

► En matière de santé comme dans d'autres secteurs, une masse croissante de données numérisées provenant de diverses sources est disponible et exploitable. C'est l'un des domaines où le potentiel du *Big data* apparaît très prometteur, avec de multiples innovations au bénéfice des patients et du système (accélération de la recherche et développement, connaissance des maladies, des facteurs de risque, médecine personnalisée, aide au diagnostic et au traitement, rôle accru des patients, pharmacovigilance, etc.), même si des inquiétudes s'expriment aussi vis-à-vis des impacts sociétaux, économiques et éthiques que le recours croissant aux algorithmes et à l'intelligence artificielle pourrait induire. Développer l'usage de ces données constitue un objectif stratégique de tous les systèmes de santé, et de ce point de vue le Système national de données de santé (SNDS) constitue pour la France un patrimoine intéressant, mais qui demande à être complété et enrichi. ◀

Données de santé, de quoi parle-t-on ?

Le nouveau règlement général sur la protection des données (RGPD), qui s'applique dans tous les pays de l'Union européenne depuis mai 2018, donne pour la première fois une définition des données de santé, entendues comme « les données à caractère personnel relatives à la santé physique ou mentale d'une personne physique, y compris la prestation de services de soins de santé, qui révèlent des informations sur l'état de santé de cette personne ». C'est une définition large, incluant « toute information concernant, par exemple, une maladie, un handicap, un risque de maladie, les antécédents médicaux, un traitement clinique ou l'état physiologique ou biomédical de la personne concernée, indépendamment de sa source » [1]. Les données de santé peuvent en effet provenir de multiples sources :

Numérique et santé (1) Les données de santé

Dominique Polton



Institut national des données de santé (INDS),
19, rue A. Croquette,
94220 Charenton, France.
dominique.polton@indsante.fr

- elles peuvent être produites dans le cadre des soins que reçoivent les patients : il s'agit notamment des informations contenues dans les dossiers électroniques des médecins, des autres professionnels ou des hôpitaux, des données générées par des dispositifs médicaux connectés, par des systèmes automatisés de tests diagnostiques, analyses de laboratoire, imagerie numérisée, etc. ; il faut y ajouter aujourd'hui les tests génétiques, en plein développement grâce aux progrès faits dans le séquençage du génome, et qui représentent d'énormes volumes de données¹ ;

- les données administratives recueillies à des fins de gestion du système (facturation, remboursement des soins, enregistrement des décès, etc.) constituent aussi une source d'information sur la santé des individus ; en France, ces bases de données, qui couvrent l'intégralité de la population, contiennent ainsi des informations sur les médicaments délivrés, les motifs d'hospitalisation, les actes médicaux réalisés, etc.

- des recueils d'informations spécifiques auprès des patients peuvent également être organisés dans le cadre d'enquêtes, d'études de cohortes ou d'essais cliniques.

Mais au-delà de ces processus de collecte très encadrés, le développement de multiples services numériques, la santé mobile, l'internet des objets génèrent un accroissement exponentiel du volume des données produites, qu'il s'agisse de plateformes de prise de rendez-vous en ligne, de systèmes pour notifier des effets indésirables, d'applications mobiles pour mieux suivre son traitement, de sites de ventes en ligne de médicaments, de réseaux sociaux développés par des communautés de patients échangeant des informations sur leur maladie, leur

¹ Pour donner une idée des volumes, *The Cancer Genome Atlas (TCGA)*, projet de recherche du *National cancer institute*, met à disposition des chercheurs une base de données de 2,5 petaoctets (1 million de milliards d'octets) sur les mutations génétiques des tumeurs de 11 000 patients.

traitement et leur expérience (comme *PatientLikeMe* aux États-Unis, ou *Carecity* en France), etc. Et ces sources de données débordent largement la sphère de la maladie et du soin, avec les objets connectés et des applications de bien-être ou d'activité sportive, ou avec l'exploitation des traces numériques de l'activité des individus sur internet ou dans les réseaux sociaux généralistes.

Ainsi, dans le domaine de la santé comme dans de nombreux autres domaines de l'activité humaine, la généralisation de la numérisation et l'augmentation continue des possibilités de stockage et de traitement de l'information permettent d'exploiter d'énormes volumes de données issues de sources diverses, structurées (diagnostics codés, résultats de tests, remboursements de soins...) ou non structurées (comptes rendus d'hospitalisation, échanges sur les réseaux sociaux, etc.).

Big data et santé : les bénéfices attendus

De multiples bénéfices sont attendus de la collecte et de l'analyse de ces données à grande échelle, et les médias se font régulièrement l'écho des promesses d'une « médecine du futur » portée par la transformation numérique, le *Big data* en santé et les algorithmes d'intelligence artificielle.

Avant d'en venir à ces perspectives, qui suscitent beaucoup d'espoirs mais aussi de craintes, on évoquera tout d'abord des usages des données de santé numérisées qui existent déjà aujourd'hui et qui sont des leviers de transformation dans plusieurs registres : l'organisation des soins autour du malade ; l'évolution du rôle des patients ; la santé publique et l'optimisation du système de soins ; la production de connaissances scientifiques.

L'organisation des soins autour du patient

Avec le vieillissement de la population et le développement des pathologies chroniques, les systèmes de santé sont tous confrontés au même défi : organiser des parcours de soins fluides, sans rupture, avec des interventions coordonnées des professionnels de soins et éventuellement des aidants professionnels ou familiaux. Cette coordination, reconnue aujourd'hui comme un élément essentiel de la qualité de la prise en charge des patients chroniques, ne va pas de soi dans des organisations souvent cloisonnées (entre l'hôpital et la médecine de ville, entre le sanitaire et le social, etc.). Pour améliorer cette coordination en facilitant les échanges d'informations entre les professionnels qui interviennent auprès du patient, beaucoup de pays, dont la France, ont développé des projets de dossiers électroniques partagés². Au-delà de ce partage au sein de l'équipe soignante, la circulation de données numérisées ouvre la voie à d'autres avancées qui représentent un réel bénéfice pour les patients. Par exemple, l'utilisation de balances connectées permet de repérer, chez des patients âgés souffrant d'insuffisance cardiaque, le signe d'alerte que représente une

prise de poids importante et rapide, et de déclencher une intervention [2]. C'est un exemple parmi beaucoup d'autres d'un usage qui, grâce au traitement continu de données de santé pertinentes – qu'elles soient issues d'objets connectés ou transmises par les individus – peut permettre de réduire considérablement le délai de réponse du système de soins. Et les gains pour les patients peuvent être considérables : ainsi l'expérimentation, par un centre anti-cancéreux français, d'un dispositif permettant aux patients souffrant de cancer du poumon de renseigner en continu leur oncologue sur leurs symptômes par la voie d'une application web, ces informations étant traitées avec un algorithme pour détecter les rechutes, a conduit à un gain de survie de sept mois en moyenne : c'est un gain supérieur à ceux qu'apportent nombre de traitements innovants très onéreux³ [3].

Un autre bénéfice pour le patient est de réduire les pertes de chances liées à l'éloignement géographique, par l'accès à distance à une expertise spécialisée. Par exemple, dans plusieurs régions, des plateformes de télémédecine ont été organisées pour permettre à des médecins traitants, des infirmiers libéraux ou des établissements de santé de faire appel à des experts en plaies chroniques ou complexes⁴ lorsqu'ils rencontrent des difficultés dans la prise en charge de leurs patients, notamment à domicile.

Les médecins disposent aussi de nouveaux outils logiciels, qui en analysant les données de leurs patients au regard du corpus de connaissances scientifiques disponibles, peuvent les aider dans leurs décisions diagnostiques et thérapeutiques.

Ce sont là quelques exemples montrant que, dès à présent, la circulation et l'exploitation des données de santé peuvent permettre d'améliorer les soins dispensés.

La transformation du rôle des malades dans le système de santé

Au-delà de ces usages dans le cadre de la relation de soins, le numérique a des impacts plus systémiques et sociétaux. La circulation et le partage des données de santé ont en effet fortement contribué à bousculer les lignes de partage anciennes et les positions d'autorité entre soignants et soignés, et favorisé une des mutations fondamentales que connaissent les systèmes de santé depuis deux ou trois décennies, qui est l'évolution de la place des malades et des usagers.

² Avec des modalités et des fortunes diverses : ainsi le Royaume-Uni a abandonné au début de la décennie un programme ambitieux de dossiers médicaux informatisés initié en 2002. En France, le bilan mitigé du projet de dossier médical partagé (DMP) lancé en 2004 a conduit à le réorienter en 2016 ; un nouvel essai est en cours dans huit départements pilotes, qui semble aujourd'hui plus prometteur.

³ Cette expérience sera exposée dans un prochain article de la série.

⁴ Notamment escarres, ulcères de jambes, plaies du pied diabétique, plaies cancéreuses, qui touchent près de deux millions de personnes en France, en particulier des personnes âgées.

Un levier majeur de cet « *empowerment* », pour utiliser le terme anglo-saxon largement repris en France, a été le partage des données entre patients et le développement rapide de communautés en ligne mettant en commun leurs connaissances et leur expérience. Elles permettent non seulement de rompre l'isolement lié à la maladie et de s'entraider, mais aussi de construire un savoir collectif permettant d'améliorer les traitements⁵.

Les patients sont mieux informés sur leurs pathologies mais aussi sur la qualité de l'offre de soins ; ils peuvent faire des choix plus éclairés, renforcer leurs compétences dans la gestion de leur maladie. Ils ont de plus en plus d'outils pour cela, comme des applications leur permettant d'avoir accès à leur dossier médical, de calculer leurs doses d'insuline s'ils sont diabétiques, de ne pas oublier leurs médicaments, etc.

Des bénéfices collectifs pour la santé publique et l'amélioration du système de santé

Certains problèmes ont fait prendre conscience du potentiel extraordinaire des bases de données de l'assurance maladie⁶ pour la pharmaco-surveillance : à un moment où les signaux de pharmacovigilance disponibles faisaient encore débat en France, deux semaines de travail ont suffi pour mettre en évidence un triplement des hospitalisations pour les patients ayant été exposés à certains médicaments [4]. Depuis, ces données sont régulièrement exploitées sur des questions touchant à la sécurité des produits de santé. Elles ont permis de montrer, par exemple, les risques liés aux pilules de dernière génération. C'est un exemple des bénéfices collectifs qu'il est possible de tirer de ces données de santé agrégées à l'échelle de larges populations, et l'on peut en citer beaucoup d'autres :

- surveillance sanitaire, détection des risques, prévisions d'épidémies ; analyse de l'efficacité des traitements en vie réelle, pour faire évoluer les recommandations et améliorer la pertinence des soins ;
- conception de nouveaux modes de rémunération prenant en compte des indicateurs de qualité ou de performance ;
- analyse des performances du système de santé, de la conformité des soins aux recommandations de bonne pratique, des inégalités sociales et spatiales, détection des progrès possibles en termes d'accès, de qualité ou d'efficacité des soins ;
- segmentation des populations pour concevoir des programmes d'intervention ajustés aux besoins des différentes catégories de patients ;
- échanges entre praticiens ou établissements de santé pour comparer leurs pratiques sur la base de données partagées, ce qui est un puissant levier d'amélioration continue de la qualité des soins.

L'accès aux données porte aussi des enjeux de transparence pour les citoyens et de débat démocratique sur le fonctionnement du système

⁵ L'exemple de Renaloo, association de malades du rein, est très illustrative d'une action favorisée par les outils digitaux. Elle a démarré par un blog où la fondatrice a raconté sa greffe de rein, puis le site s'est enrichi d'un forum permettant l'échange entre patients, s'est transformé en 2008 en une association qui a organisé les États généraux du rein en 2012 et 2013, et mené à cette occasion une enquête auprès de 8 600 personnes greffées ou dialysées. L'association est depuis très active, et a obtenu des engagements ministériels, à la formalisation de plusieurs recommandations de la Haute autorité de santé, et à une auto-saisine de la Cour des comptes sur l'organisation des soins aux malades en insuffisance rénale chronique terminale.

⁶ Voir la description de ces bases de données dans le chapitre « La situation de la France en termes de données de santé ».

de santé, participant en cela d'un mouvement plus général d'*open data* (données libres).

La production de connaissances scientifiques

L'explosion des données disponibles est aussi une chance pour la recherche, et l'on attend de leur exploitation une accélération des progrès de la connaissance, sur les facteurs de risque des maladies, leur physiopathologie, les marqueurs permettant un diagnostic précoce, etc., notamment par la possibilité de croiser des données cliniques, biologiques et génomiques avec d'autres données personnelles ou d'environnement.

Au passage, au-delà de l'accroissement quantitatif, on assiste aussi à une transformation des modalités de cette production de connaissances : de nouveaux modèles de recherche collaborative se développent autour de données partagées (comme le projet Epidemium, lancé en partenariat par La Paillasse et le laboratoire Roche, avec une plateforme *open data* mettant à la disposition de tous ceux qui souhaitent participer des données ouvertes disponibles à l'échelle nationale, européenne et internationale)⁷

[18] (→) ; des communautés de patients s'inscrivent dans cette dynamique en montant des plateformes de recherche participative, fondées sur des données renseignées par eux⁸. À côté des essais cliniques, qui sont encore considérés aujourd'hui comme le « *gold standard* », la recherche biomédicale sera de plus en plus nourrie par des données « en vie réelle » collectées sur de grandes cohortes d'individus⁹, et exploitées avec de nouvelles méthodologies.

Big data, algorithmes, intelligence artificielle : vers la médecine du futur ?

Le potentiel des données massives de santé suscite beaucoup d'espoirs pour l'avenir, à commencer par une vie prolongée en bonne santé, grâce à une « médecine du futur » qui ne ferait que s'amorcer aujourd'hui. Cette nouvelle médecine, baptisée il y a une dizaine d'années médecine des 4 P, serait plus *prédictive*, grâce à la connaissance des facteurs de prédisposition (qu'ils soient génétiques ou liés aux comportements et à l'environnement du patient), plus *préventive* grâce à ces connaissances, plus *personnalisée* avec des straté-

(→) Voir le Repères de M. Meyer, page 473 de ce numéro

⁷ Pour plus détails : <http://www.epidemium.cc>

⁸ Comme MoiPatient, lancée par l'association Renaloo dans le champ des maladies rénales et destiné à être rapidement étendu à d'autres pathologies (site web : <http://www.moi-patient.fr>)

⁹ Par exemple, la filiale Verily de Google va suivre pendant quatre ans 10 000 volontaires équipés d'objets connectés, pour suivre à distance leur santé et identifier des facteurs de risque des pathologies.

gies thérapeutiques adaptées prenant en compte l'ensemble des données individuelles (→).

Elle serait enfin plus *participative*, avec un rôle plus actif des patients dans la décision et la gestion de leur santé.

La médecine génomique personnalisée en cancérologie est un premier exemple d'une médecine « personnalisée » ou « de précision » : grâce au séquençage à haut débit, l'analyse des milliards de données du génome est devenue rapide et peu coûteuse, et l'on traite aujourd'hui de plus en plus de cancers grâce à des thérapies ciblées en fonction de la signature génétique de la tumeur, avec des progrès de survie significatifs dans certains cas.

Le domaine du cancer, qui voit arriver les innovations à un rythme soutenu, avec des évolutions de traitement très rapides, des possibilités de combinaisons de thérapies multiples, des volumes de données à traiter potentiellement gigantesques, apparaît comme un domaine d'application privilégié pour le *Big data* et l'intelligence artificielle. Une des applications du programme d'intelligence artificielle « Watson » développé par *IBM* consiste ainsi à analyser en quelques instants les données du patient et des centaines de milliers de travaux scientifiques pour faire des propositions de stratégie thérapeutique.

Cette démarche de médecine personnalisée, où des modèles prédictifs tenant compte de la variabilité individuelle peuvent être construits en confrontant les données du patient à une masse de données issues d'autres patients et à des bases de connaissances, de façon à définir une réponse médicale optimisée, apparaît appelée à se développer dans de nombreux domaines.

Ces évolutions sont aussi porteuses d'enjeux industriels, avec un potentiel majeur de croissance, d'investissement et d'emploi.

Des bénéfices potentiels mais aussi des risques

En même temps que des espoirs, cette déferlante numérique suscite aussi des craintes sur les risques qu'elle comporte et les effets qu'elle peut induire, parmi lesquels :

- une « *ubérisation* » de la profession médicale, confrontée à l'émergence d'intermédiaires proposant des services sur internet, prise de rendez-vous, premier diagnostic, consultation, deuxième avis (« *un docteur chez vous en trois clics* », propose l'application *Docadom*), et à la numérisation d'une partie de leur valeur ajoutée dans des objets connectés et des logiciels [5]. Certains vont jusqu'à prédire la disparition de leur métier : Laurent Alexandre¹⁰ pronostique ainsi « La mort des médecins », subordonnés à des algorithmes d'intelligence artificielle prenant une place croissante dans la décision médicale, prédisant que « *les médecins signeront des ordonnances qu'ils n'auront pas conçues* » et que « *cela aboutira à un transfert du pouvoir médical des médecins vers les douze géants du numérique californiens et chinois qui contrôlent la totalité du marché mondial de l'intelligence artificielle* » [6] ;

(→) Voir le hors série n° 2 *Médecine personnalisée et innovations biomédicales*, m/s novembre 2014

- la déshumanisation consécutive des soins, qui résulteraient de plus en plus de décisions prises par des algorithmes opaques, ou de services en ligne n'assurant que des contacts ponctuels ;

- la domination sur la recherche et sur ces algorithmes des géants du numérique, forts de leurs bases de données alimentées par des milliards de clients, et la perte d'autonomie des systèmes nationaux face à la puissance de ces nouveaux acteurs mondiaux ;

- le détournement de l'usage des données dans des logiques purement commerciales : d'ores et déjà une industrie florissante de courtage en données s'est développée, proposant à la vente des listes de personnes souffrant de certaines pathologies, avec leurs identités numériques, de la même manière que se vendent des informations sur les comportements d'achat [7] ;

- la remise en cause du système de solidarité existant dans les sociétés développées en matière d'assurance santé, si les tests génétiques lèvent le « voile d'ignorance » qui conduit aujourd'hui, les risques étant perçus comme aléatoires, à accepter de les assurer solidairement. Si beaucoup de maladies génétiques sont multifactorielles et résultent d'une interaction complexe entre la prédisposition génétique des individus, leur environnement et leurs comportements, pour les maladies monogéniques la valeur prédictive est souvent beaucoup plus élevée. L'individualisation du risque pourrait, au fur et à mesure des progrès de la génétique, menacer la solidarité entre les « bons risques » et les « mauvais risques ».

Dans un contexte d'assurance privée, la question de l'usage de ces informations se pose également [8] (→) ;

- les risques pour les libertés individuelles, qui ne sont d'ailleurs pas propres à la santé. L'exploitation des traces numériques laissées en permanence par les individus dans leur vie quotidienne rend possible, comme le note Bernard Harcourt, « *des formes de contrôle et de surveillance que nous n'aurions jamais imaginées – excepté, peut-être, dans les dystopies orwelliennes* » [7, 9]¹¹. Dans le domaine de la santé, cette surveillance est souvent réalisée « pour le bien du patient » : elle n'en reste pas moins intrusive, et celui-ci peut devenir

(→) Voir l'article de D. Bardey et P. De Donder, m/s hors série n° 2, novembre 2014, page 41

¹⁰ Chirurien, entrepreneur, fondateur du site *Doctissimo* et auteur de nombreux ouvrages sur l'avenir de la médecine à l'ère du numérique.

¹¹ Le risque le plus important pouvant être, plus encore que la capacité de surveillance, le pouvoir d'influence. Ainsi lors d'une interview en 2010, Eric Schmidt, alors PDG de *Google*, dit : « Je pense que la plupart des gens ne veulent pas que *Google* réponde à leurs questions. Ils veulent que *Google* leur dise ce qu'ils devraient faire. [...] Nous savons grosso modo qui vous êtes, ce qui vous tient à cœur, qui sont vos amis. [...] Le pouvoir du ciblage individuel grâce à la technologie sera tellement parfait qu'il sera très dur pour les personnes de voir ou de consommer quelque chose qui n'a pas été, d'une certaine manière, taillé sur mesure pour eux. ». Interview citée par Eric Sadin [10].

l'objet d'une inquisition permanente visant à vérifier son « bon comportement », voire à prendre des mesures plus ou moins coercitives pour le remettre dans le droit chemin, en poursuivant les objectifs les plus louables.

Les réflexions avancent sur les conditions permettant d'engranger les bénéfiques et les avancées permises par cette explosion des données en évitant les effets pervers, fantasmés ou réels. Elles portent sur la protection des individus (le règlement européen sur la protection des données vient en particulier renforcer les droits des personnes concernées par des traitements de données), mais aussi sur les enjeux éthiques de l'intelligence artificielle et des algorithmes, question sur laquelle on voit dans la période récente se multiplier les rapports [10, 11] et les contributions¹².

La situation de la France en termes de données de santé

Compte tenu de ces enjeux, un objectif stratégique pour les pays, notamment européens, est déjà d'être en capacité de capter, d'organiser, d'utiliser les données que leur système de santé génère, pour développer des algorithmes adaptés à leur population, mais aussi pour faire progresser les démarches qualité, la surveillance sanitaire, les politiques de santé et de soins, etc.

La France a, de ce point de vue, un socle de données administratives qui constitue une réelle richesse, le système national de données de santé (SNDS). Elle s'est aussi dotée d'un plan de développement de bases de données génomiques. En revanche, elle est moins avancée que d'autres pays sur l'utilisation secondaire des données de santé produites dans le cadre des soins.

Le système national de données de santé

La France a mis en place, dans les années quatre-vingt-dix, un projet ambitieux de constitution d'un entrepôt de données « médico-administratives », c'est-à-dire collectées à des fins de gestion. Ce système national d'information inter-régimes d'assurance maladie (SNIIRAM) contient le détail de toutes les feuilles de soins ; enrichi des données de facturation hospitalière¹³, il permet de retracer, de manière détaillée et exhaustive, les parcours de soins de la totalité de la population française, soit 67 millions de personnes. Les données sont pseudonymisées¹⁴.

¹² Contributions exprimées notamment à l'occasion des débats lancés par la mission confiée au député Cédric Villani. Dans une tribune au « Monde » le 14 décembre 2017, des dirigeants d'instituts scientifiques estiment que le développement du numérique rend nécessaire la création d'un comité d'éthique pour ces technologies. http://www.lemonde.fr/idees/article/2017/12/14/il-faut-creer-un-comite-national-d-ethique-du-numerique_5229661_3232.html#rAwBYT7eEL0blQAW.99. Une autre initiative, « Ethik-IA », lancée à l'automne 2017 par une équipe d'enseignants, de chercheurs en droit numérique, en technologie de l'information et de la communication (TIC) et en sciences humaines et sociales propose une « régulation positive » de l'intelligence artificielle en santé.

¹³ Ces données de facturation des établissements hospitaliers sont centralisées dans une base de données appelée PMSI (programme de médicalisation des systèmes d'information), retraçant les diagnostics et actes principaux au cours d'une hospitalisation, tandis que le SNIIRAM retrace tous les soins reçus en ambulatoire. Les deux bases sont chaînées, ce qui permet de suivre l'intégralité du parcours de soins d'un patient.

¹⁴ C'est-à-dire que les informations d'identité de la personne (nom-prénoms, numéro de sécurité sociale) ont été retirées et remplacées par un « numéro d'anonymat », qui est toujours le même pour un même individu, ce qui permet un suivi longitudinal.

Ces données, avec aujourd'hui un historique de plus de dix ans, constituent un patrimoine remarquable : elles couvrent une large population, offrent des possibilités de suivi sur une longue période, sans perdu de vue en cours de suivi, avec une assez bonne homogénéité de codage. En outre, si des investissements importants ont été consentis pour constituer l'entrepôt, elles ont aujourd'hui un faible coût de collecte, puisqu'étant obtenues en sous-produit d'un système de gestion.

Cette base de données a déjà donné lieu à de nombreuses publications internationales sur l'état de santé et le recours aux soins, l'usage des médicaments en vie réelle, la mesure d'effets indésirables de traitements, le suivi des dépenses, etc. [12]. Elle est largement utilisée à des fins de pharmaco-surveillance par l'Agence nationale de sécurité du médicament et par l'Assurance maladie. Même si elle comporte peu de données cliniques, il est possible d'approcher les pathologies de patients en croisant un ensemble d'informations, motifs d'hospitalisations, produits de santé prescrits ou actes diagnostiques et thérapeutiques. Un réseau s'est d'ailleurs constitué pour élaborer des algorithmes par domaine pathologique [13].

Jusqu'à la période récente, les dispositifs mis en place pour accéder à ces données étaient confus et complexes. Chaque base avait sa gouvernance propre. Les équipes de recherche publiques pouvait demander des extractions du SNIIRAM, mais les délais pour les obtenir en pratique étaient souvent très longs, notamment du fait des délais d'instruction de la CNIL (Commission nationale de l'informatique et des libertés). Les appariements avec d'autres bases via le numéro d'inscription au Répertoire national d'identification des personnes physiques (NIR), plus communément appelé numéro de sécurité sociale, étaient virtuellement impossibles compte tenu des conditions très restrictives d'utilisation de cet identifiant¹⁵. Les accès n'étaient pas prévus pour les acteurs privés (industriels, bureaux d'études, etc.).

À l'issue de quelques années de débats publics et de réflexions¹⁶, la loi de modernisation de notre système de santé de janvier 2016 a réorganisé la gouvernance, clarifié la doctrine et remis à plat le dispositif d'accès [14]. Elle a créé le système national de données de santé, qui, outre les données précédemment citées, comprend également les causes médicales de décès ;

¹⁵ Un décret en Conseil d'État était nécessaire pour chaque étude.

¹⁶ En 2012, un collectif « Libérez les données de santé » s'est constitué, se réclamant de la société civile et revendiquant un droit d'accès au nom de la transparence due aux usagers sur le fonctionnement du système. À la suite de ce mouvement, une mission a fait des propositions à la demande de la Ministre. Une phase de concertation a suivi, avec la préparation de la loi de janvier 2016.

deux autres jeux de données viendront l'enrichir dans le futur¹⁷. En termes d'accès à ces données, la loi pose un principe d'ouverture large, les données pouvant être utilisées, quel que soit le statut public ou privé de l'utilisateur, pour toute recherche, étude ou évaluation, à condition que celle-ci présente un caractère d'intérêt public, qu'elle respecte l'interdiction concernant certaines finalités d'utilisation de ces données¹⁸ et que le traitement soit effectué dans des conditions qui garantissent le respect de la confidentialité des données et de la vie privée des personnes¹⁹.

Concrètement, les jeux de données extraits de ces bases médico-administratives et rendus totalement anonymes²⁰ ont vocation à être en *open data*, avec possibilité de réutilisation, sans nécessité d'autorisation préalable ; pour les données présentant un risque de ré-identification, la loi a prévu des accès permanents pour des organismes publics ou chargés d'une mission de service public, et pour les autres utilisateurs, des accès autorisés par la CNIL pour des projets spécifiques. Le dispositif mis en place a pour objectif d'accélérer les procédures, l'Institut national des données de santé (INDS) ayant pour mission d'accompagner les utilisateurs et de fluidifier les accès. Au terme de quelques mois de fonctionnement du nouveau dispositif, les résultats apparaissent encourageants : l'avenir dira si l'objectif, démultiplier les travaux menés à partir de ces bases, est effectivement atteint.

Le plan Médecine France génomique 2025

Le plan « Médecine France génomique 2025 », piloté par AVIESAN, a l'ambition d'ici 10 ans de positionner la France dans le peloton de tête des grands pays engagés dans la médecine génomique. Il prévoit de prendre en charge, à l'horizon 2020, environ 235 000 séquences de génomes par an pour les patients atteints de maladies rares et leurs familles et pour des patients atteints de cancers, puis d'intégrer progressivement des maladies communes [15].

Un appel à projet pour des plateformes de séquençage génomiques à haut débit a été lancé, et les deux premières ont été sélectionnées en juillet 2017.

Les données de santé issues du système de soins

Le SNDS est une source de données riche, mais qui renseigne essentiellement sur les *parcours de soins*. Il ne s'agit pas de dossiers médicaux : on n'y trouve pas de notions d'antécédents personnels ou familiaux, de facteurs de risques (tabac, alcool, etc.), de données cliniques (indice de masse corporelle, niveau tensionnel, etc.) ou paracliniques (résultats d'examens).

Un certain nombre de pays ont mis en place une stratégie pour constituer des bases de données en récoltant des informations qui sont générées au fil du parcours de soins des patients ; l'objectif est de promouvoir des utilisations secondaires de ces données à des fins de santé publique, de recherche ou d'amélioration du système de soins.

C'est une tradition ancienne dans les pays nordiques, qui ont des registres exhaustifs couvrant de nombreux domaines. Beaucoup de travaux ont ainsi été réalisés sur les données danoises, suédoises ou norvégiennes. Mais d'autres pays embrassent aujourd'hui cette ambition.

L'Angleterre l'a fait sur le cancer, en constituant un registre exhaustif de la population traitée pour cancer : ce registre ne se limite pas à un objectif de mesure de l'incidence et de la prévalence, mais recueille un ensemble de données sur les pratiques de soins, à des fins d'évaluation de la qualité.

La France dispose d'un certain nombre de grandes cohortes, notamment celles financées par les programmes investissements d'avenir ; appariées avec les données du SNDS, elles peuvent constituer des outils précieux [16] (→).

(→) Voir le Forum de M. Goldberg et al., m/s n° 4, avril 2012, page 430

Mais par rapport à un objectif de collecte à grande échelle de données cliniques à partir des dossiers médicaux informatisés, la France paraît moins bien positionnée que d'autres pays, du fait de l'informatisation hétérogène des établissements et des professionnels de ville, de la faible structuration des données et de l'absence de projet organisé dans ce domaine.

Une étude comparative menée récemment par l'Organisation de coopération et de développement économiques (OCDE) conclut d'ailleurs que si tous les pays investissent dans le développement de dossiers médicaux électroniques, seuls une dizaine avancent dans la possibilité d'extraire des données à des fins de recherche, de statistiques ainsi que d'autres utilisations servant l'intérêt public [17]. La France ne fait pas partie de ces pays pionniers, qui ont mis en place, à la fois du point de vue technique et opérationnel et du point de vue de la gouvernance, les conditions nécessaires au développement d'un usage secondaire des informations provenant des dossiers patients. Il est souhaitable, au regard des avancées que l'on constate dans d'autres systèmes, que nous nous dotions d'une stratégie dans ce domaine. ♦

SUMMARY

Health data (1)

Healthcare is considered as one of the most promising areas where big data can be applied to foster innovation for the benefit of patients and of the whole system.

¹⁷ Il s'agit de données concernant les personnes en situation de handicap, qui seront issues des maisons départementales des personnes handicapées et centralisées par la Caisse nationale de solidarité pour l'autonomie ; de données de remboursements d'organismes d'assurance complémentaire, pour un échantillon de population, qui seront appariées avec les remboursements des régimes obligatoires.

¹⁸ Les finalités interdites sont la promotion des produits de santé pour les industriels, et la sélection des risques ou la tarification au risque pour les assureurs.

¹⁹ Un référentiel de sécurité a été prévu par la loi de modernisation de notre système de santé et a été publié en mars 2017. Il prévoit des conditions strictes pour les environnements d'exploitation des données (authentification, traçabilité des accès, etc.).

²⁰ Au sens où l'on ne peut plus ré-identifier des individus.



Healthcare analytics have the potential to accelerate R&D, increase knowledge on diseases and risk factors, improve treatments, develop personalised medicine, help physicians with decision support systems... The access to data is also a driving force for patients' empowerment and for the democratic debate. However, there are also concerns about the societal, economic and ethic impacts of this wave of digitization and of the growing use of data, algorithms and artificial intelligence. Given the issues at stake, collecting and analysing data generated by health care systems is a strategic challenge in all countries; in that respect the French National System of Health Data (a national data warehouse linking data from several sources and giving a vision of the care pathways for the entire population, with a ten-year history) is an asset, but it has to be completed and enriched with data from electronic health records. ♦

LIENS D'INTÉRÊT

L'auteur déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

RÉFÉRENCES

- Règlement général sur la protection des données. <http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=FR>
- Inglis SC, Clark RA, Dierckx R, et al. Structured telephone support or non-invasive telemonitoring for patients with heart failure (review). *Cochrane Database Syst Rev* 2015. <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007228.pub3/epdf>
- Denis F, Yossi S, Septans AL, et al. Improving survival in patients treated for a lung cancer using self-evaluated symptoms reported through a web application. *Am J Clin Oncol* 2017 ; 40 : 464-9.
- Weill A, Païta M, Tuppin P, et al. Benfluorex and valvular heart disease: a cohort study of a million people with diabetes mellitus. *Pharmacoepidemiol Drug Saf* 2010 ; 19 : 1256-62.
- Destky AS, Garber AM. Uber's message for health care. *N Engl J Med* 2016 ; 374 : 806-9.
- Alexandre L. La mort des médecins. *Les Tribunes de la Santé* 2017 ; 54 : 43-7.
- Harcourt BE. Gouverner, échanger, sécuriser. In : Menger PM, Paye S, eds. *Big data et traçabilité numérique. Les sciences sociales face à la quantification massive des individus*. Paris : Collège de France, 2017 : 47-68.
- Bardey D, De Donder P. Tests génétiques, prévention et marchés d'assurance santé. *Med Sci (Paris)* 2014 ; 30 (hors série n° 2) : 41-3.
- Sadin E. *La vie algorithmique. Critique de la raison numérique*. Paris : L'Échappée, 2015 : 278 p.
- CNIL. Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle. Paris, décembre 2017, 76 p. https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf
- Conseil National de l'Ordre des Médecins. *Médecins et patients dans le monde des data, des algorithmes et de l'intelligence artificielle*. Janvier 2018, 66 p. https://www.conseil-national.medecin.fr/sites/default/files/cnomdata_algorithmes_ia.pdf
- Tuppin P, Rudant J, Constantinou P, et al. Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Rev Epidemiol Sante Publ* 2017 ; 65 (suppl 4) : 149-67.
- Goldberg M, Carton M, Doussin A, et al. Le réseau REDSIAM (Réseau données Sniiram). *Rev Epidemiol Sante Publ* 2017 ; 65 (suppl 4) : 144-8.
- Loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé. https://www.legifrance.gouv.fr/affichTexte.do?sessionId=BABD1692B567056C3102D67415B8C52C.tplgfr23s_3?cidTexte=JORFTEXT000031912641&categorieLien=id
- Aviesan. Médecine France génomique 2025. Rapport au Premier Ministre, 2015. http://www.gouvernement.fr/sites/default/files/document/document/2016/06/22.06.2016_remise_du_rapport_dyves_levy_-_france_medicine_genomique_2025.pdf
- Goldberg M, Coeuret-Pelissier M, Ribet C, et al. Cohortes épidémiologiques et bases de données d'origine administrative : un rapprochement potentiellement fructueux. *Med Sci (Paris)* 2012 ; 28 : 430-4.
- Oderkirk J. Readiness of electronic health record systems to contribute to national health information and research. *OECD Health Working Papers* 2017 ; 99 : 80 p.
- Meyer M. Biologie et médecine « do-it-yourself » : histoire, pratiques, enjeux. *Med Sci (Paris)* 2018 ; 34 : 473-9.

TIRÉS À PART

D. Polton

LA FONDATION PREMUP : UN OPÉRATEUR DE TERRAIN EN PÉRINATALITÉ RECONNU POUR SON EXCELLENCE ET SON INTERDISCIPLINARITÉ

La Fondation de coopération scientifique PremUp, unique en Europe, intervient sur la prévention du handicap à la naissance, par la protection de la santé de la femme enceinte et du nouveau-né.



> Grâce à *m/s*, vivez en direct les progrès des sciences biologiques et médicales

Tarifs d'abonnement *m/s* - 2017

Abonnez-vous

à *médicine/sciences*

Bulletin d'abonnement
page 490 dans ce numéro de *m/s*

