

Big data et action publique

L'usage des données massives dans les domaines tels que l'hygiène, l'épidémiologie, la vigilance sanitaire permet d'entrevoir de nouvelles pistes de recherche et de nouveaux moyens d'action pour éclairer les décisions de la puissance publique.

Du Système national des données de santé au Health Data Hub : mise en œuvre et évolution

Ouvrir l'accès aux données de santé collectées par des organismes publics afin de tirer profit des potentialités qu'elles offrent est un enjeu sanitaire majeur. Cette ouverture vise à accroître les connaissances relatives à l'offre de soins et à la prise en charge médico-sociale à destination aussi bien des professionnels de santé que des usagers et des citoyens, à contribuer à la recherche et à l'innovation en santé et enfin à favoriser la veille et la sécurité sanitaires, des objectifs dont l'épidémie de SARS-CoV-2 a révélé le caractère essentiel.

La création du Système national des données de santé (SNDS), par la loi de modernisation de notre système de santé¹, constitue la réponse du législateur, sous l'impulsion de la ministre des Affaires sociales et de la Santé. La France s'est ainsi dotée d'un dispositif de pointe par rapport à de nombreux États de l'Union

européenne (UE), s'appuyant sur un cadre institutionnel, des acteurs et des procédures bien définis.

Le Système national des données de santé a été profondément réformé en 2019 pour tenir compte des besoins des acteurs concernés en termes de données, mais aussi d'un contexte politique favorable au développement du numérique en santé.

L'accès aux données médico-administratives du SNDS : une réglementation qui repose sur un équilibre entre ouverture des données et respect de la vie privée

Le Système national des données de santé historique, tel qu'il a été mis en place en 2016, a deux missions principales : il rassemble et met à disposition trois bases exhaustives au niveau national de données de santé publique préexistantes et indépendantes les unes des autres. Il est alimenté par :

- la base de données Sniiram (Système national d'information inter-régimes de l'Assurance maladie) contenant les données relatives à toutes les dépenses de l'Assurance maladie ;

Ève Jullien

Conseillère juridique,
Sous-direction
de l'observation
de la santé
et de l'assurance
maladie, Drees

1. Loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé. Elle institue le SNDS à l'article 193, dont les dispositions ont été ensuite intégrées aux articles L. 1461-1 à L. 1461-2 du Code de la santé publique.



- la base de données PMSI (Programme de médicalisation des systèmes d'information) contenant les données d'analyse de l'activité des établissements de santé ;
- la base de données du CépiDc (Centre d'épidémiologie sur les causes médicales de décès) contenant les données relatives aux causes de décès recueillies par les collectivités territoriales.

Des données médico-sociales liées au handicap, fournies par les maisons départementales des personnes handicapées viendront bientôt compléter cette base exhaustive².

Le SNDS, à l'image d'un entrepôt, regroupe donc des données de santé au sens strict, des données sur l'état de santé d'un patient (une affectation de longue durée par exemple ou un diagnostic d'hospitalisation) mais également des données portant sur le patient ou son professionnel de santé et enfin des informations relatives aux soins qu'il a consommés.

La Caisse nationale de l'assurance maladie (Cnam) est chargée du rassemblement et de la mise à disposition des données du Système national des données de santé historique. L'Institut national des données de santé (INDS) représentait le guichet unique pour le suivi des demandes de mise à disposition en vue de réaliser une étude, une recherche ou une évaluation dans le domaine de la santé. L'INDS opérait ainsi le suivi des formalités d'accès aux données du Système national des données de santé et ce conformément au cadre réglementaire.

La plupart des acteurs, publics et privés, devaient en effet déposer un dossier de demande d'accès présentant leur projet de recherche et accomplir des formalités obligatoires auprès d'un comité scientifique d'abord, puis de la Commission nationale de l'informatique et des libertés (Cnil). Un comité de protection des personnes (CPP), si le projet porte sur une recherche impliquant la personne humaine, ou le Comité d'expertise pour les recherches, les études et les évaluations dans le domaine de la santé (Cerees), pour celles n'impliquant pas la personne humaine, émet un avis sur les aspects méthodologiques et scientifiques du dossier avant que la Cnil ne se prononce sur le respect de la protection des données. Le Comité d'expertise sur l'intérêt public (CEIP) de l'INDS produit un avis sur ce point. Cette procédure permet de faciliter l'examen par la Cnil des projets de recherche et de réduire les délais d'instruction des demandes.

En revanche, certains organismes publics ou services chargés d'une mission de service public, dont la liste est prévue par voie réglementaire, bénéficient d'un accès dit permanent au Système national des données de santé et sont dispensés de formalités.

2. Un cinquième flux contenant des données de remboursement des complémentaires santé était prévu mais il n'a pas encore été mis en œuvre.

Le Système national des données de santé s'est construit au croisement de deux logiques, une logique de valorisation de cette masse considérable de données de santé ainsi collectée par des procédures homogénéisées et transparentes d'une part, et une logique de protection des patients par l'instauration de garde-fous juridiques et techniques d'autre part.

La réglementation encadre les motifs pour lesquels les données du Système national des données de santé peuvent être utilisées et interdit qu'elles servent à la promotion commerciale de produits de santé et à une éventuelle modification de cotisations ou primes d'assurance d'une personne présentant un risque. La loi formule également l'obligation de « pseudonymiser » les données du Système national des données de santé afin de rendre impossible l'identification directe un patient.

Des règles en matière de sécurité et confidentialité qui s'appliquent aux organismes ayant accès aux données du Système national des données de santé sont également prévues. Un référentiel de sécurité, propre au Système national des données de santé, a en outre été élaboré par le ministère chargé de la santé. Il précise les règles de sécurité que tout système utilisant des données du Système national des données de santé doit mettre en place en termes de procédure d'accès, d'habilitation des utilisateurs, de formations obligatoires et de traçage des traitements effectués, les données étant mises à disposition dans des bulles sécurisées.

Enfin, l'ensemble des principes applicables en matière de protection des données sont en vigueur, non seulement les droits que les personnes dont les données sont utilisées peuvent exercer mais également les obligations encadrant la collecte des données³.

La loi d'adaptation du RGPD a maintenu les grands principes prévalant en France sur l'accès aux données de santé

Le règlement général sur la protection des données (RGPD), dont les dispositions sont en vigueur depuis le 25 mai 2018, approfondit et étend les droits et obligations à la charge des entités traitant des données à caractère personnel. Il accroît notamment la responsabilisation des acteurs en supprimant la plupart des formalités préalables et en faisant porter sur les

3. Les personnes ont le droit d'obtenir les informations relatives au traitement opéré (catégories de données collectées, identité du responsable de la collecte, objectifs du traitement, durée de conservation, destinataires, éventuels transferts ou réutilisations). Elles disposent également d'un droit d'accès (droit de connaître les données qui ont été traitées), de rectification (droit de corriger les données collectées) et d'opposition sous certaines conditions (droit de refuser que les données soient collectées).

S'agissant des obligations prévues par le RGPD, les données doivent être collectées et traitées de manière licite et loyale, pour des raisons déterminées et légitimes ; les données doivent être pertinentes, exactes et la durée de conservation doit être proportionnée à l'usage qui en est fait.

entités la charge de s'assurer, par elles-mêmes, que leurs traitements sont conformes aux dispositions en vigueur. Toutefois, s'agissant des données sensibles, dont les données de santé, le RGPD permet à chaque État de prévoir des procédures particulières.

La France, par la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles, a fait le choix de maintenir l'obligation d'obtenir une autorisation préalable de la Cnil pour traiter des données de santé dans tous les cas où aucun autre cadre ne permet d'y déroger (exceptions législatives, conformité à un référentiel...). Le rôle des instances consultatives (CPP, Cerees et CEIP) a été conservé. Cette volonté de conserver un haut niveau de protection des données de santé a conduit à l'instauration, par cette même loi, du comité d'audit du Système national des données de santé. Ce comité, présidé par le haut fonctionnaire de défense et de sécurité du ministère chargé de la santé, diligente des audits visant à s'assurer du bon respect des règles d'accès et d'usage du Système national des données de santé.

La réforme de 2019 élargit le périmètre du SNDS et modifie la gouvernance des données de santé afin que la France devienne pionnière dans la structuration des données de santé

Les évolutions du SNDS, introduites par la loi relative à l'organisation et à la transformation du système de santé⁴, s'inscrivent dans la stratégie de développement du numérique et de l'intelligence artificielle, annoncée par le président de la République en mars 2018.

La loi du 24 juillet 2019 fonde l'extension du périmètre du Système national des données de santé à l'ensemble des données de santé dont le recueil est directement ou indirectement financé par des fonds publics, en particulier les données cliniques recueillies dans le cadre du parcours de soins, ce qui inclut les activités de prévention, de diagnostic, de soins et de suivi social et médico-social. De nouvelles sources de données sont ainsi ajoutées aux quatre bases historiques, l'appariement entre ces données devant faciliter leur structuration et démultiplier leurs usages. Ces nouvelles données permettent essentiellement d'apporter de l'information sur l'état de santé précis du patient (résultats d'examen, diagnostics des médecins) et sur ses déterminants de santé (situation sociale, indice de masse corporelle, consommation de tabac...), dont l'absence limitait les usages du Système national des données de santé.

Ces modifications s'accompagnent d'une refonte de la gouvernance des données de santé. L'Institut national des données de santé (INDS) est remplacé par la Plateforme des données de santé (PDS), également désignée par son appellation anglaise Health Data Hub, dont les missions sont élargies. Si elle conserve le

rôle de guichet unique pour les demandes d'accès, la Plateforme peut en outre procéder pour le compte de tiers à des traitements de données et se voit confier un rôle de facilitateur, tant pour les usagers dans l'exercice de leurs droits que pour les porteurs de projets dans leur réalisation.

Une organisation bicéphale est instituée afin de mettre en œuvre le Système national des données de santé élargi, la Cnam et la PDS en deviennent les deux responsables. Le Cerees devient le Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (Cesrees), qui rend désormais des avis plus complets. Les demandes pour les recherches n'impliquant pas la personne humaine sont examinées par un seul comité qui se prononce sur les différents aspects (méthodologiques, éthiques et intérêt public) qui sont généralement imbriqués. Le système des accès permanent a, en revanche, été conservé.

Il est encore tôt pour évaluer l'impact des changements prévus par la loi du 24 juillet 2019 dès lors que tous les textes d'application ne sont pas encore entrés en vigueur et que la Plateforme des données de santé n'est pas complètement opérationnelle. Il importe toutefois de souligner que l'ambition de cette réforme est de faire de la France l'un des pays pionniers dans le domaine de la promotion et de la valorisation des données de santé. Au sein de l'Union européenne, seule la Finlande a mis en œuvre un système relativement comparable. Sans conteste, la refonte du Système national des données de santé confère à la France un atout alors que la Commission européenne a lancé en septembre 2019 une initiative destinée à créer un « espace européen de la donnée de santé ».

Exemple d'utilisation de données du Système national des données de santé

De nombreuses études mobilisent les données du SNDS. Cette base de données a déjà permis d'améliorer les connaissances sur le système de santé et ses acteurs mais également d'améliorer significativement la pharmacovigilance (depuis les affaires du Médiateur ou de la Dépakine par exemple) ou encore de mieux suivre les parcours de soins et leurs conséquences. On peut par exemple citer l'étude de M. Coldéfy et C. Gandé [15] qui, à partir de la mobilisation des trois sources de données actuellement présentes dans le Système national des données de santé (Sniiram, PMSI et CépiDc), montre que la réduction de l'espérance de vie des individus suivis pour des troubles psychiques sévères atteint en moyenne 16 ans chez les hommes et 13 ans chez les femmes. Si ce phénomène était connu, seule la richesse des données du SNDS permet d'en objectiver l'ampleur, de le décliner par cause de mortalité ou troubles psychiatriques ou encore de suivre son évolution temporelle et géographique. ●

Les références entre crochets renvoient à la Bibliographie générale p. 57.

4. Loi n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé (art. 41).