

# Données de santé : données sensibles

## Présentation du dossier



Jean-Christophe Thalabard<sup>1</sup>  
et Marius Fieschi<sup>2</sup>

*La capacité à recueillir de manière objective des signes sur un grand nombre de patients et à croiser ces informations (tables de contingence) a marqué au 19ème siècle la naissance d'une médecine scientifique moderne [Foucault, 1963]. La mise en place de registres de maladie au 20ème siècle, l'essor de l'épidémiologie et, dans le champ de la thérapeutique, le développement de la médecine reposant sur les preuves (Evidence Based Medicine - EBM) en représentent des prolongements naturels, que les possibilités techniques permettent d'étendre au delà du champ médical restreint en les délocalisant, les dématérialisant et en ouvrant les possibilités de partages effectifs en temps réel ou presque de cette information. La gestion des données de santé, longtemps laissée aux seuls professionnels sur un mode papier traditionnel, d'intérêt relativement limité et considérée comme une contrainte légale (cf. les archives des dossiers médicaux dans les hôpitaux), s'est transportée sur de nouveaux supports et est considérée de nos jours comme une ressource, objet d'intérêt croissant.*

Des actions incitatives, fortement stimulées par la nécessité d'améliorer les conditions d'enregistrement, de suivi et de remboursement des prestations de soins, ont vu les professionnels de santé s'informatiser tant dans le secteur de l'exercice libéral que dans le secteur hospitalier, générant au fur et à mesure des bases de données multiples aux finalités diverses, très hétérogènes dans leur qualité, et dont une partie de la gestion est de plus en plus confiée à des sociétés privées du secteur concurrentiel. Les media se font largement l'écho d'un souhait "d'une ouverture" de ces données de santé au plus grand nombre, propre à anticiper, voire éviter quelques scandales sanitaires, comme plaidé par certains [de Kervasdoué et Sicard 2013, Bar-Hen et Flahault, 2013], point de vue qui fait par ailleurs l'objet de discussions plus nuancées [Cabut, 2013, Larousserie, 2014].

Cette ouverture apparaît répondre à des objectifs variés : classification comparative de l'offre de soins permettant, en théorie, au consommateur de soins d'intervenir dans le choix de son soignant et/ ou de son lieu de prise en charge sur des critères qui lui seraient propres ; analyse des comportements de soins afin de pouvoir investir et au besoin influencer sur ces comportements dans une logique industrielle concurrentielle ; analyse des disparités géographiques en termes d'offre de soins et harmonisation des pratiques professionnelles par rapport à un référentiel de recommandations élaboré par des acteurs du soin, tel que

1. Membre du comité de rédaction de Statistique et Société

2. Professeur de statistiques, information médicale, ancien directeur du Laboratoire d'Enseignement et de Recherche sur le Traitement de l'Information Médicale, Université de la Méditerranée, Marseille

proposé par la Haute autorité de santé dans l'optique d'une amélioration de l'efficacité des pratiques (rapport efficacité bénéfique/ coût), etc. Enfin, la connaissance de la maladie est une finalité, composante importante de cette demande, que ce soit, en chronique, dans l'étude de l'histoire naturelle de l'évolution des pathologies, avec ou sans traitement, en population dans leur environnement géographique climatique, professionnel, social, ou, en aigu, dans la capacité à détecter des signaux d'émergence de formes particulières de pathologies (détection d'épidémies, malformations, anomalies génétiques, etc.)

## Une confusion dans la nature des données disponibles

Avner Bar- Hen, enseignant-chercheur statisticien, ancien enseignant à l'École des Hautes Études en Santé Publique, insiste sur l'ouverture dite "citoyenne" des données publiques quelles que soient leurs natures, dont celles relatives à la santé, tout en rappelant les différents types et soulignant bien la différence entre la fouille de données orientées santé dans des grandes bases d'information croisées (Big data) et l'accès ouvert aux données individuelles d'ordre administratif, pour peu qu'elles aient été convenablement anonymisées au préalable. La question du contrôle d'une part de l'information partageable et d'autre part des personnes habilitées à accéder aux dites informations reste un sujet de débat important. La réponse ne peut être binaire et le débat doit être libéré d'une opposition formelle entre les supposés anciens et les modernes auto-déclarés. Des décisions récentes concernant l'accès prises par la CADA<sup>3</sup> [Daël, 2013] marquent l'évolution de la jurisprudence.

Face à cette position volontiers polémique, propre à ouvrir le débat et susciter des réactions, Statistique et Société a souhaité rassembler des contributions de professionnels des bases de données de santé publiques ou semi- publiques confrontés à cette nouvelle demande.

## Les bases de données de la Sécurité sociale sont au centre de bien des controverses

La France bénéficie d'un système de santé original, né sur un principe de solidarité au décours de la seconde guerre mondiale (Sécurité Sociale et assurances complémentaires). Ce système médico-administratif produit de grandes bases de données centralisées. La revendication d'une ouverture large de ces données au plus grand nombre, formulée par certains, témoigne parfois d'une méconnaissance des possibilités existantes. Le point de vue de l'opérateur principal (Caisse nationale d'assurance-maladie des travailleurs salariés - CNAMTS) à travers la contribution de Dominique Polton et Claude Gissot du service des statistiques et de l'évaluation économique de la CNAMTS montre des avancées concrètes, illustrées par des contributions importantes en termes de connaissance, de croisements possibles, tout en soulignant la prudence nécessaire du fait des limites de ces bases.

## Spécificité des données de santé

L'existence d'une spécificité des données de santé qui déborderait la réflexion sur le secret statistique est abordée par la contribution de Jean-Pierre Le Gléau. Il rappelle la raison d'être du secret statistique, son évolution et les efforts déployés pour le contrôler sans entraver le travail des chercheurs et des utilisateurs des données, en amenant quelques pistes techniques qui pourraient s'appliquer aux données de santé. Au delà des aspects techniques, l'accès à des données sensibles touchant à la vie privée soulève les questions importantes des garanties apportées à chacun de protection de sa propre intimité. La particularité de notre époque est bien que la donnée est devenue dématérialisée et voyage facilement, exigeant une réflexion et des décisions appropriées. L'accès aux données de santé expose, en cas de divulgation consciente

---

3. Commission d'accès aux documents administratifs

ou inconsciente, les personnes concernées à des risques de stigmatisation avec toutes les conséquences éventuelles psychologiques et sociétales, potentiellement dramatiques, qui touchent à une dimension particulière des individus, au delà de l'économique, et ce d'autant plus qu'elle concerne le sujet malade, donc plus vulnérable. Il nous a paru important de donner la parole à une juriste qui a travaillé sur ces sujets à la CNIL et est aujourd'hui liée à la structure de coordination dédié à l'e- santé, l'ASIP (Agence des Systèmes d'Information Partagés en Santé), apparue dans le paysage depuis quelques années <http://esante.gouv.fr>. Jeanne Bossi nous rappelle les principes et règles encadrant les données sensibles dans le cadre français, au sein du concert européen [Directive, 2013]. Est particulièrement importante la distinction entre données directement ou indirectement identifiantes, d'autant que les avancées algorithmiques permanentes rendent de plus en plus complexe cette délimitation [Schadt et al., 2011, Schadt, 2012, White, 2013].

## **L'intérêt de données de suivi individuel sur le long terme et du croisement des bases d'information**

L'histoire de la santé est riche d'avancées liées à l'accès à de telles données, dont l'importance n'est pas forcément liée à la multiplicité des variables recueillies mais surtout à la qualité d'un recueil systématique dans une population bien définie. Dès le 18ème siècle, Spallanzani avait pu mettre en évidence des liens entre occupation professionnelle et pathologies, plus récemment S Barker a pu montrer le lien entre certaines anomalies simples staturo-pondérales à la naissance et l'apparition, plus de 50 ans plus tard, de pathologies cardio-vasculaires et métaboliques [Barker et al., 1989, Barker et al., 1993], posant indirectement la question de l'utilisation de données pour d'autres finalités que celles pour lesquelles elles avaient été initialement recueillies. La contribution de Marie Zins et Marcel Goldberg, avec l'exemple de la cohorte Constances, nous montre l'évolution des pratiques des épidémiologistes dans la constitution de grandes cohortes s'appuyant sur les bases médico-administratives existantes pour les mises à jour et l'amélioration de la qualité des données recueillies au cours du temps. L'objectif ici est bien la qualité des données recueillies dans une finalité très générale de mettre à disposition un outil pour des questionnements précis futurs, véritable entreprise technique ouverte. L'analyse d'une telle masse de données ne se résume plus à des méthodes simples bien maîtrisées mais peut nécessiter des techniques d'imputation diagnostique sophistiquées, qui ne sont plus propres au domaine de santé.

## **Des enjeux et jeux d'acteurs sociaux importants**

L'ensemble de ces questionnements est loin de se limiter à de simples aspects techniques, ou aux réticences de certains acteurs à ouvrir leurs informations. La contribution de François Briatte et Samuel Goëta nous éclaire sur les aspects sociologiques et sur les enjeux sous-jacents. Concernant l'atteinte à la vie privée notamment, toute la question est bien de savoir si les avantages d'une fouille de bases de données comme celles de la CNAMTS ayant pour finalité revendiquée une amélioration de la qualité des soins, que ce soit l'efficacité dans des choix de prise en charge au niveau individuel ou l'efficacité en terme de dépenses de santé pour rester sur le modèle économique d'un système solidaire, justifie le petit risque assumé d'une rupture de cette confidentialité pour quelques uns ou plus ? Quelle est la perception de cette question par la génération Y, voire la suivante élevée dans la banalité du cloud, de l'i-phone et des réseaux sociaux, où finalement la donnée personnelle se partage sans arrière pensée particulière en la confiant, dans un premier temps assez naïvement, à des prestataires de service ne garantissant pas la maîtrise des usages qui pourraient en être faits et le droit à l'oubli ? Quelle est alors la connaissance et la perception de la loi actuelle (loi de 1978 modifiée en août 2004) qui insiste beaucoup sur des règles qui concernent les données identifiantes que ce soit directement mais également indirectement ?

## Les évolutions en cours et à venir

Nous vivons une période d'intenses débats et échanges où la volonté d'ouverture et d'ajustement au monde actuel et à ses capacités techniques est tangible, comme le montrent les travaux de commissions parlementaires consultables sur internet et la disponibilité récente du rapport Bras [Bras et Loth, 2014]. Le présent dossier vise à apporter quelques éléments factuels propres à nourrir le débat.

### Références

- [1] Bar-Hen A. et Flahault A.. Donnons aux citoyens accès aux données de santé. *Le Monde*, 8 Mars 2013.
- [2] Barker D. J., P. D. Winter, C. Osmond, B. Margetts, and S. J. Simmonds. Weight in infancy and death from ischaemic heart disease. *Lancet*, 2 (8663): 577-580, Sep 1989.
- [3] Barker D.J., C. N. Hales, C. H. Fall, C. Osmond, K. Phipps, and P. M. Clark. Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome x): relation to reduced fetal growth. *Diabetologia*, 36 (1): 62-67, Jan 1993.
- [4] Bras P.-L. et A. Loth. Rapport sur la gouvernance et l'utilisation des données de santé. Ministère des Affaires Sociales et de la Santé, IGAS, 2014.
- [5] Cabut Sandrine Santé publique: les leçons du modèle scandinave. *Le Monde*, 1<sup>er</sup> Février 2013.
- [6] Daël. S. Avis de la Cada 20134348 du 21/11/2013, CADA Novembre 2013.
- [7] de Kervasdoué J. et D. Sicard. Plus grave que le débat sur la pilule: l'affaire des données de santé publique. *Le Monde*, 16 janvier 2013.
- [8] Directive 2013/37/EU du parlement européen et du conseil du 26/06/2013 modifiant la directive 2003/98/EC concernant la réutilisation des informations du secteur public. *Journal Officiel de l'Union européenne*.
- [9] Foucault M. Naissance de la Clinique. PUF, 1963. ISBN : 978-2-13-057865-9.
- [10] Larousserie David. Données sensibles. l'équation impossible. *Le Monde*, 9 avril 2014.
- [11] Schadt E. E. The changing privacy landscape in the era of big data. *Mol Syst Biol*, 8: 612, 2012.
- [12] Schadt E. E., M. D. Linderman, J. Sorenson, L. Lee, and G. P. Nolan. Cloud and heterogeneous computing solutions exist today for the emerging big data problems in biology. *Nat Rev Genet*, 12 (3): 224, Mar 2011.
- [13] Weitzman E. R., L. Kaci, and K. D. Mandl. Sharing medical data for health research: the early personal health record experience. *J Med Internet Res*, 12 (2): e14, 2010. doi: 10.2196/jmir.1356. URL <http://dx.doi.org/10.2196/jmir.1356>.
- [14] White S. E.. De-identification and the sharing of big data. *J AHIMA*, 84 (4): 44-47, Apr 2013.