

I. LES ATTENTES DES PATIENTS ET USAGERS ET DE LEURS ORGANISATIONS.

Les patients souhaitent pouvoir disposer d'informations sur l'offre de soins (localisation, horaires d'accueil, prix des biens et services de santé, contenu des prestations ...) et sur la qualité des soins et du service rendu.

Ce besoin est d'autant plus aigu dans un contexte de massification des pathologies chroniques et des besoins en périphérie du soin qu'elles génèrent (information, orientation, accompagnement, soutien psycho-social).

Les associations de patients souhaitent, quant à elles, avoir la possibilité de :

- Disposer d'informations sur les inégalités géographiques et sociales de santé et l'efficacité des dépenses (consommations de biens et services de santé rapportées à un territoire de santé ou à un établissement). La lutte contre les inégalités sociales est affichée comme une priorité et l'accès aux données sur la santé peut incontestablement y contribuer en permettant de les mesurer de manière précise, d'une part, et en offrant aux associations qui se mobilisent pour leur réduction une occasion de plus grande pertinence dans la définition de leurs actions et de leurs lieux d'intervention, d'autre part. A cet égard, l'introduction d'une variable sociale dans les données collectées sur les assurés, par exemple le niveau d'instruction, constituerait un progrès considérable pour améliorer la connaissance disponible aux fins d'action plus pertinemment ciblées.
- Eviter de mener des enquêtes quantitatives coûteuses et se concentrer sur les enquêtes qualitatives, en vue des pertinences thématiques, de territoires et méthodologiques des actions à mener.
- Produire des services adaptés aux préoccupations de leurs publics.
- Améliorer leurs propositions dans le cadre de la démocratie sanitaire telle qu'elle est définie par la loi du 4 mars 2002.
- Jouer leur rôle de lanceur l'alerte sur des questions de sécurité sanitaire qui les préoccupent.
- Mieux défendre les intérêts des patients et usagers. Ainsi, à partir d'un accès plus ouvert aux données sur la santé, elles pourraient permettre à leurs publics, en général considérés « à risque aggravé de santé » de mieux accéder à l'assurance en leur garantissant notamment que les éventuelles exclusions ou surprimes seront évaluées au regard des données actualisées de survie et d'invalidité les concernant.

¹ Christian Baudelot, sociologue, et Christian Saout, administrateur du Collectif interassociatif sur la santé.

Les auteurs de ce document considèrent que les citoyens usagers du système de santé et leurs organisations sont actuellement tenus dans l'impossibilité de jouer leur rôle dans ces domaines, faute :

- de volonté suffisante des pouvoirs publics de mettre en œuvre l'Open Data dans le domaine de la santé. Les « jeux de données » aujourd'hui disponibles sont tellement appauvris qu'ils ne permettent que très marginalement de répondre aux attentes évoquées plus haut,
- de disposer des ressources humaines pour exploiter les bases de données, alors même que les organisations de patients et d'usagers pourraient, seules ou avec d'autres, produire des services innovants d'accès aux données de santé anonymisées².

Pour toutes ces raisons, il est indispensable de décider d'un Open Data en santé. Tout en prenant en compte le risque de ré-identification des personnes qui se pose de façon plus aigüe que dans d'autres domaines.

Il faut donc faire le pari d'un accès aux données de l'Assurance Maladie anonymisées mais aussi à tout un ensemble d'autres sources publiques de données de santé, notamment les Registres :

- pour permettre la plus juste information des patients et usagers, dans une visée de transparence de l'action publique et de rétablissement de l'équilibre entre les parties prenantes du système de santé,
- pour favoriser la création de services innovants afin de répondre aux attentes nouvelles des patients et usagers dans un esprit de transparence.

II. LES RISQUES DE RE-IDENTIFICATION.

Dès qu'il s'agit d'accéder à des données anonymisées, dans le cadre de ce qu'il conviendrait d'appeler un « Open Data régulé », sont pointés des risques de ré-identification, liés potentiellement à ces données ou par croisement avec d'autres.

21. Quels sont ces risques ?

De notre point de vue, il convient de distinguer deux niveaux de risques de ré-identification.

211. Un premier niveau que l'on qualifiera de risque faible est celui qui découle de l'utilisation des données issues du programme de médicalisation des systèmes d'information (PMSI) après acquisition de la base auprès de l'Agence technique de l'information sur l'hospitalisation (ATIH). En effet, l'acquisition de ces données est possible depuis 2005 après obtention d'une autorisation de la Commission nationale de l'informatique et des libertés (CNIL) laquelle précise, d'une part, l'objet des travaux effectués, et d'autre part, les sanctions encourues en cas de non-respect des obligations de sécurité et de divulgation des informations³.

² Au surplus, les techniques du Big Data pourront permettre demain à de très nombreux opérateurs d'accéder à la connaissance en santé, sans que les organisations dont les usagers et les patients soient en situation de bénéficier de ces techniques, aggravant encore l'écart entre ceux qui produisent des services et des soins de santé et ceux qui y ont recours.

³ 5 ans d'emprisonnement et 300 000 euros d'amende.

Dans ces hypothèses, le risque de ré-identification est faible car la CNIL impose déjà des limites en exigeant la transformation d'un certain nombre de données : mois de sortie de l'hôpital à la place de la date de l'hospitalisation, âge à la place de la date de naissance, code postal regroupé⁴ au lieu de code postal de la commune.

Cette méthode d'anonymisation semble suffisante puisque depuis près de dix ans aucune violation de la vie privée n'a été rapportée, malgré la circulation de près de 300 CD de la base envoyés aux différents demandeurs à des fins de recherches publiques ou privées, dont la presse « grand public » qui utilise ces données pour la construction des palmarès publiés chaque année par les plus grands hebdomadaires.

212. Un second niveau doit être qualifié de risque fort : c'est celui qui résulte du croisement des données natives du PMSI avec celles des données ambulatoires issues du Système national d'information inter-régimes de l'assurance maladie (SNIIRAM), ou d'autres bases.

Dans ces hypothèses, les données traitées sont plus précises : date de naissance, date d'hospitalisation, code postal de résidence, date du décès, toutes données appariables aux données de prescriptions, d'examens, de consultations, d'affection de longue durée, etc ... En cas d'intention malveillante, il semble possible de retrouver l'identité d'un patient.

22. Des solutions complémentaires peuvent être envisagées pour pallier la survenue des risques de ré-identification.

Avec les ambitions de l'Open Data en santé telles que nous les décrivons, nous voyons bien que la survenue de risques potentiels de ré-identification est plus forte. Quelles seraient alors les solutions complémentaires et/ou nouvelles à mettre en place pour limiter ce risque ?

221. Dans les situations de « risques faibles », outre les conditions déjà exposées, il conviendrait :

- de renforcer l'anonymisation des données fournies, en supprimant par exemple l'âge du patient et en le remplaçant par une tranche d'âge⁵,
- d'envisager un renforcement des audits périodiques pour s'assurer que les données sont stockées sur des serveurs sécurisés, pour vérifier les conditions d'accès à ces données par les opérateurs et l'adéquation entre les travaux déclarés dans l'autorisation de la CNIL et les travaux effectivement réalisés,
- d'envisager la déclaration préalable, lors de la demande d'utilisation, des objectifs des exploitations des données,
- d'envisager une publication régulière, au moins annuelle, d'un rapport décrivant précisément la liste des personnes morales et individuelles⁶ procédant à l'acquisition des bases du PMSI, ainsi que de leurs objectifs déclarés

En revanche, il serait contre-productif en termes d'analyses médicales, épidémiologiques et économiques de supprimer l'identifiant patient⁷ qui permet, par des appariements, de chaîner

⁴ 6000 codes « postaux » au sens PMSI contre 36 000 codes au sens « postal ».

⁵ De notre point de vue, l'impact négatif sur la finesse d'analyse serait très limité en échange d'un renforcement considérable de la protection de l'anonymat.

⁶ Sociétés privées, sociétés savantes, établissements publics, associations de patients, assureurs, journalistes ...

le patient dans le temps et ainsi de travailler sur les parcours et trajectoires de soins, les consommations de ressources, les inégalités territoriales et sociales de santé, etc, dont nous avons aujourd'hui tant besoin pour une plus grande efficacité du système de soins et pour une expression des attentes plus fidèle aux besoins ressentis.

222. Dans les situations de risque fort, nous proposons les évolutions suivantes :

- mise en œuvre des trois recommandations du point 221,
- renforcement des moyens humains des autorités de régulation, ce qui aurait par ailleurs pour effet de réduire les délais de réponse aux demandes d'autorisation,
- obligation de transparence sur les objectifs, les protocoles détaillés, les méthodologies d'analyse et les types de résultats attendus des études autorisées.

223. Des évolutions techniques peuvent y être jointes.

En effet, des techniques informatiques permettent aujourd'hui d'altérer partiellement des données de sorte qu'elle ne peuvent plus ouvrir à la ré-identification d'une personne.

224. Des évolutions normatives ne doivent pas forcément être écartées.

En effet, si, malgré les protections évoquées aux points 221, 222 et 223, on craint une évolution significative de la sélection des risques par exemple par les opérateurs de l'assurance complémentaire, cette possibilité devrait alors être prohibée par des dispositions adaptées du code pénal et/ou du code des assurances⁸.

De même si des craintes existent sur l'accès des industries de santé aux données, malgré l'interdiction de tout contact direct entre une industrie de santé et un patient⁹, une législation adaptée devrait alors être adoptée.

Enfin, l'information sur les sanctions en cas d'infraction aux règles de la protection des données doit faire l'objet d'une promotion générale plus systématique nonobstant la mention pour le bénéficiaire de ne pas enfreindre cette réglementation à l'occasion de la notification de cette autorisation.

III. LA GOUVERNANCE DES ACCES.

Elle est aujourd'hui le fait de la double intervention de l'Institut des données de santé (IDS) et de la CNIL. L'IDS assurant l'instruction des demandes et accordant des autorisations d'extractions des données et assurant l'instruction des demandes, et la CNIL, de son côté, vérifiant que les données sollicitées correspondent bien aux finalités exposées par le demandeur. S'y ajoute au cas par cas, l'intervention de comités consultatifs sur le traitement de l'information en matière de recherche dans le domaine de la santé (CCTIRS)¹⁰. Isolément,

⁷ Numéro unique anonymisé à 2 reprises (établissement puis agence régionale de santé) par la fonction FOIN à partir du numéro de sécurité sociale, la date de naissance et le sexe.

⁸ Article 225-3

⁹ Encore confirmé récemment par l'article L. 1161-1 du code de la santé publique pour ce qui concerne l'éducation thérapeutique.

¹⁰ La note remise lors de la dernière séance de travail propose de renforcer le rôle d'organe technique des CCTIRS alors même que ces CCTIRS sont composés de membres bénévoles, la note remise lors de la dernière séance de travail se propose de renforcer le rôle de ces

le Comité d'orientation et de pilotage d'information interrégimes de l'Assurance maladie (COPIIR) examine les demandes de l'Etat et de ses établissements bien que la loi¹¹ ne lui ait pas confié une telle mission.

Il est proposé de remplacer, partiellement tout au moins, cet édifice par un Haut conseil des données de santé dont la nature juridique et le périmètre des compétences ne sont pas encore précisés. Toutefois, en miroir des organismes portant l'appellation de « Haut conseil », il faut bien admettre que cela ne plaide pas pour qu'une telle instance soit considérée comme une autorité indépendante des pouvoirs publics mais plutôt comme une commission « à la main » de ces derniers.

En tout état de cause, nous ne voyons pas l'intérêt de créer un nouveau dispositif en lieu et place de l'IDS qui assure aujourd'hui l'instruction des demandes et accorde des autorisations d'extraction des données sous réserve que la CNIL ait vérifié le bien fondé des analyses et leur finalité. Nous considérons au contraire qu'il faut parier sur un renforcement de l'IDS :

- du point de vue des moyens humains et financiers pour lui permettre d'instruire les demandes plus rapidement, d'une part, et offrir aux associations des compétences afin d'effectuer des requêtes adaptées à leurs besoins d'information, d'autre part ;
- du point de vue des compétences afin qu'il puisse examiner tous types de demandes, y compris celles des organismes à but lucratif. Les organismes défendant les intérêts des patients et des usagers ne sont pas toujours des associations de la loi de 1901. Les demandes présentées aujourd'hui au sein du COPIIR seraient alors présentées demain devant l'IDS, il n'y a en effet pas de raison d'instaurer un régime d'exception les concernant.
- du point de vue du statut de l'IDS qui doit devenir une autorité indépendante de la puissance publique étant établi que des décisions favorables à des demandeurs sont aujourd'hui retenues à la demande de la puissance publique.

instances, ce qui ne manque pas de surprendre. En outre, quelle sera la composition de ces instances renforcées et comment seront instruits les liens d'intérêts entre leurs membres et les demandeurs ?

¹¹ Mais au contraire l'a limitée à la gestion et au renseignement de la base (loi de financement de la sécurité sociale pour 2009)