



HAL
open science

L'accès aux données de santé : débats et controverses en France

Joumana Boustany, Gabriella Salzano, Christian Bourret

► To cite this version:

Joumana Boustany, Gabriella Salzano, Christian Bourret. L'accès aux données de santé : débats et controverses en France. L'ère du numérique. Enjeux des données massives, ISTE, pp. 207-225, 2017, 978-1784052881. hal-04024482

HAL Id: hal-04024482

<https://hal.science/hal-04024482>

Submitted on 31 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'accès aux données de santé : débats et controverses en France

Joumana Boustany, Gabriella Salzano, Christian Bourret

Résumé. *L'ouverture et la réutilisation des données de santé offrent des opportunités considérables pour l'enrichissement des connaissances et pour la recherche dans le domaine de la santé publique, de la gestion, et des sciences sociales. Afin de faire avancer l'état des savoirs, les chercheurs doivent pouvoir accéder et réutiliser les données recensées dans plusieurs systèmes d'information (SI) de santé. En France, le débat autour de l'ouverture de ces données est fortement controversé du fait que les informations contenues dans ces systèmes sont privées et hautement sensibles. Ce chapitre examine l'état des pratiques concernant les données de santé ouvertes en France d'un point de vue critique. Nous avons observé les principaux SI de santé afin de construire une synthèse sur le plan d'organisation, d'accessibilité et des possibilités de croisement et de réutilisation de ces données.*

1. Introduction

Aujourd'hui, chaque pays possède sa propre approche concernant l'ouverture des données de santé résultant de contextes historiques et de mentalités collectives parfois très différentes. Les données de santé sont très sensibles, avec des spécificités fortes en particulier relatives la vie privée des patients et et la confidentialité des diagnostics. D'une part, une mauvaise manipulation des données médicales peuvent avoir des conséquences extrêmement graves. D'autre part, et comme spécifié par le « Health and Social Care Centre » les données ouvertes de santé apportent de multiples avantages comme « augmenter le choix des patients et améliorer leur santé, offrir une productivité accrue, contribuer à la croissance économique » (Health & Social Care Information Center 2015). Ces données accélèrent la recherche scientifique et ont une valeur économique élevée. Selon l'estimation du McKinsey Global Institute (Manyika et al. 2013), « l'utilisation des données ouvertes et propriétaires dans le domaine des soins de santé pourraient aider à générer une valeur de 300 à 450 milliards de dollars par année aux États-Unis. La majeure partie de cette valeur se présente sous la forme d'économie de dépenses pour les fournisseurs, les payeurs et les patients ».

Le système de sécurité sociale étant en crise avec un déficit de 3,9 milliards pour 2015 en France (Godeluck 2016), l'ouverture des données pourrait aider à redresser partiellement ces comptes. Ce chapitre, basé sur une revue de la littérature ainsi que sur l'analyse de quelques systèmes d'informations majeurs de santé, étudie les conditions d'accès à cette forme de données publiques de santé en France.

2. Méthodologie

Pour traiter cette question, nous avons adopté une approche empirique qui est appropriée pour un sujet d'étude en pleine expansion comme c'est le cas avec les données ouvertes et les données massives de santé. Aujourd'hui, avec les objets connectés, Internet, les réseaux sociaux, etc., la quantité de données collectées s'est démultipliée à tel point que « les données produites en deux jours correspondent à ce que l'humanité a produit depuis ses origines jusqu'en 2003 »¹. Afin de retracer le progrès de l'implémentation des données de santé en France, nous avons donc exploré les archives des journaux français, les plateformes de données officielles, les sites web gouvernementaux ainsi que les sites de la Haute Autorité de santé (HAS). Cette méthode a permis d'appréhender les points de vue des différentes parties prenantes concernant les décisions publiques relatives aux systèmes d'information de santé et de soins. Pour l'analyse des bases de données, nous avons pris en considération des systèmes

¹ Eric Schmidt: Every 2 Days We Create As Much Information As We Did Up To 2003. Techcrunch, (2010). <https://techcrunch.com/2010/08/04/schmidt-data/>

« opérationnelles » avec une valeur ajoutée pour la recherche dans différents domaines (santé publique, sciences sociales, etc.). Nous avons exclu les bases de données privées comme la « Longitudinal Patient database » acquise récemment par imshealth.com ou Xpr-SO regroupant « le panel [en] temps réel de 3 004 pharmacies, représentatif des 22 458 pharmacies de France métropolitaine »². Ces bases de données ainsi que la plupart des données massives obtenues via les objets connectés ne sont pas accessibles ou payantes ; de ce fait il est difficile de les analyser !

Afin de réaliser cette étude, nous avons eu recours à des analyses tant quantitatives que qualitatives. Les données qualitatives ont été obtenues par l'analyse de deux types de systèmes : ceux considérés par les spécialistes comme fondamentaux pour mesurer les indicateurs de santé, générant un grand nombre de données, mais dont l'accès est restreint ou interdit. À noter qu'un nombre restreint de ces systèmes sont ouverts à l'exception des plus récents qui ont été conçus pour être accessibles à tous.

3. Revue de la littérature

Depuis 2007, une série d'actions en faveur de l'ouverture des données de santé a été entreprise au niveau international. Mais, « la France reste encore très en retard, par rapport aux pays d'Europe du Nord et d'Amérique du Nord, vis-à-vis du croisement des sources de données (absence de base de données renseignant sur les motifs d'usage et les principales caractéristiques permettant de juger du bien-fondé d'une prescription) et, surtout, du partenariat avec le monde de la recherche » (Bégaud et Costagliola 2013). L'adoption généralisée du *dossier médical personnel* a accéléré la collecte de données cliniques personnelles et sensibles comme l'historique médical, les résultats des examens cliniques et des examens de laboratoire, la lecture et l'interprétation de l'imagerie médicale, etc. La réutilisation de ces données est essentielle pour la recherche, car elle permet de générer de nouvelles connaissances et vient en appui à la prise de décision stratégique et opérationnelle.

Plusieurs auteurs ont souligné les bénéfices du partage des données de la recherche : « Les arguments pratiques et scientifiques qui militent en faveur du partage des données comprennent l'amélioration de la précision dans la recherche, l'analyse des risques et des bénéfices des options de traitement, le renforcement des collaborations, l'accélération des recherches biomédicales et la restauration de la confiance dans la recherche clinique »³ (Olson et Downey 2013). D'autres soulignent également les entraves à l'ouverture des données de santé : les différences entre les patients, les données ne sont pas structurées de la même manière, la vie privée, la conversion des données d'autant plus que chaque organisation collecte généralement ses informations dans des bases de données différentes et pour des finalités particulières. De plus, les obstacles majeurs se situent également au niveau technique. Ils portent spécifiquement sur l'hétérogénéité des organisations, des données et des standards qui sont un des critères essentiels pour la réutilisation des données. En effet, « des développements parallèles non normatifs dans de multiples secteurs entraînent une duplication substantielle des coûts et de l'effort humain » (Geissbuhler et al. 2013).

La littérature sur les systèmes d'information de santé en France dévoile un dispositif important qui aurait besoin d'améliorations substantielles : « Les différentes bases de données nationales présentent un intérêt majeur pour le système statistique de la santé en France. Cependant, cet intérêt est actuellement limité en raison des obstacles liés à leur utilisation à des fins de santé publique »⁴ (Goldberg et al. 2012).

² <https://www.xpr-so.net/pages/Public/Accueil.aspx>

³ The practical and scientific arguments for data sharing include improving the accuracy of research, informing risk/benefit analysis of treatment options, strengthening collaborations, accelerating biomedical research, and restoring trust in the clinical research enterprise.

⁴ «The very large national databases present a major interest for the French health statistics system. However, the interest is currently limited because there are obstacles regarding their utilisation for public health purposes».

Dans des rapports commandités par différentes administrations françaises, des experts pointent l'abondance des informations publiques qui n'ont aucune cohérence de formats (Picard et al. 2012) et recommandent d'améliorer la visibilité de l'offre d'information sur la santé en créant un portail institutionnel unique (Brun et al. 2011). Ils se sont interrogés quant à l'accessibilité de l'information de santé en France par les citoyens. Les systèmes d'information de santé français offrent une information très dispersée, inintelligible par les citoyens. Le rapport sur la surveillance et la promotion du bon usage du médicament en France fait une lecture terrifiante sur les structures françaises de surveillance de la consommation de médicaments (Bégaud et Costagliola 2013).

De nombreuses voix se sont élevées en faveur de l'ouverture des données en France. En septembre 2013, un rapport « Gouvernance et utilisation des données de santé » a été soumis à Marisol Touraine, Ministre des Affaires sociales et de la Santé. Ce rapport prône une ouverture des données de santé, mais uniquement lorsqu'il n'y a pas de risque de réidentification des personnes (Bras 2013). Il est important de noter que la vie privée est théoriquement protégée en France par la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. En octobre 2013, l'initiative *transparence santé*, association dissoute depuis le 1^{er} octobre 2014, a lancé une pétition pour l'ouverture des données. En novembre 2013, durant la présentation de la stratégie nationale de santé, la ministra avait réaffirmé son engagement, a créé les conditions nécessaires pour une ouverture des données avec un accès sécurisé aux données de santé et a lancé un débat public sur le sujet. Cet engagement n'a guère eu d'effet sur le terrain puisque trois ans après des spécialistes de renom réclament toujours l'ouverture des données de santé. Citons Didier Sicard (Professeur de médecine, président d'honneur du Comité consultatif national d'éthique) et Jean-Marie Spaeth (président honoraire de la CNAMTS et président honoraire de l'École nationale supérieure de sécurité sociale) qui, dans un article du monde du 27 septembre 2016, ont réclamé l'ouverture des données de santé dans ces termes : « Notre pays dispose de données très précises sur la santé. Malheureusement, il est difficile d'y avoir accès, en dépit de ce qu'en dit l'État. Cela doit changer, le débat démocratique en dépend ».

Un des rapports au Sénat sur l'*open data* a souligné le fait que « l'identification de l'information recherchée apparaît souvent difficile ; l'information disponible est lacunaire et sa qualité est inégale (en raison, notamment, de la tardiveté des mises à jour, d'une certaine instabilité méthodologique ou de sa grande technicité) ; enfin, elle ne répond pas nécessairement aux attentes des citoyens ». (Bouchoux 2014)

Les années passent, les écrits et les rapports se succèdent pour *dénoncer les freins à l'ouverture des données*. Dans son rapport du 3 mai 2016 relatif aux données personnelles de santé gérées par l'Assurance maladie, la Cour des comptes a considéré qu'il était prioritaire de sécuriser l'existant, d'améliorer la gouvernance et d'assurer la fluidité des accès selon une approche à la fois rigoureuse et ouverte, afin d'encourager l'utilisation des données de santé à des fins d'intérêt général. Ce rapport (Cours des comptes 2016) regrette que malgré une base de données de santé étendue, la France affiche un retard considérable lorsqu'il s'agit de faciliter leur accès et recommande d'améliorer la qualité des données dans un objectif de construction d'un *open data* santé et de simplifier les procédures devant la CNIL.

Comme le montre cette revue de la littérature, les acteurs politiques français manifestent leur enthousiasme à l'idée d'ouvrir des données sur la santé. Le nombre de rapports commandités par le ministère en atteste, mais les résultats sont lents à apparaître. Ce constat est toujours valable en 2017. Un étrange mouvement d'un pas en avant et deux pas en arrière semble caractériser le débat sur les données ouvertes touchant à la santé en France.

4. Les systèmes d'Information de santé en France

Plusieurs régimes pour la gestion des systèmes de santé coexistent en France. Par exemple le régime général de la sécurité sociale, le régime agricole, le régime social des travailleurs indépendants (RSI), etc. De ce fait, il existe plusieurs systèmes d'information de santé avec des objectifs et des contenus différents qui rendent difficile le croisement des données. L'*Institut des Données de Santé (IDS)* a été créé avec la « mission d'assurer la cohérence et de veiller à la qualité des systèmes d'information... dans des conditions garantissant l'anonymat fixées par décret en Conseil d'État pris après avis de la Commission nationale de l'informatique et des libertés »⁵. L'IDS a créé et favorisé la création de plusieurs bases de données⁶ qui collectent :

- des données médicalisées recueillies par un échantillon de médecins généralistes ;
- des données hospitalières transmises par des établissements publics que privés ;
- des données sur le handicap communiquées par les *Maisons départementales des personnes handicapées (MDPH)* ;
- des données sur la dépendance adressées par les Établissements d'Hébergement pour Personnes Agées Dépendantes (EHPAD) ;
- des données sur les causes médicales de décès envoyées par les médecins constatant les décès et par les mairies ;
- des données sur les cohortes pour la recherche et la sécurité sanitaire ;
- des données des assurances maladies complémentaires ;
- des indicateurs du Service d'Aide à la Décision ;
- etc.

En avril 2014, Etalab⁷ a publié, dans le cadre du débat public sur l'ouverture de données de santé lancée en novembre 2013, une cartographie des bases recensant et qualifiant les principaux jeux et bases de données sur la santé en France⁸. Cette cartographie identifie 313 bases de données, enquêtes et publications en lien avec la santé publique et plus spécialement :

- Consommation de soins et dépenses (104) ;
- Santé publique et épidémiologie (93) ;
- Offres de soins (78) ;
- Performance et opérations (38).

Pour effectuer ce recensement, Etalab a dépouillé les principaux rapports sur les données de santé en France, analysé des sites web de différents régimes de sécurité et interviewé des gestionnaires de données de santé.

Ces données peuvent être réparties en deux niveaux de granularité : « le niveau granulaire (données au niveau le plus fin qu'il est possible d'obtenir en fonction de l'origine de la donnée et du système de collecte), et le niveau agrégé (données obtenues en regroupant des données granulaires selon une ou plusieurs caractéristiques communes) »⁹. À savoir qu'une même base de données peut se retrouver dans

⁵ <http://www.institut-des-donnees-de-sante.fr/institut/la-loi/>

⁶ <http://www.institut-des-donnees-de-sante.fr/connaissance/>

⁷ Il s'agit d'un service du Premier ministre, au sein du Secrétariat général pour la Modernisation de l'Action publique, en charge de l'ouverture des données publiques et du gouvernement ouvert

⁸ <https://www.data.gouv.fr/fr/datasets/cartographie-des-bases-de-donnees-publiques-en-sante/>

⁹ Idem note de bas de page 7

les deux catégories. Outre leur contenu, l'importance des données réside également dans la possibilité de leur réutilisation. Cette dernière dépend :

- des termes et des conditions d'accès (libre accès, accès restreint ou accès réservé) ;
- du coût (payant ou gratuit) ;
- du format des données (exploitable : ods, csv, json ou non exploitable : en PDF ou consultable uniquement en ligne) ;
- des conditions de réutilisation (explicité sans restriction, explicité avec restriction ou non explicité).

Ces critères sont pris en considération dans la cartographie effectuée par Etalab. Parmi les jeux de données recensées, une centaine sont agrégées et en accès libre. En ce qui concerne les jeux de données granulaires, 70 % ont un accès restreint et 21 % un accès ouvert, les jeux de données restants étant réservés aux gestionnaires. La granularité la plus fine concerne les individus, les institutions, les entreprises ou les médicaments ; les données agrégées sont obtenues en regroupant des données granulaires, par exemple les personnes âgées de moins de 25 ans, le nombre de lits disponibles dans les urgences des publics en France.

Il est également intéressant d'*appréhender la couverture des données*. Celle-ci peut être hétérogène ou mixte, se référant à un échantillon de patients, de médecins, d'institutions, de périodes ou de territoires. Les données granulaires sont en accès libre. Elles ont trait aux services de soins et, plus particulièrement, ceux qui concernent les professionnels, les entreprises et les médicaments. Environ 50 % des bases recensées sont dans un format exploitable. Ces données proviennent des établissements de santé ou de systèmes tels que le SNIIRAM (Système National d'Information Inter-Régime de l'Assurance Maladie) et le PMSI (Programme de Médicalisation des Systèmes d'Information). Le nombre de bases de données avec un accès libre est très limité et le niveau d'hétérogénéité à différents niveaux est très élevé. Environ *un tiers offre des données agrégées avec un accès libre* sauf pour deux d'entre elles. Plus des deux tiers sont des données granulaires dont 70 % sont en accès restreint et 8 % en accès fermé. Seules 21 % sont en accès libre. Les données sous licence ouverte sont libres d'accès.

En raison du grand nombre de SI de santé, nous présenterons une sélection de bases de données afin d'illustrer notre propos.

Systèmes d'accès restreint et contrôlé

SNIIRAM

Cette base de données a été créée par la loi de financement de la Sécurité sociale en 1999. L'article 2 de l'arrêté du 19 juillet 2013 modifié le 6 octobre 2016 fixe ses finalités : améliorer la qualité des soins, contribuer à une meilleure gestion de l'assurance maladie et des politiques de santé, transmettre aux prestataires de soins les informations pertinentes relatives à leur activité, à leurs recettes et, s'il y a lieu, à leurs prescriptions. Les informations sur les patients proviennent des organismes de régimes d'assurance maladie et les informations relatives à l'activité hospitalière émanent de l'Agence technique de l'information sur l'hospitalisation (ATIH). Les données sur les patients concernent entre autres les données démographiques, les affections de longue durée, les consommations de soins, les consultations et les visites chez les médecins généralistes, les actes médicaux, les médicaments, les résultats d'analyses et les diagnostics ainsi que les dispositifs médicaux. Quant aux offres de soins, elles regroupent les spécialités des prescripteurs, des exécutants, les lieux d'exécution, l'implantation géographique ainsi que le statut conventionnel ou juridique.

Cette base de données étant gérée par la *Caisse Nationale d'Assurance Maladie des Travailleurs Salariés* (CNAMTS), son accès est fortement réglementé et l'utilisation des données doit faire l'objet

d'une autorisation de la Commission Nationale Informatique et Libertés (CNIL). L'article 4 de l'arrêté mentionné *supra* répertorie les catégories de personnes qui peuvent utiliser les données du SNIIRAM. Bien que les données soient anonymisées et que le numéro d'anonymat soit établi par codage informatique irréversible (art. 5), les chercheurs ayant accès à ces données ne peuvent pas mettre en œuvre simultanément plus d'une des quatre variables considérées comme sensibles (code commune, date des soins, mois et année de naissance, date de décès).

La modification du 6 octobre 2016 a assoupli certaines règles d'accès. Elle a notamment :

- prolongé la durée de l'autorisation à utiliser les données du SNIIRAM de 3 à 6 ans ;
- autorisé le traitement des informations par les organismes de recherche, les universités, les écoles ou autres structures d'enseignement liées à la recherche. Ces autorisations sont désormais soumises à l'approbation du bureau de l'IDS et de la CNIL.

Malgré cette réforme, l'utilisation des données du SNIIRAM est très réglementée et sommes loin des principes d'ouverture et de réutilisation des données telle que définie par les licences ouvertes (open licence).

Programme de Médicalisation des Systèmes d'Information (PMSI)

À sa création en 1982, le PMSI a été présenté comme un outil épidémiologique et comme un système favorisant un meilleur échange entre les médecins, les soignants et les administratifs. Le PMSI a été généralisé aux établissements de santé publique en 1994 et en 1996 pour le secteur privé. Depuis 2005, c'est un outil de description et de mesure médico-économique de l'activité hospitalière. Il permet de quantifier et de standardiser l'activité et les ressources des établissements de santé et de calculer ainsi les allocations budgétaires pour les établissements. Les données proviennent de différentes sources :

- PMSI MCO (médecine, chirurgie, obstétrique et odontologie) : résumé de sortie standardisé (RSS) pour les séjours hospitaliers en soins de courte durée. Ces RSS sont classés automatiquement dans des groupes homogènes de malades (GHM) ;
- PMSI SSR (soins de suite et de réadaptation) : recensement des hospitalisations réalisées dans des structures ayant une activité autorisée dans ce domaine ;
- PMSI HAD (hospitalisation à domicile) : recueil des modes de prise en charge qui sont classés en groupe homogène de prise en charge (GHPC) ;
- RIM-P (Recueil d'informations médicalisé pour la psychiatrie) : description de toute l'activité réalisée au bénéfice de malades par les établissements de santé, en hospitalisation complète ou partielle (résumé par séquence, RPS) comme en ambulatoire (résumé d'activité ambulatoire, RAA).

Cette base de données, permettant de coder, recueillir, analyser, restituer et diffuser l'information hospitalière tant dans le secteur public que privé, est gérée par l'Agence Technique de l'Information sur l'Hospitalisation (ATIH). Si certaines données de synthèses comme la cartographie de l'activité des établissements, le référentiel national de coûts, les indicateurs de santé périnatale, etc. sont accessibles sur le site *Scan santé*¹⁰, l'accès aux données détaillées et les extractions particulières sont réservés à des titulaires ayant obtenu l'accord de la CNIL qui est délivré en fonction de l'utilisation.

Quelle synthèse est-il possible de tirer de la description de ces bases de données ?

L'accès et l'utilisation de ces bases de données nécessitent une grande expertise en ce qui concerne le système de paiement de la sécurité sociale en France aux prestataires, les règles de remboursement

¹⁰ <http://www.scansante.fr>

(SNIIRAM) et le système de paiement hospitalier (PMSI). Pour faciliter l'utilisation de ces bases de données, des formations sont mises en place.

Comme les conditions d'accès sont très strictes et fortement contrôlées, les utilisateurs potentiels doivent adresser une demande d'accès à des organismes différents d'une base de données à une autre. Par exemple, des organismes de recherche français très réputés, comme l'*Institut national de la santé et de la recherche médicale* (INSERM), n'ont toujours pas la possibilité d'accéder directement aux données individuelles anonymisées de la SNIIRAM. Ils peuvent uniquement demander au gestionnaire du système de leur extraire un échantillon de données. Le délai moyen pour obtenir ces données est de 17 mois ; parfois il peut aller au-delà de deux ans¹¹. Cependant, les informations recensées par ces systèmes ont une valeur ajoutée pour la recherche. Par exemple, il est possible de calculer la consommation d'un traitement médical en France selon le profil du bénéficiaire et de surveiller cette consommation.

Croisement des données

Malgré la diversité des objectifs et des contenus de ces bases de données, certains croisements restent possibles :

1. Au niveau du patient, entre :
 - SNIIRAM et PMSI : l'activité hospitalière et les données de dépenses, les données sur les demandes de soins ambulatoires,
 - SNIIRAM et l'Enquête santé et protection sociale (ESPS) : les déterminants sociaux, les besoins en matière de soins de santé et les problématiques d'accès, les données sur les demandes de soins depuis 2012 ;
2. Au niveau des hôpitaux entre PMSI et La Statistique annuelle des établissements de santé (SAE) : Caractéristiques des hôpitaux ;
3. Au niveau des données agrégées : entre toutes les bases de données.

Par exemple, le Ministère du Travail, de l'Emploi, de la Formation Professionnelle et du Dialogue Social a publié des informations sur les distances et le temps d'accès aux soins en France métropolitaine. Ces données sont accessibles à tout public sur la plateforme data.gouv.fr. Cela représente un état d'accessibilité spatiale aux soins de santé en France, réalisé le 1er janvier 2007 et en avril 2011. Les bases de données permettant l'analyse de l'accès aux soins sont multiples et des domaines spécifiques sont concernés comme les soins de ville ou les soins hospitaliers. Pour les soins ambulatoires, les données sont extraites de la SNIIRAM.

Pour les soins hospitaliers, une méthodologie innovante pour le suivi et la géolocalisation des hôpitaux spécialisés repose sur les dossiers PMSI, SAE et le Fichier national des institutions sanitaires et sociales (FINESS). Pour l'équipement matériel lourd, les données SAE et SNIIRAM ont été mobilisées.

Vers un système unifié ouvert ?

L'article 193 de la loi sur la modernisation du système de santé du 26 janvier 2016 a mis en place le Système National des Données de Santé (SNDS). Le décret portant création de cette institution est paru le 28 décembre 2016 avec une entrée en vigueur au 1^{er} avril 2017. Le SNDS regroupe les données du SNIIRAM, du PMSI, du Centre d'épidémiologie sur les causes médicales de décès (CépiDC), des données médico-sociales de la Caisse nationale de solidarité pour l'autonomie (CNSA) et un échantillon représentatif des données de remboursement des complémentaires. Il sera géré par la CNAMTS et

¹¹ Accès aux données de santé. Propositions du comité d'experts de l'Institut des données de santé (IDS) adoptées par la majorité de ses membres. (2013) <http://www.institut-des-donnees-de-sante.fr/site-download-process/56-d2cq8k.html>

l'Institut des données de santé (IDS) sera remplacé par l'Institut national des données de santé (INDS) qui est censé devenir le guichet unique pour l'accès aux données. D'après ce décret, les données du SNDS seront anonymisées d'une manière irréversible et « l'Institut national de la santé et de la recherche médicale peut, en tant que coordinateur national d'infrastructures de recherche utilisant des données de santé, être chargé, dans le cadre d'une convention conclue avec la Caisse nationale de l'assurance maladie des travailleurs salariés, d'assurer la réalisation des extractions et la mise à disposition effective de données du système national des données de santé, pour des traitements mis en œuvre à des fins de recherche, d'étude ou d'évaluation ». Ce même décret définit les personnes qui disposeront d'un droit d'accès :

- les personnels ou les prestataires de la CNAMTS habilités et nommément désignés par son directeur général en raison de leurs missions ;
- les personnels ou les prestataires de l'INDS et de la recherche médicale dont les missions le justifient, habilités et nommément désignés par son président-directeur général.

Alors que les intentions d'ouverture des données de santé sont souvent mises en avant par les autorités administratives, dans les faits, ces intentions sont assorties de conditions d'accès très restrictives et des procédures très complexes¹².

Systèmes d'accès restreint

Enquêtes décennales de santé (EDS)

Cette enquête est réalisée par l'*Institut National de la Statistique et des Études Économiques* (INSEE). La première enquête a eu lieu en 1960-1961. L'objectif est de décrire l'état de santé de la population, comprendre la consommation annuelle de soins de santé et de prévention, la morbidité déclarée ainsi que la perception de la santé. Elle couvre des domaines assez larges : conditions de vie, protection sociale, état général de santé, handicap, description des maladies, recours au médecin et à l'hôpital, antécédents chirurgicaux, interruptions d'activité pour raisons de santé, habitudes alimentaires et prévention. La dernière enquête a été menée en 2002-2003 en conformité avec l'*European Health Interview Survey* (EHIS). Elle a généré trois ensembles de données granulaires avec un accès réservé aux membres du conseil scientifique et les chercheurs.

Baromètre Santé

Géré par l'*Institut National de Prévention et d'éducation pour la Santé* (INPES), ce baromètre existe depuis 1992. Ces enquêtes abordent les comportements et attitudes de santé des Français (toxicomanie, comportements sexuels, etc.), les risques professionnels et les mesures de prévention prises en soins primaires (dépistage, éducation à la santé). L'enquête 2016 traite des perceptions et pratiques relatives aux maladies infectieuses. Ce baromètre est une source fiable pour les études de santé publique même si les questions ne sont pas les mêmes d'une enquête à l'autre.

Ce baromètre de santé a généré des données granulaires accessibles gratuitement avec un accès restreint aux chercheurs via un accord signé avec l'INPES. Le format des données est réutilisable avec quelques restrictions.

¹² Pour les conditions d'accès voir la Délibération n° 2016-316 du 13 octobre 2016 portant avis sur un projet de décret en Conseil d'État relatif au Système national des données de santé.

Systèmes d'information ouverts

Grâce à l'Internet, certaines informations de santé ont pu être mises à disposition en ligne. Depuis 2013, plusieurs systèmes d'information de santé ont été conçus et développés pour une publication en ligne avec une politique d'ouverture et de partage des données sous une licence ouverte. En plus de la statistique des établissements de santé, on peut aussi mentionner la *Base de données publique des médicaments* (BDPM) et *Scope santé* qui sont complètement orientées vers les utilisateurs.

Statistique Annuelle des Établissements de santé (SAE)

La première enquête a été menée en 1995 pour collecter les données de 1994. Il s'agit d'une enquête administrative obligatoire concernant tous les établissements publics de santé français (environ 1 000) et les établissements privés (environ 2 500). Elle est réalisée par la *Direction de la Recherche, des Études, de l'Évaluation et des Statistiques* (DREES). Cette plateforme permet différents types de requêtes et méthodes de collecte (tables, SQL...). Les thèmes abordés sont : types de salles, capacité, équipement, personnel (salariés, praticiens indépendants, pharmaciens...), activités (nombre de jours d'hospitalisation, nombre de séances de radiothérapie...). Cette statistique génère trois séries de données granulaires, accessibles gratuitement à tous dans un format exploitable (Excel).

BDPM : Base de données publique des médicaments

Lancée en octobre 2013, « la base de données publique des médicaments permet au grand public et aux professionnels de santé d'accéder à des données et documents de référence sur les médicaments commercialisés ou ayant été commercialisés durant les trois dernières années en France. Cette base de données administrative et scientifique sur les traitements et sur l'usage des produits de santé est mise en œuvre par l'Agence nationale de sécurité du médicament et des produits de santé (ANSM), en liaison avec la Haute Autorité de santé (HAS) et l'Union nationale des caisses d'assurance maladie (UNCAM), sous l'égide du ministère des Affaires sociales et de la Santé. » Chaque notice donne, en plus des informations usuelles concernant un médicament, deux indicateurs sur les bénéfices médicaux, le *Service Médical Rendu (SMR)* et l'*amélioration du service médical rendu (ASMR)*. Ces indicateurs contribuent à répondre à deux questions : *ce médicament est suffisamment efficace pour être pris en charge totalement par la sécurité sociale ?* et : *ce médicament apporte-t-il une réelle amélioration parmi les autres traitements disponibles ?* Ces indicateurs permettent de fixer le taux de prise en charge par l'UNCAM et de négocier le prix de vente avec les entreprises.

Scope santé

Cette base de données a été lancée en novembre 2013. Elle renseigne sur le niveau de qualité de tous les hôpitaux et les cliniques de France, soit plus de 5 000 établissements de santé. Elle est destinée principalement aux utilisateurs puisqu'ils peuvent comparer les établissements selon différents critères et leur fournit une information officielle évaluant la qualité et les indicateurs d'activités, de qualité et de sécurité des soins dans les établissements de santé. En effet, depuis 2009, les établissements de santé, publics et privés, ont l'obligation de fournir leurs résultats ; ces informations n'étaient pas accessibles au public avant 2013. *Scope santé* présente des données unifiées et intégrées fournies par des organismes qualifiés et émanant de : SAE, PMSI, HAS, les résultats de luttes contre les infections nosocomiales (Bilan LIN), le Fichier National des Établissements Sanitaires et Sociaux (FINESS).

Les informations de cette base de données sur les établissements peuvent être explorées selon différents critères : recherche sur carte avec des fonctions de géolocalisation, recherche par profil (enfants et adolescents, femmes enceintes, personnes âgées, etc.), par type de soins (hospitalisation à domicile, soins à long terme, etc.), par spécialités médicales, par équipements disponibles (scanner, etc.), par types d'établissements (public, privé, secteur privé sans but lucratif). Les utilisateurs peuvent également utiliser des filtres concernant le niveau de certification (avec ou sans recommandation, avec des réserves,

non certifiés, certification en cours...) et le niveau de performance par indicateur. Cette base de données illustre bien l'énorme tâche à accomplir avant l'ouverture des données notamment pour les rendre interopérables, pour les diffuser et pouvoir les visualiser selon différentes modalités, pour rendre leur consultation adaptée à l'utilisateur final et au citoyen.

La transparence de ce type d'indicateurs a déjà fait l'objet de plusieurs controverses et cette approche a été souvent questionnée. Philippe Leduc, directeur du think tank *Économie Santé*, s'interroge : « *Mais, est-ce bien normal que cela soit les patients qui, de fait, vont décider de l'avenir d'un service d'un établissement de santé en évitant d'aller là où la note octroyée laisse frémir ? Non, assurément.* » (Leduc 2013)

5. Vers une ouverture des données de santé ?

Cette revue de la littérature et des systèmes d'information de santé français démontre la complexité du problème des données de santé. L'ouverture des données s'effectue dans un contexte fragmenté. Plusieurs agences exécutives sont impliquées ; leurs périmètres de compétence se télescopent. Les systèmes qui produisent des données essentielles pour développer des indicateurs et pour réaliser des études sont très mal accessibles même pour les chercheurs. Les économistes et les chercheurs ont besoin d'utiliser ces données et de les croiser afin d'exploiter tout leur potentiel dans le respect de la vie privée. En effet, « la CNAMTS seule n'a pas les moyens d'exploiter au maximum ces données » (L'Hote 2014).

« La France possède tous les ingrédients nécessaires pour mettre en place un système d'information de santé fiable au niveau national. Le principal challenge est l'absence d'une réelle politique nationale de santé » (Goldberg et al. 2012). Un leadership politique serait-il essentiel pour améliorer le contexte et accroître la transparence et l'efficacité des systèmes de santé ? La création de l'INDS avec des conditions d'accès complexes conduira-t-elle à l'ouverture des données tant souhaitée par de nombreux chercheurs ?

« Les acteurs -responsables politiques et administratifs, décideurs des régimes de protection sociale, professionnels de santé, usagers, organismes agissant à des fins d'études et de recherches en santé publique- sont confrontés à des enjeux qui nécessitent une connaissance plus exhaustive et mieux partagée du système de santé » (IDS 2013).

Les efforts déployés par l'IDS (demain l'INDS) peuvent aider les autorités publiques à prendre conscience des difficultés d'accès et de la dispersion des systèmes d'information, à formaliser les problèmes et à travailler pour rationaliser les interactions entre de multiples organes administratifs. La diversité et l'hétérogénéité de ces données constituent un obstacle majeur à l'interopérabilité et de ce fait à l'exploitation de ces données. D'ailleurs, l'ouverture des données nécessite souvent un travail préliminaire important, travail d'autant plus important que les indicateurs sont calculés à partir de multiples sources elles-mêmes hétérogènes, gérées par différentes agences exécutives !

L'analyse résumée plus haut démontre qu'il existe actuellement de multiples ressources ; que les formats distincts conduisent à une haute hétérogénéité à différents niveaux. Chaque système poursuit un large éventail d'objectifs. L'ouverture des données sur la santé concerne en priorité la consommation et les dépenses de santé, la santé publique, les soins ambulatoires et le rendement (financier, opérationnel et qualitatif) du système de santé français. Les données ouvertes sont principalement des données agrégées par rapport aux données granulaires. Un grand nombre de ces données granulaires sont d'accès restreint.

Lorsqu'une information est libre, elle est généralement gratuite. Les données granulaires ouvertes concernant les patients sont toujours anonymisées et d'accès restreint ou fermé. En 2013, des bases de données granulaires (*Scope santé* et *BDPM*) ont été créées sous une licence ouverte. Depuis, les autorités administratives ont laissé espérer une meilleure ouverture des données, mais cet espoir est, pour le

moment, sans effet réel sur le terrain. La création de l'INDS sera-t-elle le signe d'une réelle ouverture des données relatives à la santé en France ?

6. Conclusion

Dans ce chapitre, nous avons présenté des positions, des décisions et des controverses concernant les données ouvertes sur la santé en France. Nous avons brièvement décrit l'état d'ouverture de l'accès aux données de santé. Cela met en évidence des réticences à l'ouverture et une faible réutilisation des données de santé à court terme. Le processus qui pourra rendre les différents systèmes interopérables sera long. Il suppose un soutien politique en faveur de l'ouverture des données. Il se poursuit par la transformation des systèmes existants afin de rendre leurs données ouvertes et interopérables.

Les frontières et les limites entre les différentes catégories de données -personnelles ou publiques, anonymes ou indirectement identifiables, données massives et données ouvertes- ne sont pas claires. En effet, le 21 octobre 2013, le Parlement européen a voté une révision majeure de la directive de 1995 pour renforcer le contrôle des citoyens sur leurs données personnelles. Le 27 avril 2016, le Parlement européen et le Conseil de l'Union ont adopté un nouveau règlement sur la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données. Ce règlement (contrairement à une directive) est directement applicable à l'ensemble des États membres de l'Union européenne (UE) à compter du 25 mai 2018. Le point 35 définit les données de santé ; l'article 9 interdit « le traitement des données à caractère personnel qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique ». Malgré quelques exceptions (consentement explicite, sauvegarde des intérêts vitaux de la personne, etc.), l'accès aux données reste limité à des catégories de professionnels. Ce règlement ne facilite donc pas l'ouverture des données de santé.

Dans un communiqué de presse publié le 9 juillet 2014, la ministre chargée de la santé a annoncé qu'elle souhaitait accélérer le processus d'ouverture des données de santé. Elle veillera en particulier à ce que les résultats des enquêtes et recherches financées par des fonds publics soient d'accès gratuit. Le 4 juillet 2016, lors de la présentation de la stratégie *e-santé*, elle a rappelé que l'impulsion en faveur de l'ouverture maîtrisée des données de santé doit être pérenne. Elle devrait créer au ministère une fonction d'*administrateur des données de santé*, dont l'une des missions serait d'*accélérer* l'open data en santé. À l'issue de la réflexion engagée sur le big data (sur les mégadonnées) elle lancerait un plan big data en santé afin d'aller plus loin encore dans la valorisation des données. Mais si des actions ont été menées sur le terrain, les résultats ne sont pas toutefois probants. À savoir que cette ouverture réclamée par le monde de la recherche suscite beaucoup de controverses comme on a pu le constater pour les données sur les établissements et les médicaments. Qu'en sera-t-il des données individuelles ouvertes, même si elles vont être dé-identifiées ?

Les intentions de la Ministre sont louables, mais tiennent-elles compte de la complexité extrême des questions, souvent contradictoires, que soulèvent le système de santé français, l'interaction des acteurs, leurs intérêts divergents et l'application des règlements européens ? Quelles seront les données vraiment ouvertes ? Dans quels délais ? Beaucoup de questions qui espèrent encore une réponse...

Bibliographie

- Bégaud, Bernard ; Costagliola, Dominique (2013) Rapport sur la surveillance et la promotion du bon usage du médicament en France. Ministère des Affaires sociales et de la Santé, éd. Paris.
- Bouchoux, Corinne (2014) Rapport d'Information fait au nom de la mission commune d'information sur l'accès aux documents administratifs et aux données publiques. Sénat, éd. Paris.
- Bras, Pierre-Louis (2013) Gouvernance et utilisation des données de santé. Ministère des Affaires sociales et de la Santé, éd. Paris.
- Brun, Nicolas ; Hirsh, Emmanuel ; Kivits, Joëlle ; Emery, Gregory ; Garcia-Viana, Audrey ; Lopez, Antonia (2011) Nouvelles attentes du citoyen, acteur de santé. Ministère du Travail, de l'Emploi et de la Santé, éd. Paris.
- Geissbuhler, A. ; Safran, C. ; Buchan, I. ; Bellazzi, R. ; Labkoff, S. ; Eilenberg, K. et al. (2013) Trustworthy reuse of health data: a transnational perspective. In : International journal of medical informatics, vol. 82, n° 1, p. 1–9.
- Godeluck, Solveig (2016) Déficit de la Sécurité sociale. L'effet de la crise de 2008 est presque effacé. In : Les Echos, 15 mars 2016.
- Goldberg, Marcel ; Jouglu, Eric ; Fassa, Maniane ; Padieu, René ; Quantin, Catherine (2012) The French health information system. In : Statistical Journal of the IAOS, n° 28, consulté le 30 novembre 2015.
- Health & Social Care Information Center (2015) Supporting transparency and open data. HSCIC.
- IDS (2013) Rapport au parlement 2013. Des avancées et des propositions raisonnées.
- Leduc, Philippe (2013) Scope Santé met à nu les hôpitaux et les cliniques. In : Les Echos.
- L'Hote, Quentin (2014) L'open data et la santé en France. In : Les Echos, 17 juillet 2014.
- Manyika, James ; Chui, Michael ; Groves, Peter ; Farell, Diana ; Kuiken, Steve Van, Almasi Doshi, Elizabeth (2013) Open data. Unlocking innovation and performance with liquid information : McKinsey & Company.
- Olson, Steve ; Downey, Autumn S. (2013) Sharing clinical research data. Workshop summary. Washington, District of Columbia : The National Academies Press.
- Picard, Robert ; Vial, Antoine ; Laval, Didier ; Varenne, Dominique (2012) De l'information du patient à la démocratie sanitaire. Enjeux et conditions d'un usage efficient des technologies : rapport. Paris : ministère de l'Économie et des Finances.