Big data, santé et droit : quelle combinaison idéale ?

Patricia Savin, Arnaud Tessalonikos
Avocats associés, DS Avocats

L'essentiel

La création de données massives (*big data*) en santé, particulièrement sensibles pour la protection de la vie privée de l'individu, implique de nouvelles perspectives notamment en matière de médecine prédictive et de médecine participative. Mais elle induit aussi de nouveaux risques relatifs à la collecte, la fiabilité, l'hébergement et l'exploitation de ces données. Au croisement entre le droit de la santé et le droit des nouvelles technologies, le dispositif légal existant se doit d'être renforcé et précisé.

Mots-clés : *big data* ; données de santé ; objets connectés ; télémédecine ; e-santé.

« L'humanité produit chaque année un volume d'informations numériques de l'ordre du zettaoctet, soit presque autant que le nombre d'étoiles dans l'Univers [1]. »

Particulièrement liée à l'essor d'une société connectée entièrement tournée vers l'information et la transparence, la création de données massives (*big data*) dans le domaine de la santé implique autant de nouvelles perspectives qu'elle met à l'épreuve les cadres juridiques traditionnels. Ainsi l'accroissement des offres d'applications numériques tendant à l'amélioration du bien-être de l'individu et de ses performances, *via* à titre d'exemple l'autosurveillance, constitue-t-il un vivier d'une importance considérable pour la recherche scientifique et médicale, mais aussi de nouveaux débouchés économiques. C'est la raison pour laquelle les firmes américaines IBM et Apple ont lancé une nouvelle plateforme, *Watson Health Cloud,* destinée à héberger les données anonymisées collectées par le logiciel Apple HealthKit et par le module Apple ResearchKit afin d'inclure les clients qui le souhaitent dans des recherches et essais médicaux [2]. L'exploitation de cette nouvelle manne économique n'est toutefois pas sans risque puisque ces données sont considérées par la réglementation française comme étant personnelles et sensibles. La Commission nationale de l'informatique et des libertés (Cnil) définit les données de santé comme « *des données individualisées recueillies auprès des professionnels de santé et relatives à leurs prescriptions et à leur pratique médicale* [3] ». Se pose la question d'actualiser cette définition au regard de l'essor des objets connectés. À l'instar de la Conférence nationale de santé [1], faut-il considérer que ces données constituent « *des informations sur l'état de santé et les maladies d'un individu ou d'une population donnée mais aussi sur des éléments qui peuvent déterminer l'état de santé et les maladies* [4] » ? La lecture de ces définitions fait apparaître à quel point ces données sont sensibles pour la protection de la vie privée de l'individu, droit fondamental protégé à la fois par des normes constitutionnelles [5] et convention-

1- La Conférence nationale de santé est un organisme consultatif placé auprès du ministre chargé de la santé. Elle exerce trois missions visant à formuler des avis sur l'amélioration du système de santé, élaborer chaque année un rapport sur le respect des droits des usagers du système de santé, contribuer à l'organisation de débats publics sur les questions de santé.

La question des modalités d'hébergement des données, de leur réutilisation à d'autres finalités, voire de leur appropriation par le fabricant des applications mobiles ou autre objets connectés est posée.

nelles [6]. L'accroissement de leur nombre et la tentation de les exploiter plus intensivement semblent donc ouvrir un large champ de possibilités mais également de risques. Dès lors, il devient nécessaire d'encadrer juridiquement cette exploitation de manière à aboutir à une protection efficace de la vie privée.

Le *big data* : un vivier riche de possibilités comme de risques nouveaux

Annonciateur d'une pratique médicale renouvelée et de nouvelles opportunités scientifiques, le *big data* constitue néanmoins un facteur de risques qu'il convient d'identifier afin de pouvoir les encadrer.

Des perspectives considérables liées à la collecte des données de santé

Le docteur Éric Baseilhac, directeur des affaires économiques du Leem (Les entreprises du médicament), l'un des principaux syndicats de fabricants de médicaments, considère que « *l'usage des* big data *dans la santé est une révolution qui va s'imposer et avoir un impact extrêmement structurant sur le système de santé en général* [7] ». Leur exploitation pourrait stimuler l'essor de la médecine prédictive puisqu'ils permettront, par l'amélioration de la connaissance des modes de vie, de recouper les données récoltées et d'analyser les risques de développement de certaines maladies chez les individus. C'est dire le potentiel de valorisation de ces données par ceux qui en assurent la collecte. La facilitation de l'accès et de la transmission des données de santé est également susceptible d'entraîner un changement de paradigme concernant la pratique médicale elle-même. Le code de la santé publique (CSP) a ainsi récemment vu pénétrer en son sein la notion de « télémédecine ». L'article L. 6316-1 du CSP, créé par une loi du 21 juillet 2009 [8] définit cette notion comme « *une forme de pratique médicale à distance utilisant les technologies de l'information et de la communication.* […] *Elle permet d'établir un diagnostic, d'assurer, pour un patient à risque, un suivi à visée préventive ou un suivi post-thérapeutique, de requérir un avis spécialisé, de préparer une décision thérapeutique, de prescrire des produits, de prescrire ou de réaliser des prestations ou des actes, ou d'effectuer une surveillance de l'état des patients* ». Le *big data* laisse donc entrevoir des possibilités importantes en termes de médecine participative, laquelle se fonde sur une démarche d'autoévaluation, ou *quantified self*, pratique d'ailleurs encouragée par le développement des applications de santé mobile fondées sur des données produites sur l'initiative des individus eux-mêmes dans le but de les comparer à différents indicateurs de santé. Surtout, le développement de telles pratiques médicales serait susceptible de réduire considérablement les coûts affectés à la santé au stade de la prévention des maladies et de la prise en charge des patients. Ce phénomène pose toutefois la question des modalités d'hébergement des données, de leur réutilisation à d'autres finalités, voire de leur appropriation par le fabricant des applications mobiles ou autre objets connectés. Si le champ de possibilités est immensément large, cela induit nécessairement un certain nombre de questionnements quant à la gestion des risques induits par la création et l'utilisation de ces données considérées comme « sensibles ».

Des risques liés à l'exploitation incontrôlée de ces données massives

Deux principaux risques sont particulièrement accrus du fait de l'utilisation des *big data*. En premier lieu, la perspective d'une utilisation massive des données de santé implique de se poser la **question de leur fiabilité**, tant sur une échelle macroscopique que microscopique. Pour certains types de dispositifs de collecte des données, la fiabilité constitue une exigence essentielle imposée par la directive 93/42/CEE du Conseil du 14 juin 1993 relative aux dispositifs médicaux. Selon celle-ci, les articles destinés à être utilisés chez l'homme à des fins de diagnostic, de traitement ou d'aide au diagnostic ou au traitement peuvent être, sous certaines conditions, considérés comme des dispositifs médicaux[2]. Sur la base de cette définition, un logiciel ou une application mobile pourrait ainsi se voir attribuer le régime applicable aux dispositifs médicaux s'il est destiné à être utilisé à des fins médicales. Cela supposerait que le matériel en question ne constitue pas une simple base de données et ait pour but de donner un résultat propre à un individu sur le fondement d'informations individuelles. La qualification d'un logiciel ou d'une application en dispositif médical entraîne dès lors la subordination de leur

2- Conformément aux termes de la directive 1er, §2.a qui qualifie de dispositif médical « *tout instrument, appareil, équipement, matière ou autre article, utilisé seul ou en association, y compris le logiciel nécessaire pour le bon fonctionnement de celui-ci, destiné par le fabricant à être utilisé chez l'homme à des fins : 1. de diagnostic, de prévention, de contrôle, de traitement ou d'atténuation d'une maladie ; 2. de diagnostic, de contrôle, de traitement, d'atténuation ou de compensation d'une blessure ou d'un handicap ; 3. d'étude ou de remplacement ou modification de l'anatomie ou d'un processus physiologique ; 4. de maîtrise de la conception ; et dont l'action principale voulue dans ou sur le corps humain n'est pas obtenue par des moyens pharmacologiques ou immunologiques ni par métabolisme, mais dont la fonction peut être assistée par de tels moyens* ».


© georgejmclittle

Pour certains types de dispositifs de collecte des données, la fiabilité constitue une exigence essentielle imposée par la directive 93/42/CEE du Conseil du 14 juin 1993.

mise sur le marché au respect de certaines exigences essentielles ayant trait notamment à leur faculté à atteindre les performances déterminées par le fabricant. Qu'en est-il toutefois des logiciels n'appartenant pas à ce régime ? Des questions importantes se posent quant aux performances de ces dispositifs en termes de fiabilité des informations collectées. Il faut dès lors se demander s'il ne serait pas opportun d'encadrer ces applications de santé par l'adoption de référentiels ou leur certification obligatoire par un organisme agréé. Si l'exploitation statistique des informations récoltées à l'échelle individuelle est envisagée comme une voie permettant également de prévoir ou d'analyser la survenance de certains phénomènes globaux, cette finalité prescriptive ne saurait exister qu'à condition que la fiabilité des données exploitées soit garantie au préalable. À ce titre, l'exemple du logiciel Google Flu Trends, destiné à évaluer la propagation dans le monde de la grippe H1N1, est particulièrement intéressant tant il permet de démontrer que le déploiement des possibilités offertes par l'exploitation des *big data* est intimement lié à la nécessité de disposer d'indicateurs et de données fiables[3]. L'essor des *big data* fait aussi craindre un risque d'atteinte au **droit de chacun au respect de sa vie privée** si des données sensibles sont captées, cédées, utilisées ou divulguées de manière incontrôlée. Or ce droit fondamental doit trouver à s'adapter au mouvement généralisé de transparence initié par l'open data, tendant à la divulgation, par l'État ou par les organismes publics, de données de santé anonymisées dans un souci de transparence, comme cela existe déjà en matière environnementale, et à des fins statistiques dans le but de favoriser les retours d'expériences quant aux politiques de santé. À ce titre, il faut mentionner le système national d'information interrégimes de l'Assurance maladie (Sniiram), créé en 1999 [9], qui rassemble des informations anonymisées issues des feuilles de soins et des prestations remboursées par les caisses d'assurance maladie. L'article 33 de la loi n° 2011-2012 du 29 décembre 2011 relative au renforcement de la sécurité sanitaire du médicament et des produits de santé [10] permet l'accès aux données de ce fichier pour la réalisation d'études de vigilance et d'épidémiologie en fonction des finalités poursuivies par ces études et de la contribution qu'elles sont susceptibles d'apporter par leur qualité scientifique. Ces évolutions importantes sont toutefois considérées comme insuffisantes par les acteurs qui réclament un accès plus poussé à ces bases de données. Un rapport Bégaud-Costagliola remis au ministère des Affaires sociales et de la Santé en septembre 2013 [11] sur la « *surveillance et la promotion du bon usage du médicament en France* » préconise ainsi un accès plus large aux données de santé. Cette voie est à l'étude, le ministère ayant ouvert un débat sur l'ouverture de données publiques de santé en novembre 2013.

L'encadrement juridique de l'exploitation des données de santé par la Cnil et le code de la santé publique

Des garanties générales sont prévues par la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, dite loi informatique et libertés[4] ainsi que par le code de la santé publique afin d'assurer une place centrale à la personne concernée par l'utilisation des données de santé. Ce cadre juridique met également en place des normes spécifiques destinées à capter l'activité des hébergeurs de données de santé.

3- Selon une étude du magazine *Science* (vol. 343, 14 mars 2014, www.sciencemag.org), le Google Flu trends, dont les indicateurs étaient fondés sur les mots clés de recherche, aurait surestimé, de manière récurrente, la propagation de la grippe H1N1 puisque ses prévisions aurait dépassé de 50 % les estimations du centre américain de contrôle et de prévention des maladies (*Center for disease control and prevention*, CDC), qui elles étaient basées sur les motifs de visites médicales.
4- La loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés a été refondue par la loi du 6 août 2004 pour tenir compte des modifications apportées par la directive 95/46/CE du Parlement européen et du Conseil du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et la directive 2002/58/CE du Parlement européen et du Conseil du 12 juillet 2002 concernant le traitement des données à caractère personnel et la protection de la vie privée dans le secteur des communications électroniques.

Les garanties générales prévues par la loi informatique et libertés et le code de la santé publique en matière de traitement de données numériques

La loi informatique et libertés met en place un régime de protection des données à caractère personnel et d'encadrement de leur « *traitement*[5] » dans une base de données. Concernant les données considérées comme sensibles, son article 8-I prévoit qu'il est interdit « *de collecter ou de traiter des données à caractère personnel qui font apparaître, directement ou indirectement, les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses ou l'appartenance syndicale des personnes, ou qui sont relatives à la santé ou à la vie sexuelle de celles-ci* ». Dans le cas des données de santé, les traitements peuvent toutefois être autorisés dans des conditions limitativement énumérées par ce même article, notamment lorsque la personne concernée a donné son consentement exprès ou lorsque le traitement porte sur des données à caractère personnel rendues publiques par la personne concernée [12]. De même, peuvent être autorisés les traitements nécessaires à des fins de médecine préventive, de diagnostics médicaux ou d'administration de soins ou de traitements lorsqu'ils sont mis en œuvre par un membre d'une profession de santé, ou par une autre personne soumise à l'obligation de secret professionnel, ainsi que les traitements nécessaires à la recherche dans le domaine de la santé [13]. Dans le but de garantir une utilisation transparente et consentie des données de santé, la loi permet aux personnes concernées par le traitement de leurs données de bénéficier de droits particuliers tenant à l'accès au fichier (ce droit étant également reconnu par l'article L. 1111-7 du CSP[6]), à la rectification et à la suppression des données collectées [14] ainsi qu'à l'opposition [15] contre le traitement informatique dont elles font l'objet. En tout état de cause, la personne devra être informée de la finalité du traitement, les données ne devant pas être utilisées de manière incompatible avec celle-ci [16]. L'usage ultérieur de données à des fins statistiques ou de recherche scientifique ou historique est toutefois considéré comme compatible avec ces finalités [17]. Enfin,

> Du big data au Big Brother, il n'y a qu'un petit pas pour l'Homme, qui ne doit pas être accompli. Il faut prévoir et faire respecter des garanties juridiques en ce sens.

afin de garantir le « *droit à l'oubli* » des personnes concernées par le traitement de leurs données personnelles, la loi prévoit que les données ne sont conservées sous une forme « *identifiante* » que pendant une durée strictement nécessaire aux finalités pour lesquelles elles sont collectées et traitées [18]. Concernant plus spécifiquement les traitements de données de santé, certaines garanties spécifiques, prévues par le code de la santé publique, s'ajoutent aux garanties générales prévues par la loi informatique et libertés. L'article L. 1110-4 du CSP rappelle, tout d'abord, les principes fondateurs posés par la loi informatique et libertés en son article premier. Ce préalable posé, le code de la santé publique met en place d'importants dispositifs de partage d'informations entre professionnels et établissements de santé, comme le dossier pharmaceutique (DP) susmentionné et le dossier médical personnel (DMP), prévu aux articles L. 1111-14 et suivants du CSP, reconnaissant la possibilité pour les professionnels de santé, quel que soit leur mode d'exercice, de partager les données de santé de leurs patients, sous réserve de l'autorisation de ceux-ci [19]. Afin de garantir la fiabilité et l'accès limité aux données de ces deux fichiers, des mécanismes d'authentification sont mis en place concernant à la fois le patient et les professionnels de santé. Un identifiant de santé des bénéficiaires de l'assurance maladie est créé [20] et l'utilisation de la carte de professionnel de santé est rendue obligatoire pour permettre l'identification des professionnels de santé [21].

La création d'un régime spécifique à l'activité d'hébergeur de données de santé

La loi du 4 mai 2002 relative aux droits des malades [22] introduit, au sein de l'article L.1111-8 du CSP, une procédure d'agrément des hébergeurs de données de santé à caractère personnel afin de garantir la sécurité de celles-ci lorsqu'elles ne sont pas hébergées par le professionnel ou l'établissement de santé. Cet agrément, dont les modalités ont été fixées par le décret du 4 janvier 2006 [23], est délivré pour une durée de trois ans par le ministre chargé de la Santé, qui se prononce après avis de la Cnil et du comité d'agrément créé auprès de lui [24]. L'activité d'hébergement des données de santé déposées par les professionnels, les établissements de santé ou la personne concernée, et recueillies à l'occasion des activités de prévention, de diagnostic ou de soin se trouve ainsi fortement réglementée. S'agissant de l'information de la personne

5- Aux termes de l'article 2 de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, le traitement désigne « *toute opération ou tout ensemble d'opérations portant sur de telles données, quel que soit le procédé utilisé, et notamment la collecte, l'enregistrement, l'organisation, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, ainsi que le verrouillage, l'effacement ou la destruction* ».

6- L'article L. 1111-7 du code de la santé publique dispose que « *toute personne a accès à l'ensemble des informations concernant sa santé détenues, à quelque titre que ce soit, par des professionnels et établissements de santé, qui sont formalisées ou ont fait l'objet d'échanges écrits entre professionnels de santé* […] »..

concernée par le traitement quant aux modalités de conservation et d'utilisation de ses données de santé, les dispositions de l'article L. 1111-8 mettent en place une obligation légale imposant à l'hébergeur des données d'établir un contrat avec le patient concernant la prestation d'hébergement. Les sous-traitants de l'hébergeur ayant accès aux données de santé à caractère personnel devront également être déclarés et apporter un niveau de garantie équivalent à celui de l'hébergeur principal [25]. Néanmoins, les professionnels et établissements de santé sont autorisés à échanger des données de santé sans le consentement exprès de la personne concernée lorsque l'accès y est limité au professionnel de santé ou à l'établissement de santé qui les a déposées [26]. Dans ce secteur extrêmement régulé, au croisement entre le droit de la santé et le droit des nouvelles technologies, la Cnil semble œuvrer en faveur d'une simplification de la règle de droit par le biais d'une mise en ligne de fiches pratiques à destination des professionnels de santé et des hébergeurs de données sensibles. Des précautions élémentaires importantes y sont référencées concernant les garanties de confidentialité et d'intégrité des données. Il peut être intéressant pour l'hébergeur de données de santé de s'en imprégner afin de tenter d'éviter un risque d'engagement de sa responsabilité qui peut être envisagée sur le terrain contractuel ou délictuel[7], et pénal[8]. L'engagement de sa responsabilité s'en trouvera d'autant plus facilité que les intérêts qu'il est censé préserver sont fondamentaux.

➲ Si une ouverture des données de santé à des fins de recherche semble être largement acceptée par la communauté scientifique, se pose évidemment le problème de l'anonymisation des données et de l'éventuelle ré-identification des personnes concernées. De même, comment garantir que les données, communiquées par l'individu lui-même ou par l'État, ne seront pas ultérieurement détournées de leur finalité première, à des fins commerciales par exemple ? Afin d'assurer l'encadrement de leur utilisation et surtout leur maîtrise par la personne concernée, le dispositif légal existant se doit d'être renforcé et précisé. Les technologies numériques accompagnent de longue date le développement de la médecine. La connexion des objets entre eux, *via* l'internet, couplée à des possibilités d'acheminement et de stockage de l'information sans précédent, vont nous permettre de disposer d'une efficacité encore améliorée des soins. Or, il est clair que, du *big data* au Big Brother, il n'y a qu'un petit pas pour l'Homme, qui ne doit pas être accompli. Il convient donc de prévoir et de faire respecter des garanties juridiques en ce sens. ■

Références

1- « Le *Big Data* : un enjeu économique et scientifique », par Fabrice Demarthon le 15 novembre 2012. Disponible sur https://lejournal.cnrs.fr/big-data#footnote1_5wxlfe9.
2- *Le Monde*, « IBM s'associe à Apple pour analyser les données de santé », par Jean-Baptiste Jacquin, le 14 avril 2015.
3- Cnil, 4 février 1997, délibération n° 97-008 portant adoption d'une recommandation sur le traitement des données de santé à caractère personnel.
4- CNS, avis, 19 oct. 2010, sur les données de santé informatisées : www.sante.gouv.fr.
5- Article 2 de la Déclaration des droits de l'Homme et du citoyen.
6- Article 8 de la Convention de sauvegarde des droits de l'Homme et des libertés fondamentales.
7- « E-santé : jusqu'où ira le *Big Data* pour nous soigner ? », par Adeline Raynal, publié le 23 mars 2015, http://www.spotwork.fr.
8- Article 78 de la Loi n°2009-879 du 21 juillet 2009 portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires.
9- Loi n° 98-1194 du 23 décembre 1998 de financement de la Sécurité sociale pour 1999, article 21. Parution au *Journal officiel* du 27 décembre 1998.
10- Article L. 5121-28 du code la santé publique.
11- Rapport du 16 septembre 2013, établi par MM. Bernard Bégaud et Dominique Costagliola dans le cadre de la mission « pharmacosurveillance » confiée par la ministre des Affaires sociales et de la Santé, Marisol Tourraine.
12- Article 8-II de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.
13- *Ibid*.
14- Article 40 de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.
15- Article 56 de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.
16- Article 6, 2° de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.
17- Article 6, 2° de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.
18- Article 6, 5° de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.
19- Article L. 1111-15 du code de la santé publique.
20- Article L. 1111-8-1 du code de la santé publique.
21- Article L. 1110-4 du code de la santé publique.
22- Article 11 de la loi n° 2002-303 du 4 mars 2002 relative aux droits des malades et à la qualité du système de santé.
23- Décret n° 2006-6 du 4 janvier 2006 relatif à l'hébergement de données de santé à caractère personnel et modifiant le code de la santé publique (dispositions réglementaires).
24- Article R. 1111-15 et R.* 1111-10 du code de la santé publique.
25- Article L. 1111-8 alinéa 10 du code de la santé publique.
26- Article L. 1111-8 alinéa 5 du code de la santé publique.

7- L'article 9 du code civil prévoit en effet que le manquement à l'obligation de confidentialité des données personnelles constitue un manquement à une obligation de moyens renforcée (Crim., 30 octobre 2001, n° 99-82136). Ce manquement ne nécessitant pas de rapporter la preuve d'un préjudice (Civ. 1re, 28 avril 2011, n° 10-17909), la responsabilité délictuelle de l'hébergeur en cas de divulgation des données ne s'en trouve que facilitée.

8- Les hébergeurs agréés sont en effet astreints au secret professionnel dans les conditions et sous les peines prévues à l'article 226-13 du code pénal. De même, au titre de l'article 226-21 du code pénal, « *le fait, par toute personne détentrice de données à caractère personnel à l'occasion de leur enregistrement, de leur classement, de leur transmission ou de toute autre forme de traitement, de détourner ces informations de leur finalité telle que définie par la disposition législative, l'acte réglementaire ou la décision de la Commission nationale de l'informatique et des libertés autorisant le traitement automatisé, ou par les déclarations préalables à la mise en œuvre de ce traitement, est puni de cinq ans d'emprisonnement et de 300 000 euros d'amende* »