

Une loi pour mieux utiliser les données de santé



Entretien avec André LOTH

Directeur de projet à la DREES - Direction de la recherche, des études, de l'évaluation et des statistiques du Ministère de la Santé

Statistique et Société a consacré dans son numéro 1 de 2015 un dossier à l'accès aux données de santé. Depuis, a été votée et promulguée la loi de modernisation de la santé – loi du 26 janvier 2016 dite « Loi Touraine ». Cette loi, par son article 193, instaure de nouvelles dispositions pour l'accès aux données de santé. André Loth, qui a beaucoup œuvré à la préparation de ce texte de loi, nous explique en quoi les nouvelles dispositions doivent permettre un essor des études quantitatives sur les problèmes de santé en France.

Statistique & Société : Comment votre parcours professionnel vous a-t-il amené sur ce sujet ?

André Loth : Le hasard du calendrier politique a mis sur le devant de la scène un sujet qui se trouvait dans le droit fil de mes travaux antérieurs. Dès 1989 j'ai travaillé, au ministère de la Santé, à la généralisation du « Programme de médicalisation des systèmes d'information » (PMSI) des hôpitaux. J'ai ensuite fréquenté les données de santé dans divers postes, à la Caisse Nationale de l'Assurance Maladie, au Centre Hospitalier de Lille... Lorsque la polémique sur les données de santé a éclaté en 2013, j'avais déjà conduit la rédaction d'un rapport interne sur ce sujet : c'est donc naturellement que j'ai participé à la rédaction du rapport demandé par la ministre à Pierre-Louis Bras, inspecteur général des affaires sociales, en 2013 [1], puis à la commission « Open Data » mise en place par la ministre en 2014 [2], et enfin à la préparation du texte qui est devenu, après discussion au Parlement, l'article 193 de la loi de modernisation de notre système de santé. Ce texte a ainsi bénéficié d'une réflexion préalable approfondie au sein de ce ministère, qui a permis de dégager des principes solides puis de construire un consensus.

S&S : Pouvez-vous d'abord rappeler de quelles données de santé il s'agit ?

AL : La raison initiale du projet de loi c'est de rendre plus accessibles les données dites médico-administratives : ce sont notamment les résumés de séjour hospitaliers et les feuilles de soins, données recueillies initialement pour gérer et contrôler (financer les hôpitaux, rembourser les soins de ville...). Mais l'article 193 a aussi modifié le chapitre de la loi informatique et libertés, qui porte sur toutes les données de santé utilisées à des fins de connaissance (de « recherches, études ou évaluations »). Cela comprend les données médico-administratives mais aussi les registres, enquêtes, cohortes, panels constitués à partir de données recueillies initialement pour

1. Journal de la Société statistique de Paris, tome 129, n. 4 (1988), p. 227-243 ; reproduit dans le présent numéro.

soigner (dossiers médicaux) ou celles recueillies auprès des personnes lors d'une recherche. En France, on dispose de bases de données médico-administratives particulièrement riches, car exhaustives dans leur champ et centralisées au plan national : le recueil des résumés de séjour hospitaliers du PMSI a commencé au début des années 1990 ; l'ajout des codes des actes et des médicaments dans les feuilles de soins a commencé un peu plus tard ; une loi de 1998 a prévu de rassembler le tout dans un Système national d'information inter-régimes de l'assurance maladie (SNIIRAM). La montée en charge a pris quelques années mais toutes les données d'une même personne sont « chaînées » depuis 2009, ce qui constitue une richesse unique au monde parmi les grands pays. La loi nouvelle y ajoute les causes de décès et prévoit d'ajouter les données sur le handicap et un échantillon de données de l'assurance maladie complémentaire.

S&S : Vous avez parlé de « principes solides » qui fondent la loi de 2016. Quels sont-ils ?

AL : Deux principes :

1° Toutes les données de santé anonymes et disponibles doivent être mises en « open data », et le problème est d'en rendre disponibles davantage.

2° Les données non anonymes doivent pouvoir être utilisées pour le bien collectif sans que cela porte atteinte à la vie privée des personnes concernées.

Ces deux principes étaient déjà contenus dans le rapport Bras [1], mais n'avaient pas pour autant été acceptés dans le contexte polémique de 2013-2014 : en particulier, certains estimaient que le souci de la vie privée n'était qu'un prétexte pour bloquer les données et en réserver l'usage à l'administration ou à des « mandarins ». Les discussions de la commission « Open Data en santé » [2] ont permis d'avancer vers un consensus, et d'inscrire ces principes dans la loi, qui crée deux filières bien distinctes : « open data » pour les données anonymes, accès contrôlé et sécurisé pour les données qui ne le sont pas.

S&S : Est-on au clair sur la notion de données anonymes ?

AL : Non, pas autant qu'il le faudrait. Le mot « anonyme », qui est ambigu, ne figure pas dans l'article 193 : on y parle de données « constituées de telle sorte que l'identification, directe ou indirecte, des personnes concernées y est impossible » (nouvel article L1461-2 du code de la santé publique). Pour garantir l'anonymat (en ce sens) d'un jeu de données, il ne suffit pas d'en retirer les noms, les adresses, et les numéros de sécurité sociale, et de les remplacer par des identifiants non significatifs¹ : dans de nombreux cas, des informations contenues dans les données elles-mêmes permettent de retrouver des personnes. Par exemple, les dates d'hospitalisation dans un établissement donné, combinées avec l'âge et le lieu résidence, même approximatifs, peuvent permettre à des tiers d'identifier un malade, et grâce aux autres données de savoir quelle est la maladie dont cette personne souffre, ce qui n'est pas acceptable. C'est le problème de la « ré-identification » [4]. Les données de santé mises en « open data » doivent être constituées de telle façon que le risque de ré-identification soit négligeable (sachant qu'il ne peut pas être rigoureusement nul).

S&S : Parlons d'abord de cette filière « open data ». Comment la développer ?

AL : Le premier enjeu est d'arriver à produire à moindre coût des données où le risque de ré-identification soit aussi faible que l'exige la loi. C'est un domaine de recherche pour des statisticiens, et cela concerne aussi la CNIL. L'autre enjeu est de pousser l'activité des producteurs de statistiques dans ce domaine : bien souvent, ils sont plus enclins à réaliser eux-mêmes des études qu'à établir des bases agrégées de sommes et de ratios, ou des requêtes pré-formatées, permettant un large accès aux données de base.

1. On parle alors de données « pseudonymisées »

La loi a prévu aussi² une procédure intermédiaire, dite homologuée (par la CNIL), pour des données agrégées ou échantillonnées qui présenteraient un risque de ré-identification non nul mais résiduel comme c'est le cas aujourd'hui pour les données mises à la disposition des chercheurs en sciences humaines et sociales par le Réseau Quêtelet. Il faudra se mettre d'accord avec la CNIL sur la définition du seuil et des procédures allégées.

Je signale que la loi a par ailleurs étendu le principe d'« open data » aux données nominatives sur l'activité des professionnels de santé (tarifs) déjà publiées par le site ameli.fr.

S&S : La seconde filière concerne donc l'accès aux données à caractère personnel, indirectement nominatives. Cet accès doit être réglementé. Quelles sont les dispositions que la loi met en place pour définir un bon équilibre entre les droits d'accès et la protection de la vie privée ?

AL : La loi santé pose trois groupes de règles : sur les personnes qui peuvent accéder à ces données, sur les finalités, sur les modalités d'accès.

Les règles sur les personnes d'abord. Le point le plus discuté a été : le secteur privé pourra-t-il avoir accès à ces données ? Il était admis dans le projet du gouvernement que des cabinets d'étude privés pourraient traiter les données au même titre que les laboratoires de recherche publics, mais quid des autres ? La réponse a finalement été oui, à l'exception de deux catégories d'acteurs : les assureurs et les industriels de la santé, qui devront passer par des intermédiaires agréés³.

Les finalités : ici intervient la notion « d'intérêt public ». L'accès aux données de santé à caractère personnel, quelles que soient les précautions techniques prises, présente un risque résiduel d'atteinte à la vie privée⁴ : c'est pourquoi cet accès est réservé aux recherches, études et évaluations qui présentent un caractère d'intérêt public, ce qui n'exclut pas que ces travaux puissent avoir simultanément un intérêt privé : recherches sur les médicaments par exemple. Le caractère d'intérêt public de la recherche sera matérialisé par une publication des résultats, pour que ceux-ci soient partagés par la collectivité. La loi prévoit que cette publication intervienne « dans un délai raisonnable » après la fin de la recherche⁵, et qu'elle soit assurée par l'Institut national des données de santé (INDS) si un éditeur privé ne peut être trouvé. Un collègue au sein de l'INDS donnera à la CNIL un avis sur le caractère d'intérêt public des projets dans les cas où la question se posera ; un comité d'experts sera chargé quant à lui de vérifier que la finalité et la méthode envisagée de traitement nécessitent bien l'accès aux données demandées sans pour autant s'ériger en juge des méthodologies. Enfin, la loi exclut deux finalités : la promotion des produits de santé et le ciblage de groupes d'assurés (voir plus haut).

Les modalités enfin : les conditions d'accès doivent assurer que les données accessibles soient uniquement celles nécessaires pour la recherche autorisée⁶, ne tombent pas sous les yeux de personnes non autorisées, ne soient pas modifiées, et que les exploitations qui en sont faites puissent être retracées. Un « référentiel de la sécurité » est en préparation entre le ministère et la CNIL, et fera probablement l'objet d'un débat public : l'exemple du centre d'accès sécurisé créé pour les données de la statistique publique relevant de la loi de 1951⁷ sera certainement mis à profit. Par ailleurs, des accès permanents seront mis en place pour les administrations qui en ont un besoin constant.

2. Au VI de l'article 54 de la loi informatique et liberté dans sa nouvelle rédaction.

3. Le seul moyen que ces entreprises auraient pour accéder elles-mêmes directement aux données serait de « démontrer » que les traitements qu'elles mettent en œuvre rendent impossible toute utilisation des données pour promouvoir des produits de santé ou pour cibler des groupes d'assurés : ce qui est une condition très forte. Voir le nouvel article L1461-3 du code de la santé publique.

4. Ne serait-ce que parce que les personnes qui exécutent les traitements peuvent nécessairement voir les données.

5. Ibidem : « L'accès aux données est subordonné [...] à l'engagement du demandeur de communiquer [...] dans un délai raisonnable après la fin de la recherche, de l'étude ou de l'évaluation, la méthode, les résultats de l'analyse et les moyens d'en évaluer la validité. »

6. Principe du « need to know »

7. CASD : centre d'accès sécurisé aux données (CASD) du Genes (Groupe des écoles nationales d'économie et statistique)

S&S : Les chercheurs se plaignent souvent de devoir passer par des procédures bureaucratiques longues pour accéder aux données. Qu'en sera-t-il ?

AL : On ne peut pas organiser un accès « encadré » sans que les porteurs de projets, publics ou privés, aient à montrer qu'ils satisfont aux conditions prévues par la loi. L'exigence de confinement des données rendra sans doute un peu plus contraignantes les modalités d'accès aux données du PMSI notamment, qui étaient assez laxistes, mais on verra à l'usage si les comités consultatifs chargés d'instruire les demandes pour le compte de la CNIL permettent bien d'obtenir des délais acceptables.

S&S : L'accès aux données sera-t-il gratuit ?

AL : Pour les données en « open data », oui ; ce sera le cas aussi pour les données en accès encadré, lorsqu'il s'agira de travaux demandés par l'autorité publique, ou de recherches réalisées exclusivement pour les besoins de services administratifs (nouvel article L1461-5 du code de la santé publique). Pour les autres recherches, études ou évaluations, notamment celles émanant du secteur privé, la question de la tarification est à l'étude.

S&S : Un citoyen peut-il s'opposer à ce que les données de santé qui le concernent soient utilisées pour des études ou des recherches ?

AL : Le texte de la loi Informatique et Liberté modifiée prévoit l'obligation d'informer les personnes concernées des traitements des données qui les concernent, et de leur droit d'opposition (articles 56 et 57). En ce qui concerne le système national des données de santé, c'est auprès des caisses d'assurance maladie que pourra être exercé ce droit d'opposition⁸. Les garanties de confidentialité mises en place par la loi devraient rassurer les citoyens, mais chacun aura le droit de refuser que ses données servent aux traitements autorisés par la CNIL !

S&S : La loi de santé contient des dispositions relatives au « numéro de sécurité sociale », alias numéro d'inscription au répertoire (NIR), le répertoire en question étant le « répertoire national d'identification des personnes physiques » (RNIPP), tenu par l'Insee. Pouvez-vous nous expliquer l'intérêt de ces nouvelles dispositions ?

AL : Pour des raisons historiques remontant aux années 1970, l'utilisation du NIR est limitée de façon rigoureuse dans la loi Informatique et Libertés. Pour pouvoir utiliser le NIR, même sous forme « hachée », une administration ou un organisme de recherche devait obtenir un décret en Conseil d'Etat : autant dire que c'était hors de portée d'un projet de recherche ou d'évaluation ordinaire. Cette situation était particulièrement dommageable aux études en santé. En effet, dans le domaine de la sécurité sociale et de la santé, le NIR a été choisi comme identifiant depuis l'origine. Les grandes bases de l'assurance maladie ont été apurées de leurs doublons depuis les années 2000 par confrontation au RNIPP : c'est pourquoi le chaînage des informations relatives à une même personne y est de très bonne qualité. Dans les bases de données issues de ces systèmes et préparées pour la recherche, le NIR est remplacé par un « NIR haché » selon un algorithme qui empêche le retour en arrière et permet des appariements sécurisés. Le souhait de nombreux chercheurs en santé qui recueillent des données originales sur un échantillon de malades est de pouvoir apparier leurs propres données avec les données médico-administratives de ces mêmes patients. C'était pratiquement impossible jusqu'à présent à cause de la restriction à l'usage du NIR. Désormais, la CNIL pourra autoriser ce genre d'opération dans le cadre de « modules de bonne pratique ».

8. Article L1461-7 6° du code de la santé publique. Le droit d'opposition ne s'applique pas aux administrations publiques disposant de droits d'accès permanents aux bases médico-administratives.

S&S : Alors, un bel avenir est promis aux études de santé en France ?

AL : Je le crois : ce pays dispose, je l'ai dit, de bases de données médico-administratives exceptionnelles, et nous pouvons être assez fiers d'avoir mis en place un dispositif d'accès qui concilie au mieux toutes les exigences légitimes. Cela dit, sa mise en œuvre dépendra des moyens qui pourront lui être consacrés, et de la mobilisation effective des acteurs. A ce stade, le rôle de la CNIL va être décisif dans la mise en place des référentiels prévus par la loi.

Références

- [1] Rapport sur la gouvernance et l'utilisation des données de santé (« rapport Bras ») – La Documentation Française, octobre 2013
- [2] Rapport de la commission « Open data en santé » - La Documentation française, juillet 2014
- [3] Etude annuelle 2014 du Conseil d'Etat : le numérique et les droits fondamentaux – La Documentation française, septembre 2014
- [4] Dossier Solidarité et Santé n°64 : Données de santé : anonymat et risques de ré-identification – DREES, juillet 2015