheur, le problème c'est que, selon l'esprit de l'Open Data, tout le monde peut avoir accès aux données de la recherche, gratuitement et sans restriction. Il existe des règles qui peuvent être mises en œuvre pour rendre les données accessibles. Mais rien ne permet d'en assurer le respect en cas d'usage commercial, discriminatoire ou même criminel.

D'ailleurs, les principaux détenteurs de données aujourd'hui ne sont pas les instituts de recherche mais les géants d'Internet (les fameux GAFAs pour Google, Apple, Facebook et Amazon) et les légions de start-up qui gravitent

autour. Les ordinateurs, les téléphones portables et leurs applications ainsi que tous les objets (voitures, aspirateurs, réfrigérateurs...) connectés produisent des informations sur les habitudes de consommation en ligne, la navigation sur Internet, les déplacements physiques, les paramètres de santé, etc. Ces sociétés, dont les plus grosses disposent de dizaines voire de centaines de milliards de dollars, possèdent également les outils de l'intelligence artificielle capables de donner du sens et, surtout, de la valeur à ces monceaux de données. Et l'usage qui en est fait n'est pas seulement scientifique, loin de là.

Suisse convoitée Dans ce monde des données reines, celles issues de la recherche médicale sont parmi les plus convoitées (on estime qu'environ un quart des fuites de données concerne le domaine de la santé). Et en particulier celles de Suisse. « *Les données de santé produites en Suisse ont des propriétés exceptionnelles, note Christian Lovis. Il n'y a que cinq hôpitaux universitaires sur le territoire, ce qui*

fait que l'homogénéité des informations médicales sur la population est unique, sans parler de leur densité. Pour certaines compagnies privées, c'est un rêve de pouvoir accéder à cette mine de données et de les combiner avec celles d'un opérateur de télécommunications presque unique, d'un quasi-monopole de seulement deux grands acteurs dans la grande distribution, d'un seul gros distributeur pour les médicaments, etc. »

Dans le cas particulier de la médecine, il existe des garde-fous, dont l'anonymisation des patients enrôlés dans les études médicales. La nouvelle directive du FNS, loin d'être dogmatique, prévoit d'ailleurs la possibilité de conserver la confidentialité des données dans certains cas, et en particulier si l'anonymat ne peut pas être assuré. Cette disposition ne rassure pas Christian Lovis pour autant.



Christian Lovis

Professeur au Département de radiologie et informatique

ON ESTIME QU'ENVIRON UN QUART DES FUITES DE DONNÉES CONCERNE LE DOMAINE DE LA SANTÉ.

«Il faut distinguer l'anonymat de la «dé-identification», explique le chercheur genevois. Anonymiser, c'est comme flouter la photographie d'un visage pour le rendre méconnaissable. Dé-identifier, c'est se borner à enlever le nom de la personne représentée. Dans le cas d'un dossier médical, on a beau «flouter» tout ce que l'on veut, si l'on désire conserver les informations nécessaires à la recherche scientifique – ce qui est tout de même le but –, il restera toujours assez de paramètres pour retrouver le patient. En d'autres termes, il est possible de dé-identifier les données médicales individuelles, mais pas de les anonymiser.»

Un article paru en 2013 dans le *Journal of the American Medical Informatics Association* montre ainsi que quatre mesures du taux de cholestérol dans le sang sont suffisantes pour identifier un patient sur 100 000 de manière unique. Cela ne signifie pas encore que l'on dispose de son nom, mais d'une sorte de profil ou d'«empreinte digitale» qui lui est propre. Une compagnie d'assurances, qui connaît les dates des examens médicaux des affiliés, ou tout autre détail exploitable, peut ainsi facilement retrouver dans ses bases de données les individus qui ont été enrôlés dans telle ou telle étude.

Privacy by design Pour Christian Lovis, il semble dès lors préférable de parler de dé-identification et d'assumer, en toute transparence, l'impossibilité pratique de mettre en place des méthodes fiables et robustes d'anonymisation non réversibles.

«Il n'y a pas d'autres solutions, à mon avis, que de développer des approches que l'on appelle Privacy by Design, ou confidentialité dès la conception, explique Christian Lovis. C'est également ce que prône le Swiss Personalized Health Network (SPHN)**, à l'origine duquel se trouvent des chercheurs des Écoles polytechniques fédérales, des universités et des hôpitaux universitaires suisses.»

Le SPHN, dirigé par l'Association suisse des sciences médicales et le FNS, est une initiative nationale qui cherche à promouvoir la médecine et la santé personnalisées en Suisse. L'idée est de mettre sur pied une infrastructure décentralisée

«DANS LE CAS D'UN DOSSIER MÉDICAL, ON A BEAU 'FLOUTER' TOUT CE QUE L'ON VEUT, IL RESTERA TOUJOURS ASSEZ DE PARAMÈTRES POUR RETROUVER LE PATIENT.»

et protégée permettant l'utilisation des quantités énormes de données de santé pour la recherche et l'innovation. Les données seraient ainsi partageables, sous conditions, plutôt que partagées sans condition, ajoute Christian Lovis. Pour protéger les données des patients, ce dernier préconise aussi d'expérimenter des solutions encore innovantes comme celles des *blockchains* (à l'origine de la monnaie virtuelle *Bitcoin*) ou des *hashgraph* (une sorte de *blockchain* perfectionnée) qui sont des technologies informatiques décentralisées permettant d'effectuer, entre autres, des transactions de manière très sûre et anonyme.

«Il y a vingt ans, le débat sur l'ouverture des données n'existait tout simplement pas, conclut Christian Lovis. Les dossiers des patients étaient, en quelque sorte, conservés dans des coffres-forts. Aujourd'hui, les données ont une valeur, elles se monnaient. On trouve des dossiers médicaux en vente sur le Dark Web. Les économistes débattent sur la manière dont il faut définir la donnée: est-ce du capital? Du travail? Une marchandise? Dans ce contexte un peu sauvage, je ne demande pas de revenir aux coffres-forts d'antan. Mais je défends au moins la notion de données «partageables» plutôt que simplement ouvertes. Cela inclut une forme de gouvernance, de contrôle qui me semble indispensable.»

* Consulter par exemple: opengovdata.org/

ou le site de la Commission européenne: goo.gl/szEwBn

** Swiss Personalized Health Network: <https://goo.gl/YpDGdL>