

# Le Health Data Hub, levier pour la valorisation des données de santé

Par **Stéphanie COMBES**

Directrice du Health Data Hub (HDH)

En cherchant à mobiliser tout le potentiel des données de santé afin d'améliorer le système de santé, le Health Data Hub se situe à la ligne de crête de l'innovation.

Son objectif est de permettre et de simplifier l'accès aux bases de données de santé (principalement au Système national des données de santé (SNDS), la base de données de l'Assurance maladie) afin de les réutiliser à des fins de recherche.

Face aux nombreux enjeux soulevés, la plateforme a pris le parti d'inclure son écosystème et les citoyens dans son projet, ainsi que de s'ouvrir à une perspective internationale : pouvoir aligner le Health Data Hub avec l'ensemble des acteurs impliqués et les perspectives à venir.

## La France est dotée de l'une des bases de données médico-administratives les plus volumineuses du monde

Cette base de gestion, liée aux remboursements des actes et des soins des bénéficiaires de l'ensemble des régimes d'assurance maladie obligatoire, a longtemps été sous-exploitée. À la suite de plusieurs rapports, l'appel à une meilleure accessibilité aux données de santé a été entendu en 2016. La loi Modernisation de notre système de santé a créé l'Institut national des données de santé (INDS), qui est chargé de favoriser le dialogue entre les acteurs et de simplifier les procédures d'accès à ces dernières.

Le 29 mars 2018, dans la continuité des recommandations du rapport du député Cédric Villani sur l'intelligence artificielle, le Président de la République annonçait la création d'un *hub* des données de santé, une structure partenariale dont l'objet est de garantir un accès simplifié aux données de santé au travers d'une plateforme technologique sécurisée disposant d'outils d'analyse répondant à l'état de l'art, le tout dans le respect des droits du citoyen. Le projet HDH devient, à ce titre, une des actions de la stratégie nationale en intelligence artificielle.

Pendant tout l'été 2018, à la demande de la ministre des Solidarités et de la Santé, Agnès Buzyn, une mission pilotée par trois experts du domaine des données de santé est conduite qui remettra son rapport le 12 octobre. La direction de la Recherche, des Études, de l'Évaluation et des Statistiques (DREES) se verra alors confier la mise en œuvre de la feuille de route qui y est proposée et, huit mois plus tard, la loi Organisation et transformation du système de santé

(OTSS) sera votée. Son article 41 élargit le Système national des données de santé (SNDS) et crée officiellement le HDH.

Le Health Data Hub (HDH) est donc constitué en groupement d'intérêt public, dont la convention constitutive a été approuvée par arrêté ministériel du 29 novembre 2019. L'Assemblée générale du Health Data Hub réunit ses cinquante-six membres, répartis en neuf collèges : l'État, les caisses d'assurance maladie, les organismes d'assurance maladie complémentaires, les organismes de recherche, les établissements de santé, les professionnels de santé, les agences, les opérateurs et l'autorité publique indépendante, les représentants des usagers du système de santé et les industriels.

## Le HDH vise donc à garantir un accès aisé et unifié, transparent et sécurisé aux données de santé pour améliorer la qualité des soins et l'accompagnement des patients

Chargé par la loi de réunir, organiser et mettre à disposition les données du SNDS élargi et de promouvoir l'innovation dans l'utilisation des données de santé, le HDH conduit quatre activités principales :

- il est un guichet unique pour accompagner les porteurs de projets dans leurs démarches administratives. Il assure, par ailleurs, le secrétariat du CESREES qui donne un avis sur les projets préalablement à leur transmission à la CNIL ;
- il met à disposition une plateforme sécurisée à l'état de l'art et offrant des capacités avancées de stockage, de calcul, de rapprochement et d'analyse des données ;

- il offre un accès à des bases de données partenaires, constituant le « catalogue de données ». Construit de manière progressive et itérative, ce catalogue est porté par un arrêté pris après avis de la CNIL. La réplique des bases mises au catalogue est possible pour en permettre l'accès et l'interfaçage : leur croisement avec d'autres bases de données, en accord avec les acteurs à l'origine de ces données. Ce catalogue est mis à jour régulièrement ;
- il anime l'écosystème dans le but d'accélérer l'innovation en favorisant le partage des expériences et des connaissances.

L'apport du HDH, service créé à destination des acteurs en charge de la collecte et de la mise à disposition des données et de ceux qui les utilisent, s'est concrétisé deux ans et demi après sa création par la mise en œuvre de :

- deux appels à projets avec plus de 350 répondants ;
- une plateforme technologique homologuée, sécurisée à l'état de l'art et actuellement utilisée ;
- soixante-six projets impliquant pour plus de la moitié des établissements de santé et ayant déjà abouti, pour certains, à de premières publications ;
- une dizaine de bases dans la première version du catalogue ;
- près d'une centaine de partenaires dont certains avec des associations d'usagers du système de santé ;
- et la création d'une direction citoyenne.

Le HDH bénéficie également d'une reconnaissance auprès des acteurs de l'écosystème et des acteurs européens et internationaux, comme la Commission européenne, Harvard ou le MIT.

### Focus sur un projet innovant : INNERVE

Le projet INNERVE, porté par la société Quantmetry, a été sélectionné dans le cadre du deuxième appel à projets du HDH, en partenariat avec le Grand défi « Comment améliorer les diagnostics avec l'intelligence artificielle ? ». Il vise à développer un logiciel devant permettre d'affiner le diagnostic des neuropathies des petites fibres. Les petites fibres sont des fibres nerveuses qui permettent de ressentir la douleur et la température, Leur dégradation entraîne des douleurs et une perte de sensibilité. Identifier celle-ci demande un temps considérable au médecin qui doit explorer les images capturées lors de l'examen médical. Le recours à des algorithmes doit permettre, avec une efficacité similaire, de réduire considérablement le temps de détection et de libérer du temps médical. Aujourd'hui, le projet fournit des résultats de 70 % d'efficacité.

En parallèle de la création du HDH, le SNDS a été élargi à toutes les données de santé qui bénéficient d'un financement de la solidarité nationale dans le but de contribuer à une utilisation plus large de ces données. Initialement composé de certaines données

médico-administratives telles que les feuilles de soin, la facturation hospitalière et les causes médicales des décès, le SNDS comprend donc désormais des données de registres, de cohortes de recherche, ou d'entrepôts de données hospitalières, etc.

Le HDH a été chargé par un décret publié en juin dernier de la coresponsabilité du traitement de la base principale avec la CNAM et de la responsabilité du traitement du catalogue du SNDS. Ces deux sous-ensembles du SNDS présentent un intérêt majeur pour l'écosystème.

Ainsi, la base principale désigne la réunion de données couvrant l'ensemble de la population : des données en provenance de l'Assurance maladie (base SNIIRAM), des établissements (base PMSI) ; des causes médicales de décès (base du CépiDC de l'Inserm), des données relatives au handicap (en provenance des MDPH – Données de la CNSA) en cible. Elle a vocation à être enrichie en continu avec l'intégration d'autres bases de données nationales et notamment, dans le cadre de ce premier arrêté, des bases relatives à l'épidémie de Covid-19 : vaccin Covid et SI-DEP (système d'information de dépistage).

Le catalogue désigne, quant à lui, une collection de bases de données non parfaitement connue à l'avance. Construit itérativement, il permet de s'adapter aux enjeux et besoins de l'écosystème. Son contenu est fixé par le Comité stratégique des données de santé, qui, créé en juillet dernier, est présidé par le ministre chargé de la Santé, tout comme est fixé l'ajout de nouveaux flux à la base principale. La première composition de ce catalogue a été fixée par un arrêté publié tout récemment.

Il est à noter que seule une copie des données sera transférée au HDH : l'acteur dépositaire de la source initiale la conserve et peut mettre à disposition ces données par ses propres moyens, s'il le souhaite. L'intérêt de copier ces bases dans la plateforme du HDH est de pouvoir enrichir ces bases en les croisant avec la base principale du SNDS et de réduire les délais d'accès en déléguant au HDH les actions d'hébergement, de *data management*, l'enrichissement du système par d'autres sources, la mise à disposition aux personnes habilitées de ces données dans des espaces informatiques maîtrisés, de contractualisation, etc. Les modalités sont précisées par voie de convention entre le dépositaire des données et le HDH.

Le catalogue et la plateforme du HDH permettent de mutualiser les investissements humains, technologiques et financiers inhérents au partage des données de santé dans le respect de la réglementation et des droits des personnes.

Les dix premières bases du catalogue constituent un premier aperçu de l'horizon des possibles en réunissant des bases de nature variée : des bases administratives, des bases nationales, des cohortes ou registres, des entrepôts parfois thématiques pour adresser des priorités de santé publique et également une base d'origine privée qui présente également un grand potentiel.

## Des avancées significatives malgré la persistance de freins au partage des données

En dépit d'avancées, le HDH a rencontré différentes difficultés tout au long de sa mise en place, qui ont pu ralentir celle-ci, comme :

- les délais importants de publication des textes et donc des retards dans le transfert de la base principale du SNDS. En l'absence de celle-ci, le HDH met à disposition les données projet par projet, ce qui a pour effet d'augmenter considérablement les délais (environ un an en moyenne, pouvant parfois aller jusqu'à deux ans). Rendre effectif ce transfert en 2022 représente donc un enjeu majeur pour passer à l'échelle et accompagner un nombre croissant de projets ;
- la réticence à un partage large qui continue de guider ceux qui constituent les bases de données, en l'absence de modèles pérennes et lisibles de financement des bases de données de santé malgré leur intérêt aujourd'hui considéré unanimement comme incontestable ;
- l'absence d'un réseau mature d'entrepôts des données de santé dans les établissements de santé ou l'absence d'une base structurée sur les diagnostics médicaux en ville, ce qui privent les différents acteurs d'un accès à un patrimoine de données extrêmement important.

Les freins liés au partage et aux usages de la donnée de santé ont des conséquences pour notre pays. Les dépasser renforcerait notre recherche, augmenterait notre compétitivité, ainsi que notre attractivité et notre souveraineté, et ce dans un contexte de compétition internationale exacerbée. Les échanges entre l'écosystème hospitalier et le HDH se sont d'ailleurs accélérés ces derniers mois dans le but d'organiser une collaboration plus efficace afin de multiplier les projets d'envergure.

Une des pistes d'intérêt pour surmonter ces freins est celle de la dynamique européenne. À l'échelle européenne, la France est considérée comme étant l'un des pays les plus engagés dans le développement des usages des données de santé. La dimension internationale est au cœur de la philosophie du HDH, lequel est fortement impliqué dans la construction de plusieurs textes au niveau européen dans le domaine des données.

Le HDH est également engagé dans deux projets majeurs au niveau européen :

- en consortium avec des établissements de santé (les Hospices civils de Lyon, le Centre Léon Bérard Unicancer, l'hôpital Paris Saint-Joseph, le CHU de Nancy), l'ANS et la plateforme Bordeaux PharmacoEpi (BPE), il a été lauréat d'un appel à projets de l'Agence européenne du médicament visant à constituer une base clinique multicentrique ;
- le HDH est aussi le pilote d'un consortium composé de quinze partenaires européens stratégiques (huit plateformes nationales de données de santé, trois infrastruc-

tures de recherche, deux agences européennes (dont l'EMA) et des associations) afin de candidater à la préfiguration d'un espace européen des données de santé (EHDS). Cet EHDS constitue l'une des priorités de la politique de santé européenne ; il est considéré comme fondamental pour développer la recherche et améliorer la santé des citoyens. La réponse est attendue avant l'été.

La coopération européenne apparaît comme une opportunité à saisir pour exister au niveau mondial. En matière de santé, les pays de l'UE sont confrontés à des défis similaires à ceux de la France : transition épidémiologique, pandémies, augmentation des dépenses de santé...

Ces problèmes complexes font de la santé une question transfrontalière et ne seront résolus qu'en mettant à profit un partage des données à grande échelle considéré comme fondamental pour développer la recherche et améliorer nos systèmes de santé. Les collaborations européennes permettent d'atteindre un seuil critique, d'accroître l'avantage concurrentiel pour les entreprises et *start-ups* du secteur et de garantir une résilience et une visibilité accrues au niveau international. Elles peuvent générer des données de santé plus nombreuses et de meilleure qualité, réduire le risque de biais technologique, créer de nouvelles possibilités de croisements entre les données de santé et d'autres données, permettant de construire des algorithmes d'IA plus performants au service de la recherche ou de la performance industrielle. La dimension internationale de PariSanté Campus pourra renforcer la visibilité de la France au niveau international en réunissant ses forces vives derrière une identité unique et lisible (voir l'Encadré de la page suivante).

## Conclusion

La France est aujourd'hui considérée comme un acteur légitime et pertinent au niveau européen. Elle dispose de perspectives pour s'imposer pleinement comme un leader naturel sur le sujet de la donnée de santé. Des progrès considérables ont été faits ces dernières années pour consolider la culture de la donnée auprès des acteurs de la santé. Des efforts, à porter avec PariSanté Campus, dont le HDH est membre fondateur, et le soutien de ses opérateurs et de son écosystème, sont encore nécessaires pour consolider un réseau d'acteurs partenaires sur l'ensemble du territoire capable de se mobiliser pour mener des projets collectifs, pour consolider le patrimoine de données de santé au travers de la mise en place d'entrepôts de données dans les établissements hospitaliers ou d'un observatoire de la médecine de ville interopérables avec le Health Data Hub y compris en termes de gouvernance et de valorisation, ou encore pour concevoir et mettre en œuvre une véritable politique publique de la donnée de santé permettant de définir un cadre de partage lisible et opposable au niveau national et tenant compte de la réglementation européenne.

## **Le HDH, le moteur d'un consortium candidat à la préfiguration de l'Espace européen des données de santé (EHDS)**

Depuis 2019, la Commission européenne a identifié la constitution d'un espace européen des données de santé comme l'une des priorités de la politique de santé européenne. Il s'agit de faciliter l'accès aux différents types de données disponibles au sein des États membres dans le but de développer la recherche et d'améliorer la santé des citoyens.

La France est considérée comme étant l'un des pays les plus engagés dans le développement des usages des données de santé et dans de multiples projets. Aujourd'hui, le Health Data Hub, qui a inscrit parmi ses quatre axes stratégiques l'objectif de « positionner la France comme un leader dans l'usage des données de santé », est une structure-exemple dans la construction de plusieurs textes clés au niveau européen dans le domaine des données.

Il s'agit notamment :

- du projet de Data Governance Act (ou DGA), qui, adopté par la Commission européenne le 26 octobre 2020, vise à élaborer un cadre de gouvernance des données favorisant leur réutilisation et la création d'un véritable marché unique des données ;
- du projet de règlement sur l'espace européen des données de santé (EHDS) qui aura pour objet de construire un système de gouvernance solide, ainsi que des règles concernant l'échange des données de santé, mais aussi des infrastructures et une interopérabilité solide.

L'EHDS permettra également de répondre aux trois constats qui peuvent être faits à l'échelle européenne lorsqu'il s'agit de l'utilisation des données de santé : la fragmentation de ces données, la multiplicité de leurs conditions d'utilisation et la diversité des modèles de gouvernance permettant d'y accéder.

Le texte, dévoilé par la Commission européenne, le mardi 3 mai 2022, aura un impact majeur sur l'utilisation des données de santé en Europe, en ce qu'il précise les règles européennes en la matière. En ce qui concerne l'utilisation secondaire des données de santé, le projet de règlement met en place un cadre de gouvernance visant à simplifier et à harmoniser les conditions d'accès. Ce cadre de gouvernance comprend notamment des règles et une procédure d'accès aux données imposant des délais pour l'autorisation et la mise à disposition des données, la mise en place d'organismes nationaux chargés de l'accès aux données, la définition des rôles et des responsabilités des acteurs impliqués.

Par ailleurs, le projet de règlement propose également un cadre pour l'utilisation transfrontière des données de santé avec la mise en place d'une infrastructure européenne HealthData@EU, et guide les acteurs pour faciliter l'identification et la réutilisation des données au niveau européen. Enfin, le texte prévoit la mise en place d'une gouvernance européenne à travers la création d'un European Health Data Space Board.

En 2021, la Commission européenne a par ailleurs lancé un appel à projets dans le but de construire une première version test de cet EHDS. Dans cette optique, le HDH a constitué un consortium rassemblant les plateformes nationales de données de santé de plusieurs États membres (la Finlande, la Norvège, le Danemark, l'Allemagne, la Belgique, la Hongrie ou encore la Croatie), ainsi que de certaines agences européennes, comme l'Agence européenne du médicament (EMA) et le Centre européen de prévention et de contrôle des maladies (ECDC), et d'infrastructures de recherche (BBMRI, Elixir).

Financé à hauteur de 5 millions d'euros par l'Union européenne, ce projet permettra de construire et de tester une première version de l'espace européen des données de santé en interconnectant des plateformes de données, qu'il s'agisse de plateformes nationales, agences européennes et infrastructures de recherche, dans un réseau de nœuds. Le projet sera en mesure de tester un parcours utilisateur pour la création, le déploiement et la gestion de projets de recherche au niveau européens sur les données de santé. Pour ce faire, le consortium construira une infrastructure IT reliant l'ensemble des nœuds du réseau et définira des normes communes pour la sécurité, les catalogues de métadonnées, la qualité, l'interopérabilité des données et les exigences juridiques sur la base de cas d'usage proposés à la Commission européenne.

Ces cas d'usage concrets de recherche seront conduits par le consortium à l'échelle européenne, avec pour objectif de démontrer tout le potentiel de la réutilisation transnationale des données de santé pour la recherche, l'innovation, l'élaboration des politiques et de la réglementation. Ces projets de recherche pourraient couvrir des thématiques telles que la lutte contre le cancer, l'étude des maladies rares, l'évaluation de l'impact de la crise de la Covid-19, etc.

L'ensemble de ces travaux, prévus sur une durée de deux ans à partir de septembre 2022, serviront de matière à l'élaboration du règlement sur l'espace européen de données de santé actuellement à l'étude et préfigurent la mise en place pérenne de l'EHDS par la Commission européenne d'ici quelques années.

L'annonce des lauréats de l'appel à projets est attendue pour le début de l'été 2022.