# OUVERTURE DES DONNÉES DE SANTÉ – LA SAGA CONTINUE

## *AVAILABILITY OF HEALTH DATA – THE SAGA CONTINUES*

Par **Norbert PAQUEL* & Denis BERTHAULT***

**RÉSUMÉ**

Le débat sur l'ouverture des données de santé révèle plusieurs confusions. Ces données s'étendent désormais aux données médico-administratives, notamment le SNIIRAM (de l'Assurance maladie), puis les données des assurances complémentaires et, demain, celles issues de dispositifs médicaux et de services privés. Pour assurer ouverture et protection, la loi crée un Système National, entrepôt géant contrôlé par l'Etat, la CNAM, la CNIL et un institut (INDS) qui reste à créer. Les règles d'ouverture, visant santé publique et recherche, limitent l'accès des acteurs privés. Or, la demande croît rapidement. Le risque est de freiner l'ouverture et donc les avantages sanitaires et économiques attendus, et de ne pas pouvoir faire face aux services nouveaux. Une réflexion nationale, européenne, internationale s'impose.

**MOTS-CLÉS**

Données publiques, Ouverture, Données de santé, Définition, Quantified self, Anonymisation, SNIIRAM, NIR, SNDS, Conditions de réutilisation, Tarification, Europe.

***SUMMARY***

*Confusion dominates the debate on healthcare opendata in France. Mainly because, at a medical level, the data now includes administrative information from the SNIIRAM, from the mandatory Heath Insurance systems and, tomorrow, from medical devices and private services. A new statutory National System, controlled by Government, the National Health Insurance and other public entities has been created to host all the datas in a huge repository to ensure data openness and privacy protection. The reuse rules favour Public Health and Research and limit private actors access. In front of a growing demand, slowing data openness would limit the expected sanitary and economic gains and prevent being able to deal with new services. National, European and international reflections are much needed.*

***KEYWORDS***

*Open data, Health data, Definition, Quantified self, Anonymization of data, Reuse restrictions, Europe.*

L'espace social de la santé a, comme tous, ses rites et ses vocabulaires et les membres du club qui ne sont pas juristes finissent par confondre les règles du club – un des plus fermés – avec la Loi.

* Délégué général, EDISANTE et Directeur du développement des contenus en ligne, Lexisnexis
norbert.paquel@gmail.com ; denis.berthault@lexisnexis.fr

Aussi, le débat sur les données de santé, qui fait quelque bruit depuis quelques temps, révèle progressivement que le Roi est nu et qu'on ne sait guère de quoi l'on parle.

Qu'est-ce qu'une donnée de santé ? Qui la contrôle ou qui en est propriétaire ?

En fait, jusqu'ici, les données de santé étaient définies par leur créateur : était réputée telle toute donnée relative à une personne et produite par un professionnel de santé dans l'exercice de sa profession. On oubliait alors prudemment que la plus grande base au monde de telles données – le SNIIRAM de l'assurance maladie (1,2 Mds de feuilles de soins, toutes les sorties d'hôpital, les données de mortalité, etc.) – ne ressortait pas des mêmes règles que les donnée d'un dossier hospitalier, par exemple.

Or, pendant que le débat se développait, la frontière de la santé devenait floue et de multiples sources nouvelles apparaissaient : d'abord, les services médico-sociaux et de soins courants gèrent des données qui sont d'une part liées à l'état de santé, d'autre part révélatrices souvent de cet état ; ensuite ce sont toutes les données qui sont produites par les dispositifs médicaux installés autour de la personne ou implantés dans la personne ; ce sont aussi celles que le citoyen et ses proches créent eux-mêmes, du « quantified self » aux réseaux d'associations de patients ou en général dans les réseaux sociaux.

Il devient très difficile de tracer une frontière et de définir des textes communs pour un ensemble de plus en plus divers et vaste. On en revient donc sans doute à constater que toutes ces données, relatives à la personne, doivent avant tout relever des textes fondamentaux sur la protection de la vie privée. On en revient donc aussi à se poser la question de la spécificité des « données de santé » – au fond, doivent elles être mieux protégées que les autres données personnelles – et est-ce d'ailleurs possible ?

En même temps, la pression pour un accès aux données anonymisées se fait forte car chacun comprend que leur analyse doit être menée avec de nombreuses approches, à la fois pour améliorer le système et sa gestion, réduire les dysfonctionnements, faire progresser la recherche, permettre aux citoyens de se grouper, aux complémentaires et aux industriels de mieux faire leur travail, permettre la multiplication de services offerts par des grandes sociétés et une foule de start-ups.

Ces interrogations n'ébranlent guère des acteurs qui débattent du problème non à partir d'une réflexion sociétale mais en fonction de rivalités institutionnelles pour accéder à des gisements connus – au premier rang desquels le SNIIRAM – et à partir d'analyse des intentions des utilisateurs potentiels de données pourtant anonymisées. Les tensions opposent l'administration d'Etat, les Caisses et avant tout la CNAMTS, les complémentaires santé, les chercheurs, l'industrie pharmaceutique, les producteurs d'équipements médicaux. Et pendant ce temps là, Apple, Google, Facebook et d'autres collectent des informations et créent, en même temps, comme aussi Microsoft ou des start-ups, des services qui vont proposer des contrats de gestion de données de santé.

Dans cet univers complexe, d'aucuns prétendent asseoir un système fondé sur une protection nationale – au demeurant non coordonnée actuellement au niveau européen – et sur des intentions : l'accès aux entrepôts doit répondre à un intérêt public.

De toutes ces interrogations mais aussi, selon des critiques de plus en plus nombreuses, à partir avant tout du point de vue de l'Etat, les rédacteurs de la Loi de santé, ont tenté de tirer un Article – l'Article 47.

Au départ, le projet de Loi semblait annoncer une ouverture large des données, et deux tabous semblaient devoir être levés : l'interdiction d'utiliser le NIR (le N° de SS) dans les données de santé gérées par exemple dans un dossier d'hôpital et l'interdiction d'utiliser le NIR pour rapprocher des données, l'obligation de placer des barrières presque infranchissables pour les acteurs privés – qu'il s'agisse d'assureurs, d'industriels, de professionnels de santé ou d'associations de citoyens.

Le projet de Loi consacre un volumineux Article aux données de santé, mais force est de constater que de très nombreuses incertitudes subsistent tandis que des tendances fortes sont toujours à l'œuvre.

La première est la volonté de l'Etat (une fois réglées les relations avec l'assurance-maladie, traditionnellement difficiles) de constituer et accroître un gigantesque entrepôt protégé par de multiples contrôles. Même, de nouvelles données s'ajoutent ainsi au SNIIRAM, d'abord intégré à un Système national des données médico-administratives et maintenant, dans la version actuelle du projet, dénommé Système national des données de santé (SNDS). Manifestement, les responsables publics continuent de penser que l'Etat est le seul acteur fiable, honnête et puissant capable de gérer et protéger les données de santé – le périmètre pourra s'étendre et on comprend que les complémentaires n'apprécient guère que leurs données y soient incluses, sous la seule gestion de la CNAMTS et sous le contrôle d'un Institut National des Données de Santé à définir.

La seconde tendance est le maintient d'une méfiance envers l'acteur privé et en particulier l'acteur privé à but lucratif. Alors que les principes d'ouverture des données publiques ne peuvent autoriser de différences d'accès qu'en fonction du risque concernant des données personnelles, c'est ici le statut qui est le critère.

La troisième tendance est la crainte de l'ouverture. Elle apparaît dans l'insistance sur l'ouverture des données anonymes ne permettant aucune réidentification (il y a bien sûr les tarifs des médecins, qui n'ont aucune raison d'être qualifiées de données

de santé) opposée à la protection sévère des données permettant une réidentification indirecte. Or, aujourd'hui, même les enregistrements d'une base de type SNIIRAM, par exemple, ayant fait l'objet d'une double pseudonymisation, peuvent conduire à une réidentification par des acteurs mettant en jeu des moyens… déraisonnables. Ce type de barrière non précisé a toujours permis de fermer l'accès de façon arbitraire.

Ainsi, la conjugaison des résistances du gestionnaire (légitimées par la lourde responsabilité de la gestion d'une base géante multi-sources et par les coûts de traitements) se combinera à un système complexe pour freiner l'ouverture (comité d'experts, INDS, CNIL).

En réalité, et les différentes parties prenantes du débat que la ministre avaient organisé le relèvent, la Loi se préoccupe avant tout de la santé publique et de la recherche, telles qu'elles sont perçues et pratiquées par les acteurs publics. Ce n'est pas un mal en soi et on souhaiterait qu'effectivement les bases soient utilisées de façon plus efficace qu'elles l'ont été jusqu'ici, mais ce n'est pas de « l'open data ».

Les conditions d'accès du secteur privé sont restrictives car il faut que les projets présentent un intérêt public significatif et que la réalisation du traitement soit confiée à un organisme d'étude public ou privé accrédité. C'est aussi ce que confirment les finalités annoncées : information du public, des professionnels, des établissements, veille sanitaire, recherche, définition et évaluation des politiques de santé et de protection sociale, connaissance des dépenses. Les rédacteurs précisent que les données indirectement nominatives ne peuvent être utilisées que pour recherche ou étude dont la finalité ne soit pas manifestement et essentiellement commerciale. Ils précisent : « Quant à l'intérêt général, sa prise en compte paraît permettre voire impose des possibilités d'accès davantage filtrées pour des besoins de nature marchande, que pour des finalités du service public, elles-mêmes d'intérêt général ». C'est là mêler une vraie question de sécurité et une appréciation quasi morale sur les intentions des acteurs.

Enfin, la question de la redevance est posée – car elle a été réintroduite dans le texte par rapport au projet initial. Une fois encore, la redevance n'est pas établie sur des critères de qualité et suivi des données mais sur la nature des demandeurs. Les rédacteurs précisent d'ailleurs, prudents, que les directives européennes prévoient que « le service public et les services d'intérêt général au sens communautaire peuvent bénéficier de certains avantages proportionnés aux sujétions nées de leur mission de service public ». Bien évidemment, les problèmes de la réutilisation des données ne sont pas abordés.

Au total, on comprend mal comment les impacts en matière de démocratie sanitaire, d'avantages économiques et d'innovation pourront être aussi importants que l'annoncent les rédacteurs de la Loi.

Les données de santé dont disposent aujourd'hui la CNAMTS et les administrations sont effectivement un matériau essentiel pour le citoyen, le professionnel de santé, les acteurs du secteur. Leur ouverture offre de grandes opportunités. Elle est urgente, pour pouvoir entre autres faire face, précisément, à la collecte sauvage, massive mais souvent biaisée ainsi qu'aux offres non régulées – et difficilement régulables – en direction des citoyens eux-mêmes. C'est sur ce plan qu'il faut placer la réflexion, en travaillant à de vrais principes d'ouverture et de contrôle, concernant l'ensemble des acteurs. Il faut aussi parvenir à des accords européens et internationaux dans le domaine, ce qui est un chantier immense. ■