DONNÉES DE SANTÉ ET SOUVERAINETÉ NUMÉRIQUE

Pierre-François LAGET, *Médecin Responsable du Département d'Information Médicale, Centre Hospitalier de Lisieux*

Mai 2014. Article n°IV.5

Il est inutile de revenir sur les enjeux que représentent les données massives ("Big Data"), enjeux commerciaux autant que stratégiques. Le gouvernement français ne s'y est pas trompé, qui a inclus leur valorisation comme l'une des sept ambitions stratégiques définies par la Commission Innovation 2030 (1).

Les questions soulevées par ces données restent néanmoins nombreuses. Problèmes avant tout politiques et stratégiques:

- par la prééminence de fait acquise par les méga-acteurs américains du numérique,
- par les questions juridiques soulevées (à qui appartiennent les données individuelles?),
- par les possibles atteintes aux libertés publiques, renforcées par les différentiels de traitement juridique d'un pays à l'autre. Une donnée considérée comme privative par l'un pourra être publique, et donc commercialement exploitable, par l'autre.

Un exemple proposé par Pierre Bellanger (2, 3) résume bien l'ensemble des problématiques soulevées. Il concerne le monde des transports.
En effet la connexion des véhicules aux réseaux de données est déjà une réalité, même si, temporairement, elle ne concerne que des utilitaires ou des produits de haut de gamme. Ces données seront exploitées de façon de plus en plus extensive pour gérer les flux de circulation et donc optimiser le trafic, avec de multiples effets économiques positifs, ne serait-ce que pour limiter le temps perdu à bord des véhicules. L'étape suivante sera leur utilisation par les acteurs de l'urbanisme, lorsqu'il s'agira de définir de nouvelles voies de circulation ou des réaménagements urbains.
Mais il est une contrainte essentielle : l'intégration de cette connectivité aux outils mobiles. Donc le système d'exploitation des véhicules – qui constituera une part importante de leur valeur ajoutée – doit inclure, dès leur conception, une possibilité d'interaction très large avec les outils de connectivité dont disposent les utilisateurs.

Donc, concrètement, les systèmes d'exploitation des acteurs majeurs du Web seront en position de domination quasi-exclusive. Ils deviennent un standard de fait dans des domaines – les systèmes embarqués – où ils n'avaient pas a priori vocation à le devenir. On assiste ainsi à un

phénomène d'auto-renforcement de l'hégémonie, désastreux en termes de valeur ajoutée technologique et surtout de souveraineté.

Un tel exemple est multipliable sans limite. Il pose la question générale suivante: *comment définir des écosystèmes numériques assurant une interopérabilité générale tout en autorisant un contrôle souverain de l'état de droit qui est le nôtre?*

Précisément des éléments de réponse ont été apportés dans un domaine où la maîtrise des données est fondamentale: celui des données de santé. Des solutions simples ont été mises en oeuvre, et nous pourrions peut-être nous en inspirer dans d'autres espaces.

CARACTÉRISTIQUES GÉNÉRALES DES DONNÉES DE SANTÉ

Quelles que soient leurs finalités, ces données présentent trois caractères fondamentaux:

1) Elles ont un statut privé, reconnu par la loi, qui sanctionne les manquements à cet égard.
2) Ce statut privé connaît des exceptions légitimes, reconnues également par la loi.
3) Elles offrent la possibilité de servir l'intérêt général, notamment:
- en permettant des études de santé publique à large échelle;
- en fournissant les éléments permettant de définir une politique nationale en matière d'offre de soins et de financement de la santé.

Or ces caractéristiques sont générales. Elles peuvent être extrapolées à de multiples domaines, et, dans le cas de l'exemple choisi (les données relatives aux transports), le parallélisme est frappant:
- nos déplacements relèvent strictement de la vie privée;
- leur analyse est néanmoins nécessaire pour optimiser le trafic;
- certains acteurs légitimes peuvent pouvoir y accéder à titre exceptionnel dans un cadre établi par les autorités régaliennes.

Nous exposerons maintenant les principes fondamentaux mis en œuvre dans les données de santé, en nous limitant aux données produites par les établissements de soins qui sont de loin les plus complexes. Nous envisagerons ensuite des extensions à ce système.

PREMIER NIVEAU DE STRUCTURATION DES DONNÉES DE SANTÉ: L'ÉTABLISSEMENT PRODUCTEUR

Historiquement, le premier producteur de données a été le médecin indépendant, la notion de secret médical remontant à l'Antiquité. Par la suite le système s'est complexifié, avec la mise en œuvre de l'Assurance maladie puis la loi du 31 Juillet 1991 (articles L6113-7 et L6113-8 du code de la santé publique), qui impose aux établissements de santé l'évaluation de leur activité.

Dans les années suivantes vont se mettre en place les premiers outils permettant de recenser l'activité, et grâce à l'environnement mis en œuvre par le PMSI (Programme de Médicalisation des Systèmes d'Information) des bases de données intra-hospitalières recenseront un ensemble réglementairement normé contenant un ensemble de données relatives:
- aux informations administratives des patients,
- à leurs pathologies prises en charge,
- et aux actes médicaux qui leur sont délivrés.

La confidentialité de l'ensemble de ces informations sera couverte par un dispositif légal rigoureux, n'admettant que certaines dérogations strictement encadrées.

SECOND NIVEAU: LES ÉCHANGES NOMINATIFS INTER-ÉTABLISSEMENTS

La loi ne prévoit qu'un nombre très limité de cas où des informations nominatives peuvent être envoyées à l'extérieur de l'établissement. Concrètement le seul rencontré en pratique est celui des registres des cancers, bases de données permettant une analyse épidémiologique et un suivi précis de pathologies lourdes.

Ces registres sont présents au niveau régional, dans un nombre très restreint d'établissements (souvent les CHU). Des responsables nommément désignés se mettent en rapport avec les médecins responsables de l'Information Médicale des établissements de leur région (5). Les contraintes légales pesant sur ces registres sont évidemment fortes en matière de transmission et de stockage de ces données hautement sensibles:

- par les règles strictes édictées par la CNIL;
- par l'obligation pour ces registres d'être requalifiés tous les 4 ans par un Comité National co-présidé par l'INSERM et l'Institut de Veille Sanitaire (4).

Dans ces cas de figure, la sécurité des données est assurée:

- par l'encadrement légal associant différentes instances indépendantes;
- par les exigences strictes, vérifiables et auditables, en matière de qualité;
- par la définition précise des responsabilités;
- par la garantie que les personnes concernées ont, par leur formation, une connaissance approfondie des enjeux de ces données, ainsi que des systèmes de sécurisation et de chiffrement;
- par le fait que ces personnes sont peu nombreuses.

Nous sommes donc en présence d'un exemple abouti de qualité liée à la maîtrise des processus.

TROISIÈME NIVEAU: L'ENVOI DES DONNÉES SUR UNE PLATE-FORME NATIONALE ET LEUR EXPLOITATION PUBLIQUE

La sécurité des données est assurée par un dispositif performant assurant l'anonymisation des données tout en assurant leur caractère exploitable.

Le principe, né en grande partie des contraintes de sécurité imposées par la CNIL, en est le suivant (6):

- L'ensemble des informations relatives au séjour d'un patient (incluant les pathologies prises en charge, les actes effectués, les caractéristiques du séjour, les données techniques permettant la valorisation des séjours...) sont codées selon des règles nationales, et regroupées dans une base de données nominative, comportant les identifiants du patient. Cette base de données est conservée dans des conditions sécurisées au sein de l'établissement **et n'en sort jamais** (les seules personnes extérieures légalement habilitées à la consulter sont les médecins contrôleurs de l'Assurance maladie).
- Les normes applicables à cette base de données étant fixées par voie réglementaire, les éditeurs de logiciels équipant les matériels hospitaliers ont l'obligation de s'y soumettre afin de pouvoir transmettre les données produites.

- Cette base de données est anonymisée dans l'établissement en effectuant un hash sur trois identifiants invariants: le N° de SS ["NIR"]; la date de naissance; le sexe. Le hash, effectué selon un standard ouvert et reconnu, est bien sûr irréversible dans l'état actuel de nos connaissances.
- De plus, un certain nombre de précautions sont prises pour "flouter" certaines données sans nuire à leur exploitabilité. Ainsi les dates de début et de fin d'hospitalisation sont remplacées par les mois correspondants et la durée du séjour. De même la date de naissance est réduite à l'année de naissance, et les codes postaux des très petites communes sont compactés par des regroupements.
- Cette base de données anonymisée est transmise de manière sécurisée à une plate-forme nationale, l'ATIH (7).
- Au sein de l'ATIH, le hash est à nouveau hashé deux fois avec un système de salage secret.
- Cette base de données anonymisée, comportant les enregistrements de la vingtaine de millions d'hospitalisations annuelles en France, devient utilisable:
    - pour déterminer les montants financiers alloués à chaque établissement
    - et pour une multitude d'objectifs ayant trait à l'épidémiologie, à la santé publique, aux études médico-économiques de toute sorte... bref à l'ensemble de la politique de santé nationale.

De ce dispositif, au demeurant assez simple et peu exigeant en ressources, il résulte une conséquence capitale: même si il est impossible de remonter à l'identité du patient, les hash successifs sont reproductibles et comme par ailleurs ils sont appliqués à un ensemble d'invariants, **il devient possible de suivre les parcours successifs des patients, quelles qu'en soient les dates, mais sans connaître leur identité.**

IMPLICATIONS DE CE MODÈLE

Ce modèle est ainsi construit sur deux étages:
- un étage décentralisé, au niveau local, concentrant les données sensibles;
- un étage centralisé, national, ne contenant plus que des données anonymes et sans risques pour la vie privée.

Il en résulte:
- un cantonnement des données sensibles en de multiples lieux de stockage, placés sous le contrôle d'un nombre restreint de personnes dont les responsabilités et les processus de production sont strictement définis et maîtrisés;
- donc une étanchéité rendant illusoire une compromission à grande échelle;
- alors que parallèlement, les données susceptibles de compromission car centralisées n'offrent aucun intérêt pour un attaquant, puisqu'elles sont disponibles!

On voit donc immédiatement la supériorité d'un tel système par rapport à celui du Dossier Médical Personnel (DMP) qui centralise les données sensibles. Même si, comme c'est le cas, des précautions draconiennes sont prises pour assurer sa sécurité, une compromission peut être économiquement valorisable - rendant ainsi la menace tangible.

UN SCHÉMA GÉNÉRALISABLE

La transposition à l'exemple des transports est parfaitement possible:
- les informations recueillies par les véhicules restent à bord et ne peuvent être exploitées que par un processus sous contrôle régalien;

- les informations émises par les véhicules sont anonymisées en utilisant un système de hash d'invariants (numéro d'immatriculation, numéro de série...);
- toutes informations utiles au trafic peuvent alors être exploitées;
- les échanges entre les véhicules et les équipements mobiles grand public sont régis par une norme incluant un système de chiffrement afin d'éviter leur récupération par des acteurs non habilités.

AU TOTAL : LA SOUVERAINETÉ DES DONNÉES COMME AXE STRATÉGIQUE

Il devient indispensable de mener une politique vigoureuse visant à cantonner les données issues de la vie privée, car les enjeux sont considérables.

Enjeux économiques majeurs d'abord.

La protection des données peut parfaitement devenir un modèle économique à part entière, et la France dispose de toutes les compétences techniques requises pour proposer des solutions et des services visant à protéger la vie privée de l'usager et garantir le secret des affaires.

Ensuite parce que sans protection des données, la confiance dans l'économie numérique n'est pas possible. Rappelons que le commerce en ligne n'a pu se développer que lorsque les utilisateurs ont pu avoir confiance dans les dispositifs de paiement. Et que dire de l'internet des objets, qui pourront être autant de chevaux de Troie… Enfin on peut s'attendre à ce que l'économie « quaternaire », ou « économie des solutions » recèle des possibilités considérables de création de valeur tout en mettant la qualité et la durabilité au cœur des modèles économiques concernés (8). Et sans confiance, il n'est pas d'économie quaternaire possible.

Enjeux stratégiques ensuite.

La France a tout intérêt à se positionner comme **le pays des droits du numérique.** Si une volonté politique en ce sens est claire, lisible, mise en œuvre et surtout crédible, il y aura là un facteur d'attractivité venant s'ajouter à d'autres avantages comparatifs comme la qualité de nos infrastructures.

Si nous arrivons à mettre en place une telle politique, nous aurons créé un **marché de la souveraineté,** contre-pouvoir direct face à ce qui devient clairement un impérialisme numérique.

Nous aurons surtout contribué utilement à ce qui sera un des conflits majeurs dans les années à venir: retrouver une souveraineté sur nos données personnelles, sans laquelle le pouvoir des multinationales du numérique deviendra insupportable.

1) http://www.redressement-productif.gouv.fr/commission-innovation-2030-installee

2) Pierre Bellanger, *La souveraineté numérique*, Éditions Stock, 2014, 264 p. (ISBN 978-2918866213)

3) Interview par Xavier Delaporte ("Place de la Toile", France Culture, 12 Avril 2014)

4) http://www.invs.sante.fr/Espace-professionnels/Comite-national-des-registres/Liste-des-registres-qualifies-au-1er-mars-2014

5) Leurs rôles et missions sont définis par les articles R. 6113-1 à R. 6113-10 du code de la santé publique.

6) Versions successives de l'arrêté du 22 Février 2008 relatif "au recueil et au traitement des données d'activité médicale et des données de facturation correspondantes".

7) Agence Technique de l'Information sur l'Hospitalisation http://www.atih.sante.fr/

8) Pour une bonne synthèse sur ce sujet: M. Debonneuil, « Bienvenue dans l'économie des solutions », http://www.paristechreview.com, Avril 2014.