

L'Open Data Santé : rendre publiques les données de la sécurité sociale

Après le phénomène de Big Data, c'est désormais celui d'Open Data qui fait son apparition. Le Big Data fait référence au mouvement de numérisation de nos données qui est en œuvre depuis quelques décennies. Depuis le début de l'ère de la numérisation dans les années 2000, notre capacité de stockage des données a explosé et le développement du Big Data franchi une étape supplémentaire. Aujourd'hui, au cœur des enjeux de transparence liés au contexte de modernisation de l'action publique, c'est plus particulièrement l'*ouverture* des données publiques – l'Open Data – qui est en cours d'évolution. En témoigne la création prochaine d'un poste d'administrateur général des données qui sera « *autorisé à connaître les données détenues par l'administration de l'État et ses opérateurs¹* ».

Venu des Etats-Unis et du Royaume-Uni, qui se sont engagés dans le processus d'ouverture des données en 2009 et 2010 respectivement, ce mouvement a pris de l'ampleur. En témoigne la Charte en faveur de l'ouverture des données publiques, signée en juin 2013 par les membres du G8. En France, la mise en place, sous Nicolas Sarkozy, de la mission Etalab, chargée de créer une plateforme des données publiques, a donné une impulsion au processus d'Open Data. A titre d'exemple, BPI France a récemment ouvert quatorze années de données statistiques sur les PME avec le lancement de son think-tank « Lab ». La Saône et Loire et la ville de Paris ont aussi mené leurs premières expérimentations avec les plateformes Open Data 71 qui donne accès à des millions d'informations publiques, et ParisData qui rassemble des projets axés sur la réutilisation de données ouvertes. Mapize, par exemple, permet de visualiser le nombre de naissances, décès et mariages par arrondissements et par années.

La France accuse, néanmoins, un retard certain en matière de santé. L'industrie de la santé repose sur quatre sources de données : les données de R&D d'entreprises pharmaceutiques, les données cliniques des hôpitaux, les rapports des caisses d'assurance maladie et les données détenues par divers acteurs exclus du secteur industriel, le patient lui-même par exemple². Les expériences internationales ont fait la démonstration des bénéfices directs et indirects liés à l'ouverture des données publiques de santé. En France, la mission Etalab a fait un pas en avant en publiant récemment une cartographie des données publiques disponibles dans le domaine de la santé, recensant plus de 260 bases ou jeux de données³. Toutefois, les blocages ne

¹ Communiqué de presse, Ministère de la Réforme de l'Etat, 21 mai 2014.

² Policy Paper, « Faire entrer la France dans la troisième révolution industrielle : le pari de l'innovation », Institut de l'Entreprise, mai 2014.

³ Plateforme du gouvernement data.gouv.fr

sont pas encore levés, et les données « publiques » restent encore très confidentielles...

Si la santé est un bien public, n'est-ce pas le rôle de l'Etat que de mettre gratuitement à la disposition de tous les données recueillies ? Il nous faut évoluer vers un système de santé « intelligent », défini selon un arbitrage équilibré entre impératif économique, sécurité sanitaire et respect du droit à la vie privée.

Une tendance internationale à l'Open Data : des enjeux sanitaires et de croissance

Les Etats-Unis et le Royaume-Uni, ainsi que les pays scandinaves, ont sauté le pas de l'ouverture des données. Par exemple, les patients américains peuvent désormais avoir des informations sur ce que les hôpitaux facturent⁴ : sur le site data.gov, on peut recueillir des informations sur le coût d'un soin d'une fracture de la hanche et du pelvis au St John's Hospital dans l'état de New York⁵. Par ailleurs, les expériences concluantes menées en Europe du Nord et au Canada sont nombreuses.

L'ouverture des données publiques en matière de santé constitue d'abord un enjeu sanitaire essentiel. A quoi ressemblerait le futur si le progrès médical ne reposait plus exclusivement sur les découvertes de laborantins mais bien davantage sur l'exploitation systématique de l'information en temps réel ?

Du point de vue de la recherche, accéder librement aux données de santé ouvre un champ d'analyse considérable pour les professionnels de la santé. En rendant possible un suivi précis et en temps réel de la consommation des médicaments, l'amélioration de la surveillance épidémiologique – qui se fera désormais à l'échelle d'une population entière et non plus sur des échantillons restreints de populations – permettra de détecter plus tôt les dérives d'usage d'un médicament et assurera une meilleure transparence en pharmacovigilance. En France, l'Institut de veille sanitaire (InVS) a commencé à développer un réseau de surveillance basé sur des structures capables de fournir des informations en continu sur l'état de santé de la population. Le réseau OSCOUR recueille les données des quelque 600 unités d'urgence existantes et les transmet automatiquement au service d'urgence de l'InVS. Cette collecte des données, dans le but d'améliorer la surveillance, devrait être menée à plus large échelle. Au Royaume-Uni, la mise en place en 2012 du Clinical Practice Research Datalink (CPRD) – qui croise de multiples données économiques, profil du patient, ses motifs de consultation, etc. récoltées grâce à l'Open Data – a permis de construire un système performant et efficace pour la recherche médicale. Aux Etats-Unis, le

⁴ <http://www.fondapol.org/politique-2-0/sante-open-data-pour-un-meilleur-rapport-qualite-prix/>

⁵ <https://data.cms.gov/Medicare/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3>

système du Group Health Cooperative (GHC), par exemple, permet d'obtenir des informations sur deux millions de sujets et de remonter jusqu'en 1980 !

Du point de vue du patient, l'ouverture des données représente une véritable mine d'informations et la possibilité de mieux se faire soigner. Nous pourrions d'abord disposer de renseignements comparés sur les prestations des hôpitaux, celles des médecins, sur les différents traitements... En gardant à l'esprit que tout ne peut pas être comparable – le taux de mortalité d'un hôpital doté d'un service de réanimation ne peut pas être comparé avec celui d'un hôpital sans ce type de service – l'accès aux données permettra une meilleure visibilité. En France, à l'heure actuelle, un patient est dans l'impossibilité de savoir, après un séjour à l'hôpital, ce qui est facturé à l'assurance et ce qui est à ses frais. De même, il n'est pas possible de connaître les tarifs exercés par chaque hôpital pour une prestation spécifique⁶. Autant de paramètres qui permettraient, grâce à l'accès à l'information, une meilleure sélection des soins.

L'ouverture des données de santé constitue enfin un enjeu de croissance. Aux Etats-Unis, l'ouverture des données a facilité la recherche génomique – qui est l'étude du génome par le biais des maladies qui nous affectent – et aurait permis de créer 310 000 emplois dans ce domaine et de rapporter 20 milliards de dollars de ressources nettes à l'Etat. Libérer l'accès aux données influe directement sur la création d'emplois en favorisant l'innovation et en favorisant la création de start-ups. Le site internet américain mydrugcost.com propose des applications en ligne sur le coût des médicaments à partir de données qui proviennent de la Bibliothèque nationale de médecine. Au Royaume-Uni, l'ouverture des données a permis l'éclosion d'applications iPhone comme *Genes in Space* qui repose sur le traitement délocalisé de données : en jouant sur leur Smartphone, les utilisateurs de l'application contribuent à la recherche contre le cancer en aidant les scientifiques à analyser des données en temps réel. L'Open Data a également contribué à attirer des chercheurs et à faciliter la concurrence entre acteurs publics et privés pour conduire des analyses. En France, il est clair que l'Open Data permettrait par ailleurs de mieux débusquer les inefficiences du système actuel et de mettre fin à de nombreux gaspillages de la sécurité sociale... Potentiel économique important, créateur d'emplois et de services, générateur d'économies publiques, l'Open Data santé peut ainsi participer, à son échelle, à améliorer l'activité économique et à mettre la France à niveau dans la compétition internationale.

Le blocage français : l'encadrement strict de l'usage des données et l'étendard du respect de la vie privée

⁶ <http://www.fondapol.org/politique-2-0/sante-open-data-pour-un-meilleur-rapport-qualite-prix/>

En France, le droit d'accès aux documents a vu le jour en 1978 avec la loi sur l'accès aux documents administratifs. Depuis, l'influence européenne et la prise de conscience politique ont fait progresser ce droit. En 2011, la mission Etalab puis le portail national data.gouv.fr qui permet l'accès à plus de 355 000 informations publiques gratuites et réutilisables⁷ ont vu le jour. Toutefois, **l'usage des données reste extrêmement encadré.** La collecte de données est autorisée par la loi Informatique et Libertés dans des cas strictement définis (article 8 et 9). Les données médicales concernant les patients et les professionnels de la santé ne peuvent faire l'objet de cession ou d'exploitation, et même le consentement du patient ne saurait effacer le secret professionnel.

Par ailleurs, toutes les données ne sont pas accessibles. L'ensemble des données de la Caisse nationale d'Assurance maladie (CNAM) est stocké dans le SNIIRAM, la base centrale pour les données de santé individuelles. Détenteur de données anonymes issues des remboursements effectués par l'ensemble des régimes d'assurance maladie⁸, le SNIIRAM représente un potentiel d'informations considérable. Par an, le SNIIRAM regroupe 1,2 milliard de « feuilles de soins », 500 millions d'actes médicaux et 11 millions de séjours hospitaliers (en médecine, chirurgie et obstétrique), avec une potentielle capacité de mémoire de 14 ans⁹. Mais la complexité de l'architecture des données rend leur utilisation difficile et l'extraction des données du SNIIRAM doit suivre un processus précis avant d'être validée par la CNIL, en coordination avec l'Institut des Données de Santé¹⁰. Hormis quelques exceptions près – dont bénéficient par exemple la Haute Autorité de Santé ou l'Institut de Veille Sanitaire depuis un arrêté du 19 juillet 2013¹¹ – l'accès à ces données est extrêmement restreint et l'opacité règne. En témoigne l'affaire Celtipharm.

Celtipharm est une société statistique de traitement des données pour le compte de l'Agence nationale de sécurité du médicament (ANSM) et d'industriels. Afin de pouvoir analyser plus précisément la consommation de médicaments, Celtipharm a développé, il y a plus de trois ans une technologie d'analyse en temps réel. Celtipharm souhaite traiter le flux de données (anonymisées) que les pharmacies transmettent aux assurances maladies pour les demandes de remboursements. Les données anonymisées étant 'codées', Celtipharm requiert l'accès au système permettant de les décoder pour pouvoir identifier les médicaments et produire des statistiques. Depuis deux ans et demi et en dépit de l'autorisation de la Commission nationale de l'informatique et des libertés (CNIL), le projet est bloqué. La CNIL elle-même a fait l'objet d'une attaque en référé au Conseil d'Etat pour avoir donné cette

⁷ http://www.etalab.gouv.fr/pages/Qui_sommes_nous_-5883786.html

⁸ Cnam-TS, MSA, RSI et les seize autres régimes spéciaux.

⁹ BRAS Pierre-Louis, Rapport sur « La gouvernance et l'utilisation des données de santé », septembre 2013.

¹⁰ Voir en annexe le schéma de l'articulation des compétences des différents organismes (source : Institut des Données de Santé).

¹¹ <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000027830713>

autorisation. La décision des juges du Conseil d'Etat est tombée mardi 27 mai 2014 : les autorisations de la CNIL d'exploiter des feuilles de soin électroniques anonymisées ont été jugées légales. Les prochains mois nous diront si cette décision va permettre d'ouvrir une brèche concernant l'accès aux données de la CNAM.

La position de la CNAM pose en effet un problème du point de vue juridique.

D'une part, l'article L1110-1 du Code de santé publique prévoit que « le droit fondamental à la protection de la santé doit être mis en œuvre par tous moyens disponibles » et que les organismes d'assurance maladie doivent assurer « la meilleure sécurité sanitaire possible ». D'autre part, la France, en refusant l'accès libre et gratuit aux données de santé, ignore l'esprit du principe de droit communautaire de réutilisation libre et gratuite des données publiques. Issu de la directive européenne 2003/98/CE, ce principe fut transposé en France par l'ordonnance du 6 juin 2005. Si la présente directive ne contient aucune obligation de publicité – la décision d'autoriser ou non la réutilisation est laissée à l'appréciation des États membres – elle vise toutefois à faciliter la réutilisation des données détenues par des organismes du secteur public des Etats-membres (article 1). La CNAM, dans son refus de coopérer, ignore ce principe.

Par ailleurs, la demande citoyenne d'accès aux données de santé est réel.

Du point de vue sanitaire, on recense chaque année en France 18 000 morts causées par le mauvais usage du médicament¹². Du point de vue des consommateurs, outre le répertoire AMELI qui constitue un premier embryon d'Open Data en répertoriant des professionnels de la santé, nous disposons de (trop) peu d'informations pour choisir un médecin ou un établissement de soins. L'opinion publique semble néanmoins de plus en plus réceptive à ce problème de l'accès des données santé. L'avis de la conférence de citoyens mise en place par l'Institut Montaigne montre que les Français souhaitent plus de transparence dans le domaine de la santé¹³ et plaident en faveur de l'Open Data sur les données concernant les coûts, le fonctionnement et les résultats de notre système de santé. Les citoyens souhaitent pouvoir accéder facilement aux informations relatives à la performance des médecins et des établissements de santé publics et privés pour pouvoir orienter leur choix. Selon un récent sondage Ipsos, neuf Français sur dix considèrent que leur médecin traitant leur prescrit trop de médicaments ! Ce sont donc des outils permettant l'aide à l'interprétation de données brutes qu'il faut donc développer. Pour répondre à cette demande, le site La Fourmi, comparateur de tarifs médicaux, avait vu le jour en 2011. Mise en demeure par la CNAM, sous prétexte que ces données ne pouvaient être utilisées sans l'accord de

¹² http://www.lemonde.fr/sante/article/2013/05/27/les-medicaments-causent-au-moins-18-000-morts-en-france_3418273_1651302.html

¹³ Avis de la conférence de citoyens, 9 décembre 2012, Institut Montaigne.

chacun des professionnels concernés, les dirigeants de La Fourmi se sont vus contraints de fermer leur site¹⁴...

Dans le monde médical, seuls les syndicats de médecins se montrent réticents à la logique d'Open Data. Ils invoquent la protection de la vie privée mais craignent tout simplement la mise en concurrence. Sans transparence, les comportements prescripteurs à risque restent ignorés¹⁵... Toutefois, l'intervention croissante de la société civile dans le débat public devrait, à court terme, accélérer le processus. En témoigne, en vue de la future loi de santé publique, le rapport « Pour l'an II de la démocratie sanitaire » remis à Marisol Touraine, Ministre des Affaires Sociales et de la Santé, en février 2014 qui prend acte de la volonté du citoyen d'être au cœur des décisions de santé grâce au développement d'outils de participation directe des usagers.

Organiser l'Open Data : (i) garantir l'anonymisation des données et (ii) le droit de propriété des patients sur leurs données personnelles

Depuis la remise du rapport « La gouvernance et l'utilisation des données de santé » à Marisol Touraine, en octobre 2013, le débat sur l'Open Data est rouvert.

Outre le coût de mise en œuvre très élevé de l'Open Data et le risque d'utilisation biaisée des données par les laboratoires, le rapport pointe du doigt **le risque de ré-identification à partir des données, portant de facto atteinte à la vie privée**¹⁶. Selon le Ministère de la Santé, l'exploitation des données ne devrait donc être envisagée que « *dans un cadre respectueux du secret médical et des règles garantissant l'anonymat. Seules les données non sensibles de ce double point de vue [auraient] vocation à être mises à disposition de la collectivité* ». Car malgré le progrès constant du digital, l'imperfection des systèmes informatiques demeure et le risque d'une dérive « orwellienne » inquiète. Dans les scénarii extrêmes liés à l'Open Data, banquiers et assureurs moduleraient leurs offres en fonction de notre état de santé, dont ils seraient parfaitement informés. Censés être les bénéficiaires de l'Open Data, les citoyens deviendraient donc les victimes d'un système qui les dépasse.

¹⁴ <http://www.lequotidiendumedecin.fr/actualite/securete-sociale/interdit-par-la-cnam-un-site-comparateur-de-tarifs-medicaux-s-adresse-au->

¹⁵ <http://www.eric-verhaeghe.fr/mais-qui-bloque-lopendata-sante-en-france/>

¹⁶ BRAS Pierre-Louis, Rapport sur « La gouvernance et l'utilisation des données de santé », septembre 2013.

Ce discours alarmiste, récurrent dans l'administration, semble toutefois cacher d'autres raisons. Selon Didier Sicard, ancien président du Comité consultatif national d'éthique, le gouvernement craint que l'ouverture des données ne lève le voile sur une mauvaise gestion des dépenses de santé¹⁷. La mainmise de l'Etat sur les données de santé relève avant tout d'un réflexe de protection. En ce qui concerne les obstacles objectifs et rationnels que sont l'anonymisation des données et de la protection de la vie privée, il existe en France suffisamment de garanties pour y faire face sereinement et permettre à l'Open Data de devenir la règle et la confidentialité des données l'exception.

En France, le respect de la privée (i) est inscrit à l'article 9 du Code civil depuis la loi du 17 juillet 1970. Depuis, la loi Informatique et Libertés du 6 janvier 1978 est venue rappeler que les données de santé constituent des données sensibles dont le traitement et la collecte sont en principe interdits (article 8) et a souhaité renforcer la garantie de protection de la vie privée en créant la CNIL. **Le respect de la vie privée est donc assuré par l'anonymisation des données dont la CNIL est la garante** : si la CNIL distribue des autorisations pour pouvoir collecter des données, elle peut aussi les retirer.

Ce processus d'anonymisation des données est possible dans le domaine de la santé, comme il a déjà eu lieu dans les pays pratiquant l'Open Data. En France, l'ouverture des données du PMSI (programme de médicalisation des systèmes d'information) – l'hébergeur des données des hôpitaux – avait suscité le même débat : aucun scandale lié au respect de la vie privée n'est à déplorer depuis son ouverture. En pratique, les données recueillies par les hôpitaux sont soumises au secret médical et sont transmises sous forme de fichiers anonymisés aux Agences de Santé Régionales, qui les transmettent à leur tour à l'agence chargée de la constitution de bases de données nationales¹⁸.

D'autre part, les données de santé des patients leur appartiennent (ii). Ce droit de propriété n'est cependant pas garanti en France puisque nos données sont détenues par l'Assurance Maladie. C'est pourtant aux patients que devrait revenir la possibilité de décider si oui ou non ils souhaitent partager leurs données. C'est le cas au Royaume-Uni. Si en temps normal, le médecin généraliste transmet au HSCIC (Health and Social Care Information Centre) les données du patient, le patient peut choisir de ne pas communiquer ses données personnelles en remplissant un « opt out form ».

¹⁷ <http://www.opendatasante.com/2013/05/14/pour-didier-sicard-la-ministre-de-la-sante-ne-souhaite-pas-ouvrir-les-donnees-entrevue/>

¹⁸ « Pour une meilleure utilisation des bases de données nationales pour la santé publique et la recherche », Haut Conseil de la Santé Publique, mars 2012

Trois propositions pour la mise en œuvre pratique de l'Open Data en France

Ainsi, afin de permettre une utilisation effective des données de santé, respectueuse des droits de chacun, GenerationLibre émet les trois propositions suivantes :

1. **Faire de la transparence la règle, et de la confidentialité l'exception. Il faut déverrouiller l'accès en temps réel aux données brutes du SNIIRAM (base centrale pour les données de l'Assurance Maladie), du GIE SESAM-Vitale (programme de dématérialisation des feuilles de soin pour l'Assurance Maladie) et de l'ASIP (Agence des Systèmes d'Information Partagée de santé).** Si l'accès à ces trois opérateurs est primordial, la liste pourrait être étendue aux (nombreux) autres acteurs de la santé qui participent à en faire un domaine fermé : ATIH, HAS, INVS, DREES, INPES, IRDES, ANSM... A terme, il faudrait créer **une plateforme commune d'accès unique à l'ensemble des données.**
2. **Renforcer les pouvoirs de la CNIL pour (i) autoriser, au cas par cas, le retrait de certaines données sensibles de la base publique et (ii) veiller à l'anonymisation des données.**
3. **Fournir aux patients des « opt-out forms ».** Sur le modèle de ce qui est en place au Royaume Uni¹⁹, chaque patient devrait pouvoir décider de ne pas partager ses informations personnelles en remplissant un formulaire à remettre à son médecin généraliste.

¹⁹ www.connectingforhealth.nhs.uk/systemsandservices/scr/staff/aboutscr/comms/pip/optout.pdf

ANNEXE

Articulation des compétences des organismes donnant un avis sur les demandes nécessitant une extraction particulière de données issues du SNIIRAM (cas de l'offre non standard)

