

# Données de santé : libérer leur potentiel

NOTE D'ACTION - FÉVRIER 2024



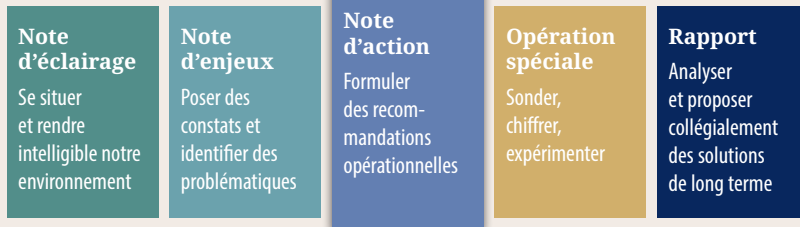
*Think tank* de référence en France et en Europe, l'Institut Montaigne est un espace de réflexion indépendant au service de l'intérêt général. Ses travaux prennent en compte les grands déterminants économiques, sociétaux, technologiques, environnementaux et géopolitiques afin de proposer des études et des débats sur les politiques publiques françaises et européennes. Il se situe à la confluence de la réflexion et de l'action, des idées et de la décision.

NOTE D'ACTION - Février 2024

## Données de santé : libérer leur potentiel



*Les notes d'action de l'Institut Montaigne identifient un enjeu spécifique et formulent des recommandations opérationnelles à destination des décideurs publics et privés.*



Les professionnels de santé et les patients coproduisent chaque jour de nombreuses données de santé. Ces données sont collectées initialement dans le cadre du soin à des fins d'utilisation primaire de la donnée de santé. **Il est pour autant possible de donner une seconde vie à la donnée de santé à travers son utilisation secondaire** qui ouvre un immense champ d'innovation et de progrès pour le secteur de la santé. **Accélération de la recherche, dépistages précoces, aide au diagnostic, santé prédictive, meilleur accès aux soins, personnalisation de la prise en charge des patients : le potentiel des données de santé promet des avancées spectaculaires pour améliorer la santé de tous.** Elles représentent également un atout précieux dans le pilotage de nos politiques de santé, notamment *via* l'analyse des besoins de santé d'une population, l'évaluation de l'efficacité de nos politiques publiques et la gestion de l'offre de soins.

La France a été un pays pionnier des données de santé avec la création du Système national des données de santé (SNDS) en 2016. Depuis, de nombreuses bases locales et thématiques, fondées sur un établissement de santé, une pathologie, un projet de recherche, ont contribué à compléter l'offre de données de santé exploitables dans notre pays. Les pouvoirs publics ont bien identifié l'enjeu que représente ces données, faisant de leur utilisation secondaire une priorité. **Gouvernance nationale renforcée, moyens spécifiques dédiés, développement des entrepôts de données hospitaliers : une stratégie nationale de l'usage des données de santé est déjà en construction.**

Notre système de santé ne peut pas se permettre de rater le tournant de l'utilisation secondaire des données de santé faute de quoi il ne sera pas en capacité de se transformer, d'innover et d'adopter un virage résolument préventif, au plus près des besoins de santé du patient. Ce tournant ne sera possible qu'à la condition de relever des **défis d'ordre technique, juridique, économique et citoyen.**

Ainsi, cette note avance cinq propositions pour accélérer l'utilisation secondaire des données de santé et favoriser leur potentiel :

- 1. Mettre en place des règles de standardisation des bases de données afin d'assurer leur interopérabilité.**
- 2. Développer un réseau d'entrepôts de données de santé dédiés à la médecine de ville.**
- 3. Intégrer au sein des établissements de santé, des *Data Pôles* chargés de définir une stratégie des données de santé.**
- 4. Faire émerger un modèle économique de la donnée grâce à des contrats et tarifications types permettant de favoriser le partage des données de santé et de rémunérer les producteurs de bases.**
- 5. Créer une culture de la donnée auprès des professionnels et des citoyens grâce à des campagnes de sensibilisation et la création d'un registre national d'opposition.**

## Emma Ros

---

Emma Ros est chargée de projets santé à l'Institut Montaigne depuis septembre 2021, où elle a porté des publications sur les défis du virage ambulatoire et la territorialisation des politiques de santé. Diplômée de Sciences Po Paris et de la *London School of Economics*, elle a développé une expertise sur les enjeux de transformation des pratiques de soins grâce au numérique, de soutien aux soins primaires et d'amélioration de l'accès aux soins de santé mentale. Elle accompagne par ailleurs le déploiement de l'expérimentation SÉSAME (Soins d'équipe en santé mentale), portée par l'Institut Montaigne et qui fait aujourd'hui l'objet d'un financement dérogatoire au titre du dispositif Article 51 de la loi de financement de la Sécurité sociale de 2018.

## 1

Synthèse .....	5
<b>Introduction</b> .....	12
<b>Des données de santé de natures diverses, collectées à des fins variées et hébergées dans des bases éparpillées et multiples</b> .....	18
<b>1.1.</b> Une base nationale à vocation universelle : le Système national des données de santé (SNDS) .....	19
<b>1.2.</b> La structuration progressive de bases locales ou associées à des projets spécifiques .....	24
<b>a.</b> Des bases nationales thématiques sur des pathologies précises .....	24
<b>b.</b> Des bases locales en cours de structuration : les entrepôts de données de santé .....	28
<b>1.3.</b> Des champs encore peu couverts et des données difficilement exploitables .....	31
<b>a.</b> Un périmètre insuffisamment couvert : la médecine de ville .....	31
<b>b.</b> Des données externes trop peu exploitées .....	33
<b>1.4.</b> Des actions nationales et européennes ayant contribué à la construction d'un cadre sur l'utilisation secondaire de la donnée de santé .....	35
<b>a.</b> Une stratégie française en cours de construction .....	35
<b>b.</b> Une stratégie européenne aux contours encore peu précis .....	38

## 2

<b>Les cas d'usage de la donnée de santé à des fins d'utilisation secondaire</b> .....	41
<b>2.1.</b> L'utilisation des données de santé pour optimiser le pilotage de notre système de santé .....	43
<b>a.</b> Un outil participant à l'efficacité du pilotage du système de soins .....	43
<b>b.</b> Un élément essentiel des politiques de santé publique .....	44
<b>2.2.</b> L'utilisation des données de santé pour transformer les pratiques et l'organisation des soins .....	46
<b>a.</b> Un outil essentiel pour le diagnostic, les dépistages précoces et la prévention .....	46
<b>b.</b> Des innovations technologiques pour l'aide à la gestion des établissements et de notre système de santé .....	47
<b>2.3.</b> L'utilisation des données de santé pour la recherche .....	49
<b>a.</b> Une source d'avancées significative pour la recherche médicale .....	49
<b>b.</b> Une contribution indéniable à l'amélioration de la qualité des soins .....	50
<b>Des défis qui subsistent freinant les usages de la donnée secondaire en France</b> .....	54
<b>3.1.</b> Un défi technique : des bases de données de santé à standardiser et à rendre interopérables .....	55
<b>a.</b> Un enjeu d'interopérabilité entre les données de santé et les différentes bases .....	55

## 3

<b>b.</b> La médecine de ville : l'enjeu de la collecte et de la constitution des bases de données .....	61
<b>c.</b> Un retraitement souvent nécessaire pour la constitution des bases de données de santé .....	64
<b>3.2.</b> Un défi économique : l'insuffisante valorisation économique de la donnée et du travail de constitution des bases de données .....	69
<b>a.</b> L'utilisation secondaire de la donnée : une collecte, un traitement et un hébergement coûteux .....	69
<b>b.</b> Un système de redevance à mettre en place pour les producteurs des bases de données .....	72
<b>3.3.</b> Un défi juridique : un cadre réglementaire protecteur mais complexe à appréhender .....	76
<b>a.</b> Une procédure complexe pour accéder aux données de santé .....	77
<b>b.</b> Des délais d'instruction des dossiers jugés trop longs .....	80
<b>c.</b> Un cadre réglementaire difficile à appréhender pour les acteurs .....	82
<b>3.4.</b> Un défi citoyen : une nécessaire sensibilisation des citoyens et des professionnels de santé à l'utilisation secondaire .....	84
<b>a.</b> Des citoyens informés de l'utilisation des données de santé mais désireux de la construction d'un cadre de confiance renforcé .....	84
<b>b.</b> L'information et le consentement des patients : une obligation juridique, garante de la construction d'un cadre de confiance .....	86
<b>c.</b> Des pistes pour créer un cadre de confiance capable d'alimenter et maintenir une culture de la donnée .....	89

<b>Conclusion</b> .....	93
-------------------------	----

<b>Annexes</b> .....	94
<b>Annexe 1 - Méthodologies de références</b> .....	94
<b>Annexe 2 - Fiches pays</b> .....	96
Finlande .....	96
Danemark .....	97
Royaume-Uni .....	99

<b>Remerciements</b> .....	101
----------------------------	-----

<b>Liste des acronymes et des abréviations</b> .....	105
------------------------------------------------------	-----

Les professionnels de santé et les patients coproduisent chaque jour de nombreuses données qui viennent alimenter les dossiers médicaux, les bases de remboursement de l'Assurance maladie ou encore les systèmes d'information de nos établissements de santé. Ces données sont collectées initialement dans le cadre du soin, c'est-à-dire entre un patient et un professionnel de santé, et regroupées au sein de dossiers médicaux électroniques. **On parle alors d'utilisation primaire<sup>1</sup> de la donnée de santé.**

Ces données contiennent des informations précieuses sur l'état de santé de notre population, les parcours de soins des patients et *in fine* sur l'efficacité de nos politiques publiques. À ce titre, elles représentent un potentiel immense pour améliorer la performance et le pilotage de notre système de santé, la qualité de la prise en charge des patients ainsi que la recherche.

**Il est possible de donner une seconde vie à la donnée de santé à travers l'utilisation secondaire : il s'agit ici de tirer avantage des données récoltées dans le cadre du soin pour permettre de mieux comprendre le profil clinique d'une population, l'organisation du système de soins ou encore de trouver un traitement innovant pour une pathologie précise.** Les données individuelles des patients peuvent être ainsi regroupées et structurées au sein de bases. Cette massification des données est essentielle afin de déployer leur plein potentiel et regrouper des effectifs significatifs<sup>2</sup> de la population. **Il existe de nombreux cas d'usage pour les données de santé qui sont sources de progrès considérables, que ce soit pour le pilotage du système de santé** (analyse des besoins de santé d'une population, évaluation de la qualité des soins ou d'une politique de santé, gestion

<sup>1</sup> HAS, *Entrepôts de données de santé hospitaliers en France, Quel potentiel pour la Haute Autorité de santé ?*, 2022.

<sup>2</sup> Conseil scientifique consultatif du Health Data Hub, *Bénéfices et risques de l'utilisation des données de santé à des fins de recherche*, Position paper, janvier 2024.

d'une épidémie), **l'amélioration de la prise en charge des patients à toutes les étapes du parcours de soins** (médecine prédictive, aide au diagnostic, personnalisation des actes et des traitements) **ou encore le développement de l'innovation et de la recherche** (à travers le développement d'algorithmes d'intelligence artificielle ou la conduite d'essais cliniques).

**Cette réutilisation de nos données constitue donc un atout considérable** pour la recherche médicale, la qualité des soins et l'amélioration de notre système de santé. La Stratégie mondiale pour la santé numérique 2020-2025 de l'Organisation mondiale de la santé indiquait que « *le lien entre la production secondaire d'informations et leur utilisation pour les soins, la promotion de la santé et la prévention crée un cycle d'amélioration continue qui profite à chacun*<sup>3</sup> ». Chercheurs, soignants, patients, régulateurs du système de soins, citoyens : tous les acteurs ont un intérêt à plus et mieux exploiter les données produites par notre système de santé.

De par sa tradition centralisatrice, notre pays a été pionnier dans la structuration de ses données de santé avec la création dès les années 1970 d'un système de codage des actes médicaux et hospitaliers. Il a permis de créer les bases nationales du Système national d'information inter-régimes de l'Assurance maladie (Sniiram) et du Programme de médicalisation des systèmes d'information (PMSI), qui contiennent des informations à l'échelle nationale sur le niveau de consommation de soins de la médecine de ville et des établissements de santé. La loi de modernisation de notre système de santé de 2016<sup>4</sup> est allée encore plus loin en regroupant ces deux bases nationales au sein d'une base agrégée, unique au monde en termes de volume, de maturité et de profondeur des données : le Système national des données de santé (SNDS).

<sup>3</sup> Organisation mondiale pour la santé, *Stratégie mondiale pour la santé numérique 2020-2025*, août 2021.

<sup>4</sup> Loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé.

La loi de 2019 relative à l'organisation et à la transformation du système de santé<sup>5</sup> a par la suite créé la Plateforme des données de santé (PDS), plus communément nommée le *Health Data Hub* (HDH)<sup>6</sup>, première plateforme nationale des données de santé. Enfin, des investissements conséquents ont été réalisés à partir de 2022, afin de renforcer les entrepôts de données de santé hospitaliers existants, bases locales décentralisées, qui ont pour objectif de structurer les données des activités de soins des hôpitaux<sup>7</sup>. Ces avancées majeures ont permis de renforcer la gouvernance autour des données de santé en France. C'est pourquoi le *Panorama de la santé* de l'OCDE de 2023 classait notre pays en quatrième position en matière de gouvernance des ensembles de données de santé, derrière le Danemark, les États-Unis et la Finlande.

**Pour autant, force est de constater que les initiatives gouvernementales des dernières années, initiées pour structurer l'offre en matière de données, peinent aujourd'hui à se déployer de manière effective.**

Au même classement de l'OCDE, la France n'atteint que la 14<sup>e</sup> place en ce qui concerne la maturité de ses bases<sup>8</sup> ainsi que la possibilité d'accéder aux ensembles de données de santé et de les coupler<sup>9</sup>. Si nos grandes bases nationales et locales regroupent une grande quantité de données, elles demeurent peu utilisées, font encore preuve de niveaux de maturité inégaux et ne sont pas toutes interopérables. Par ailleurs, leur multiplicité (on recense 172 bases de données<sup>10</sup> de santé en France) et l'absence d'une cartographie recensant les bases de données existantes ainsi que

<sup>5</sup> Loi n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé.

<sup>6</sup> « Le HDH, au service de la recherche sur les données de santé depuis trois ans », site internet du Health Data Hub, décembre 2022.

<sup>7</sup> « Entrepôts de données EDS », site du Ministère du Travail, de la Santé et des Solidarités, consulté le 17 janvier 2024.

<sup>8</sup> Jillian Oderkirk, "Survey Results: National Health Data Infrastructure and Governance", OECD Health Working Papers, n° 127, 2021.

<sup>9</sup> *Panorama de la santé 2023, les indicateurs de l'OCDE.*

<sup>10</sup> « Inventaire des bases de données relatives à la santé », site internet de [data.gouv.fr](https://data.gouv.fr).

leur contenu entraîne une difficulté supplémentaire pour identifier la base de données adéquate selon le projet à mener. La France accuse ainsi un certain retard en termes de publications relatives à ses bases de données, moins nombreuses et moins accessibles que certaines bases étrangères<sup>11</sup>.

Les précédentes réformes, bien qu'ayant impulsé un mouvement autour de l'utilisation secondaire des données de santé, n'ont pas permis de créer une stratégie nationale cohérente. **D'importants défis persistent pour permettre une utilisation secondaire optimale des données de santé :**

- **Un défi technique :** la constitution d'une base de données à partir des données générées dans le cadre du soin est un processus complexe. Il implique plusieurs étapes, notamment la collecte, la standardisation et le stockage. Chaque base de données nécessite un travail conséquent de mise à niveau des données issues des dossiers médicaux afin d'assurer leur exploitation. La médecine de ville, au regard de la diversité des outils de saisie utilisés par les professionnels de santé de ville, constitue un champ encore non-couvert par des bases de données complètes. Une fois la base constituée, l'absence de standards d'interopérabilité complique la constitution d'un réseau de bases de données interopérables entre elles et rend plus difficile les initiatives de recherche.
- **Un défi économique :** l'absence de cadre économique et de réelle valorisation de la donnée de santé n'incite pas les acteurs, publics comme privés, à partager la donnée, voire à investir dans la constitution de bases de données. La construction d'une base de données implique des investissements importants. La création d'un cadre économique national autour du partage de la donnée de santé représente une priorité pour favoriser leur exploitation.

<sup>11</sup> Jérôme Marchand-Arvier, Pr Stéphanie Allasonnière, Aymeril Hoang, Dr Anne-Sophie Jannot, « Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire des données de santé », décembre 2023.



- **Un défi juridique** : les données de santé sont des données catégorisées comme sensibles. À ce titre, un cadre réglementaire complexe et rigoureux existe pour protéger les données d'usages potentiellement malveillants. Néanmoins, la complexité de ce cadre ne permet pas sa pleine appropriation par les acteurs et rallonge les délais d'accès aux données de santé pour les porteurs de projet. À titre d'exemple, on estime à 18 mois<sup>12</sup> le délai moyen d'accès au SNDS. Ces délais particulièrement longs ont un effet délétère sur la recherche française et l'attractivité de nos bases de données.
- **Un défi citoyen** : le patient-citoyen est à la fois le contributeur et le bénéficiaire final des projets impliquant l'utilisation secondaire des données de santé. À ce titre, la valorisation des usages et des bénéfices de la donnée de santé, auprès des patients comme des professionnels de santé, représente un levier majeur d'adhésion du citoyen à cette révolution du système de santé. La construction d'un cadre de confiance, assurant la transparence, la sécurité et l'éthique de cette réutilisation est donc une condition indispensable à cette pleine adhésion.

**La stratégie nationale sur la réutilisation et l'ouverture des données de santé<sup>13</sup>, dont la publication est prévue pour le premier semestre 2024, devrait participer à lever ces freins.** Des premiers éléments de cette stratégie ont été dévoilés en janvier 2024, avec la publication d'un rapport de la mission interministérielle pilotée par Jérôme Marchand-Arvier<sup>14</sup>. Par ailleurs, cette stratégie s'inscrit dans un cadre européen en pleine mutation, avec notamment la construction d'un espace européen des données de santé.

<sup>12</sup> Commission des affaires sociales, Sénat, « Données de santé : une réforme encore en cours de chargement », Rapport d'information, juillet 2023.

<sup>13</sup> Feuille de route du numérique en santé 2023-2027, « Mettre le numérique au service de la santé ».

<sup>14</sup> Jérôme Marchand-Arvier, Pr Stéphanie Allasonnière, Aymeril Hoang, Dr Anne-Sophie Jannot, « Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire des données de santé », décembre 2023.

**C'est dans ce contexte, et pour contribuer à cette réflexion, que cette note présente l'écosystème des données de santé, en rappelle les principaux usages et potentialités, avant d'identifier les freins à leur déploiement effectif.** Enfin, nous formulons des préconisations concrètes pour lever ces freins et accélérer la mise en œuvre d'une stratégie majeure pour notre système de santé.

# 1 Des données de santé de natures diverses, collectées à des fins variées et hébergées dans des bases éparpillées et multiples

Le Règlement européen sur la protection des données<sup>15</sup> (RGPD) définit la donnée de santé<sup>16</sup> comme une donnée qui renseigne sur la santé physique ou mentale d'une personne. On distingue généralement les catégories de données de santé suivantes :

- **les données médico-administratives** : elles sont générées à des fins de remboursement ou de régulation du système de santé. Elles ne comportent pas de caractère médical ;
- **les données cliniques du patient** : elles renseignent sur l'état de santé des patients et donnent des informations telles que les antécédents médicaux, les résultats aux tests médicaux, etc. ;
- **les données générées par le patient** dans le cadre du développement des applications numériques, des dispositifs médicaux connectés et le déploiement de la télésurveillance<sup>17</sup> ;
- **les données « hors santé »** : des données environnementales, socio-économiques ou socio-démographiques qui peuvent renseigner sur un déterminant de santé et être croisées avec des données de santé.

**Ces données, principalement collectées dans le cadre du soin, peuvent être par la suite stockées au sein de bases de données à des fins d'utilisation secondaire.** En 2021, on recensait 172 bases de données de santé en France<sup>18</sup> gérées par de nombreux acteurs, à l'image

<sup>15</sup> Nesrine Benyahia, « Cinq ans après le RGPD, quelles protections pour les données de santé », Institut Montaigne, mai 2023.

<sup>16</sup> « Qu'est-ce qu'une donnée de santé ? », site internet de la [CNIL](#).

<sup>17</sup> « Généralisation de la prise en charge des patients en télésurveillance pour certaines pathologies », site internet de [l'Assurance maladie](#), juillet 2023.

<sup>18</sup> « Inventaire des bases de données relatives à la santé », site internet de [data.gouv.fr](#). D'après [data.gouv.fr](#), on comptabilise 79 gestionnaires de bases de données en France.

d'acteurs institutionnels tels que la Caisse nationale d'assurance maladie (Cnam) ou l'Agence nationale de sécurité du médicament et des produits de santé (ANSM), d'instituts de recherche ainsi que d'acteurs privés, tels que les laboratoires pharmaceutiques. **Elles présentent des niveaux de maturité différents et regroupent des données de natures extrêmement variées, qui ont été collectées pour des finalités diverses (soins, remboursement, recherche).** Elles agrègent des données individuelles de patients, qui ont été au préalable pseudonymisées<sup>19</sup> (toutes les caractéristiques personnelles directes de la personne, telles que son nom ou sa date de naissance sont retirées et remplacées par un identifiant). Il faut bien noter que le plein potentiel de la donnée de santé n'est atteint qu'à partir du regroupement d'effectifs conséquents : la donnée individuelle d'un seul patient n'est jamais analysée, l'objectif étant de regrouper des échantillons significatifs de la population.

L'ensemble de ces bases de données demeure très hétérogène quant à la qualité et à l'intentionnalité des données qu'elles contiennent. Ces données ont été recueillies dans des cadres distincts, selon des finalités différentes et ont été pensées dans un cadre ou un contexte spécifique (celui du soin ou de la recherche)<sup>20</sup>. Cet enregistrement initial a un impact direct en termes de qualité et de pertinence des données.

## 1.1. UNE BASE NATIONALE À VOCATION UNIVERSELLE : LE SYSTÈME NATIONAL DES DONNÉES DE SANTÉ (SNDS)

La France a été un pays pionnier dans la structuration de ses données de santé dès les années 1970<sup>21</sup>. Nous disposons aujourd'hui d'une vaste base nationale de données de santé. **Le système national des données**

<sup>19</sup> « Recherche scientifique (hors santé) : enjeux et avantages de l'anonymisation et de la pseudonymisation », site internet de la [CNIL](#), janvier 2022.

<sup>20</sup> « Le pouvoir des données de santé », West Data Festival, mai 2023.

<sup>21</sup> Commission des affaires sociales, Sénat, « Données de santé : une réforme encore en cours de chargement », Rapport d'information, juillet 2023.

**de santé (SNDS)<sup>22</sup> est unique dans son genre, tant par le volume des données qu'il agrège** (1,5 milliard de feuilles de soins de ville et 12 à 15 millions de séjours hospitaliers) que par **la représentativité de la population couverte ou l'historique des informations** (allant jusqu'à 20 ans)<sup>23</sup>.

Les données rassemblées au sein du SNDS sont de nature médico-administratives. Leur collecte répond initialement à des finalités de remboursement des soins pour les assurés. Elles servent également, pour le régulateur du système de soins, à mesurer l'activité médicale et paramédicale des professionnels de santé et des établissements hospitaliers, afin de déterminer leurs niveaux de financement et de remboursement. **Les données du SNDS ne comportent donc pas directement des informations de nature médicale** (motif de consultation, diagnostic, parcours de soins, observance des traitements)<sup>24</sup> **et manquent de finesse clinique.**

Hébergé à la Caisse nationale de l'Assurance maladie (Cnam), le SNDS est un immense agrégateur qui rassemble les bases de données suivantes :

- **le système national d'information inter-régimes de l'Assurance maladie (Sniiram)** qui compile depuis 1999 les dépenses liées à l'activité des professionnels de santé dans l'ensemble des régimes de base de l'Assurance maladie ;
- **le programme de médicalisation des systèmes d'information (PMSI)** qui collecte des données relatives à l'activité médicale des établissements de santé. Cette base de données est gérée par l'Agence technique de l'information sur l'hospitalisation (ATIH) qui la transmet mensuellement à la Cnam pour alimenter le SNDS ;

<sup>22</sup> « Le système national des données de santé (SNDS) et l'accès aux données de santé », site internet de la *Drees*, décembre 2016.

<sup>23</sup> Assurance maladie, « Données de santé en France : état des lieux et enjeux des bases de données au service des études, de la recherche et de l'évaluation, rôle de l'Assurance maladie, Dossier de presse », avril 2023.

<sup>24</sup> *Ibid.*

- **la base de données sur les causes médicales de décès (BCMD)**, produite par l'unité CépiDC de l'Inserm qui restitue les données des certificats d'attestation de décès ;
- **les données issues des bases SI-DEP et Data vaccin Covid, utilisées dans la gestion de la pandémie de Covid-19**, récemment intégrées au SNDS<sup>25</sup> ;
- **les données du système d'information alimenté par les maisons départementales des personnes handicapées (MDPH)**, piloté par la Caisse nationale de solidarité pour l'autonomie (CNSA), qui contient les informations sur les demandes d'accompagnement pour les personnes handicapées<sup>26</sup>. Ces dernières seront intégrées dans le SNDS à compter de 2024.

**La loi relative à l'organisation et à la transformation du système de santé du 24 juillet 2019<sup>27</sup> a ouvert très largement le périmètre des données pouvant être hébergées par le SNDS.** Toute donnée associée à des soins financés par l'Assurance maladie, collectée à l'occasion d'activités de prévention, de diagnostic, de soins ou de suivi médico-social, peut désormais être intégrée au sein du SNDS. Ainsi, un catalogue d'une dizaine de bases devrait être ajouté au SNDS, notamment :

- la base Epicov, enquête statistique sur l'état de santé et les conditions de vie en lien avec l'épidémie due au coronavirus<sup>28</sup> portée par l'Inserm et la Drees ;

<sup>25</sup> Délibération n° 2022-061 du 19 mai 2022 portant avis public sur les conditions de mise en œuvre des systèmes d'information développés aux fins de lutter contre la propagation de l'épidémie de COVID-19 (octobre 2021 à avril 2022), CNIL.

<sup>26</sup> « Le système national des données de santé (SNDS) et l'accès aux données de santé, site internet de la Drees, décembre 2016. Depuis 2018 et sur la base des travaux réalisés par l'ANS, la CNSA pilote la construction d'un système d'information MDPH devant permettre d'homogénéiser les informations autour de l'accompagnement des personnes handicapées par les MDPH afin de constituer une vision nationale.

<sup>27</sup> Loi n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé. Arrêté du 12 mai 2022 relatif aux données alimentant la base principale et aux bases de données du catalogue du système national des données de santé.

<sup>28</sup> Enquête EpiCov (Épidémiologie et Conditions de vie sous le Covid-19), Drees, mai 2021.

- la base Hepather (renseignant sur les hépatites virales) construite par l'Inserm<sup>29</sup>;
- la base OSCOUR sur les données de passage aux urgences, gérée par Santé publique France<sup>30</sup>.

Ces ajouts renforceront la vocation universelle du SNDS qui va centraliser un volume toujours plus important de données de santé. **Ce modèle, peu répandu dans des pays comparables, est un véritable atout pour le système français.** Il permet de rendre l'offre de soins plus visible pour les acteurs de la santé, de faciliter les appariements entre bases et ainsi de disposer d'une plus grande complétude des données de santé.

### Danemark : des registres nationaux sur de grandes pathologies qui regroupent les données de l'ensemble des citoyens

**Le Danemark a choisi de constituer des bases nationales contenant de la donnée clinique. À la différence du SNDS français, ces bases comportent ainsi un caractère médical sur des pathologies précises.**

Au Danemark, le *Danish Health Data Authority* est l'entité responsable de la maintenance des registres nationaux<sup>31</sup>. Ces registres concernent des sujets variés allant de la santé de l'enfant aux médicaments jusqu'aux principales pathologies (cancers, maladies psychiatriques). Ces registres regroupent les données de

l'ensemble des citoyens danois. Dans chacun des registres, les patients sont identifiés par leur numéro d'identification personnel danois, permettant de relier les données des registres entre elles.

La Cnam développe de plus en plus des outils de *data visualisation*, disponibles en *open data* à partir des données du SNDS ciblés sur des pathologies<sup>32</sup> spécifiques à l'image de la *plateforme data pathologies* ou encore de la récente plateforme *data professionnels de santé libéraux*<sup>33</sup>. Ces plateformes permettent de visualiser de manière agrégée et anonyme les données du SNDS et renseignent sur la prévalence de 50 pathologies clés ainsi que sur les effectifs, les honoraires ou encore les prescriptions des professionnels de santé libéraux.

D'autres données récoltées à des finalités de remboursements des soins peuvent venir compléter les données issues du SNDS, comme par exemple celles issues des organismes complémentaires. Ces institutions collectent les données des soins de la majorité des Français qui, pour plus de 95 %<sup>34</sup> d'entre eux, sont couverts par une assurance maladie complémentaire. Le décret de 2016 créant le SNDS, prévoyait à ce titre « *qu'un échantillon de données de remboursement d'assurance maladie complémentaire* » intègre le SNDS<sup>35</sup>.

<sup>32</sup> *Data pathologies*, Assurance maladie.

<sup>33</sup> *Data professionnels de santé libéraux*, Assurance maladie.

<sup>34</sup> « *Couverture de la population par l'assurance maladie complémentaire* », site internet de la Sécurité sociale.

<sup>35</sup> Décret n° 2016-1871 du 26 décembre 2016 relatif au traitement de données à caractère personnel dénommé « *système national des données de santé* ».

<sup>29</sup> Portail Épidémiologie France.

<sup>30</sup> « Réseau OSCOUR », site de Santé publique France, novembre 2021.

<sup>31</sup> « List of national health registers », site internet *The Danish Health Authority*.

## 1.2. LA STRUCTURATION PROGRESSIVE DE BASES LOCALES OU ASSOCIÉES À DES PROJETS SPÉCIFIQUES

**Parallèlement au SNDS, de nombreuses bases de données spécifiques à certains projets scientifiques se sont développées.** Les initiatives sont diverses et portées par de multiples acteurs de la recherche et du soin. Elles permettent d'avoir une connaissance plus fine de pathologies spécifiques ou des problématiques territoriales de santé. **Néanmoins, leur multiplicité, leur hétérogénéité et le manque de documentation sur les informations qu'elles contiennent complexifient leur identification et mobilisation.** La récente mission interministérielle autour des données de santé pointe à juste titre la nécessité d'une cartographie transverse des bases de données, à même de décrire le contenu des bases<sup>36</sup>.

- a. Des bases nationales thématiques sur des pathologies précises

**Il existe un grand nombre de bases sur des pathologies spécifiques, gérées généralement par des organismes de recherche.** À titre d'exemple, sur la recherche en cancer, on peut citer les bases suivantes (*liste non exhaustive*) :

- L'Institut national du cancer (INCa) développe depuis 2019 une plateforme de données en cancérologie<sup>37</sup> regroupant des données cliniques, des données issues de programmes de soins ou des données de génétique moléculaire.

- La base OncoDataHub<sup>38</sup> a été récemment développée par Unicancer et le laboratoire Roche. Elle réunit les données en vie réelle de la base ESMÉ<sup>39</sup> d'Unicancer ; regroupant l'analyse de données de 70 000 patients ; et de la base PRM<sup>40</sup> de Roche, comportant les données de prescription des traitements anticancéreux<sup>41</sup>.
- Enfin, lancé en 2021 par Unicancer et le *Health Data Hub*, le programme Unibase a pour but de créer, sur une durée de 3 ans, une collection de bases de référence en cancérologie<sup>42</sup>. À titre d'exemple, le projet HORUS, porté par l'Institut Gustave Roussy et l'Inserm, a pour objectif de créer une cohorte de patients traités avant l'âge de 25 ans pour un cancer pédiatrique.

De nombreuses autres bases existent sur une immense variété de pathologies, à l'image de la base Orphanet<sup>43</sup> de l'Inserm renseignant sur les maladies rares et comportant les données de 42 pays. On peut également citer les bases de données de Santé publique France relatives à l'épidémie de grippe en France.

<sup>38</sup> « Onco Data Hub », site internet recherche Unicancer.

<sup>39</sup> « Épidémiologie-Stratégie Médico-Économique ».

<sup>40</sup> « Modèles de Remboursement Personnalisé ».

<sup>41</sup> « Onco Data Hub », site internet recherche Unicancer.

<sup>42</sup> « Unicancer, la Ligue nationale contre le cancer et le HDH annoncent les lauréats de la seconde vague de l'AMI UNIBASE », site internet du *Health Data hub*, mai 2023.

<sup>43</sup> « Bases de données pour la recherche en santé », site internet de l'Inserm, juin 2022.

<sup>36</sup> Jérôme Marchand-Arvier, Pr Stéphanie Allasonnière, Aymeril Hoang, Dr Anne-Sophie Jannot, « Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire des données de santé », décembre 2023.

<sup>37</sup> « La plateforme de données en cancérologie », site internet de l'Institut national du Cancer.

Corée du Sud : un réseau national  
sur des données cliniques et biologiques :  
le *Korea Biobank Network*<sup>44</sup>

**Projet initié par le ministère sud-coréen du « Bien-être et de la Santé » en 2008, le *Korea Biobank Network* est un réseau national de biobanques.** Il collecte, gère et distribue les ressources biologiques humaines de la biobanque nationale de Corée du Sud, ainsi que de dix-sept banques régionales situées dans des hôpitaux universitaires (ressources pathologiques : sang, tissus, urine, conformément à un accord de don volontaire signé par les patients). Pour accéder à ces banques de données, il faut se rendre sur le portail du *Korea Biobank Network*, qui est la plus grande plateforme ouverte partagée pour les ressources détenues par les biobanques régionales. **Ces dernières jouent ainsi un rôle important en tant qu'infrastructures publiques de soutien à la recherche.**

Enfin, il est nécessaire de citer la présence de grandes cohortes nationales, financées notamment par les investissements d'avenir<sup>45</sup>. Les études de cohorte permettent d'observer la même population dans le temps long et suivre son évolution de santé. Elles constituent l'un des outils les plus efficaces pour la recherche médicale et l'épidémiologie<sup>46</sup>.

<sup>44</sup> «Korea Biobank Network», site internet du [KBN](#).

<sup>45</sup> « Le Programme d'investissements d'avenir », site internet du [gouvernement](#), décembre 2021.

<sup>46</sup> « Qu'est-ce qu'une cohorte en santé », site internet de [France Cohortes](#).

On distingue deux types de cohortes :

- **les cohortes en population générale**, à l'image de Constances<sup>47</sup>, cohorte épidémiologique « généraliste » mise en place en 2012<sup>48</sup> et constituée d'un échantillon représentatif de 200 000 adultes âgés de 18 à 69 ans ;
- **les cohortes cliniques**, à l'image de la cohorte CANTO lancée en 2012 qui regroupe les données de 13 250 femmes atteintes de cancer du sein au stade 1 à 3<sup>49</sup>.

Depuis 2020, une initiative de l'Inserm, France Cohortes<sup>50</sup> a pour objectif de développer un centre de ressources et de services (recueil et traitement juridique, communication, valorisation) à disposition des chercheurs travaillant à partir de données de cohortes<sup>51</sup>.

**Toutes ces bases thématiques nationales, notamment celles issues de vastes projets de recherche, forment un ensemble hétérogène.** Elles présentent des modalités de gouvernance et d'accessibilité aux données variées<sup>52</sup> et la gestion de ces données demeure peu centralisée.

<sup>47</sup> Site internet de la [cohorte Constances](#).

<sup>48</sup> La cohorte Constances, Résumé du protocole, UMS 011, Inserm, mai 2022.

<sup>49</sup> « CANTO », site internet de [France Cohortes](#).

<sup>50</sup> Site internet de France Cohortes. Avec une quinzaine de cohortes associées, [France Cohortes](#) comptabilise plus de 600 000 participants et œuvre à valoriser l'exploitation scientifique des données recueillies, renforcer leur protection et faciliter l'interopérabilité.

<sup>51</sup> « Qui sommes nous ? », Site internet de [France Cohortes](#).

<sup>52</sup> Bernard Bégaud, Dominique Polton, Franck von Lenep, « Les données de vie réelle, un enjeu majeur pour la qualité des soins et la régulation du système de santé – L'exemple du médicament », mai 2017.

**b. Des bases locales en cours de structuration :  
les entrepôts de données de santé**

**Depuis quelques années, de plus en plus d'entrepôts de données de santé (EDS) se développent dans les établissements hospitaliers.** Ils agrègent, à l'échelle d'un ou plusieurs établissements, des données de systèmes d'information médicaux des différents services d'un hôpital. **Ces données sont de natures diverses et peuvent avoir des finalités variées : le pilotage de l'activité de soins, le suivi des soins ou des travaux de recherche.** L'entrepôt de données de santé est ainsi une infrastructure informatique sécurisée sur laquelle il est possible de regrouper les données issues de différents logiciels métiers de l'hôpital. La constitution de cette base permet de piloter de manière plus fine les activités de l'établissement et de mobiliser les données de l'hôpital plus facilement dans le cadre de projets de recherche<sup>53</sup>. **L'intérêt de ces bases de données réside dans le fait qu'elles recueillent des données de nature médicale, notamment celles issues des comptes-rendus de consultations, des données biologiques ou encore d'imagerie.** À titre d'exemple, l'EDS de l'AP-HP intègre aujourd'hui les données de plus de 14 millions de patients, 40 millions d'actes de soin, 45 millions de diagnostics, 103 millions de comptes rendus, 35 millions d'examen d'imagerie médicale et 2 500 millions d'analyses biologiques<sup>54</sup>.

En 2022, 30 projets<sup>55</sup> d'EDS étaient recensés par la Fédération hospitalière de France (FHF), portés par des Centres hospitalo-universitaires (CHU). Autre exemple en région, l'entrepôt des données de santé du CHU de Nantes a été créé en 2019 afin d'agréger plus de 50 millions de données, 2,3 millions de dossiers patients et 16 millions de documents médicaux.

<sup>53</sup> « EDS CITY – 002 : Formation citoyenne aux entrepôts de données de santé (EDS) », *Health Data Hub*, 12 juillet 2023.

<sup>54</sup> « L'Entrepôt de Données de Santé de l'AP-HP », site internet de l'AP-HP.

<sup>55</sup> Rapport d'information n° 873 (2022-2023), déposé le 12 juillet 2023.

**Royaume-Uni : des hubs au niveau local  
regroupant des données sur une aire  
thérapeutique dédiée**

Depuis 2019<sup>56</sup>, le *Health Data Research*, en partenariat avec le gouvernement britannique, a créé des centres d'excellence sur les données de santé, les *Hubs*, implantés dans différentes villes du Royaume-Uni.

Chacun des *Hubs* a pour objectif de développer des connaissances et des données spécifiques sur une thématique ciblée. Chaque *Hub* est le fruit d'une collaboration étroite entre le NHS, les universités, l'industrie et le public et fait partie intégrante de l'écosystème dynamique des données de recherche au Royaume-Uni. Il y a actuellement neuf *Hubs* qui couvrent plusieurs domaines de maladies et sources de données. À titre d'exemple, *DataMind* est le *Hub* dédié à la santé mentale et *Insight* est celui dédié à la santé de l'œil.

**Le développement des EDS a été soutenu par un programme de transformation numérique déployé dans le cadre du Ségur de la Santé en 2020<sup>57</sup> qui prévoyait des incitations financières pour les établissements se dotant d'EDS.** Des fonds du programme « France 2030 »<sup>58</sup> ont également été mobilisés à compter de 2022 pour renforcer ces bases de données locales *via* le lancement d'un appel à projets financé à hauteur de 75 millions d'euros. Au total, seize lauréats ont été retenus en deux

<sup>56</sup> « Health Data Research Hubs », site internet du *Health Data Research UK*.

<sup>57</sup> « Le Ségur du numérique en santé », site internet de l'Agence du numérique en santé.

<sup>58</sup> « France 2030 : accélération de la croissance des entreprises en santé numérique : les lauréats de l'Appel à Manifestation d'intérêt santé numérique », Dossier de presse, septembre 2022.

vagues pour permettre la construction d'un réseau territorial<sup>59</sup>. L'un des lauréats de cet appel à projets est l'*Ouest Data Hub*, plateforme interrégionale regroupant les entrepôts des données de santé des Hôpitaux Universitaires Grand Ouest (HUGO), soit des CHU de Rennes, Brest, Angers, Nantes et Tours. Il est ainsi le premier réseau européen de *big data* en santé avec 2,5 milliards de données structurées à partir des données de 9,3 millions de patients et 11 millions de séjours<sup>60</sup>. Par ailleurs, la création de 16 nouveaux pôles d'excellence en recherche en santé<sup>61</sup> (dont 12 Instituts Hospitalo-Universitaires et 4 bioclusters), également financés dans le cadre de la stratégie d'investissement France 2030, devrait être source d'émergence de nouveaux entrepôts de données.

Les entrepôts de données de santé constituent un potentiel majeur pour l'utilisation des données de santé car ils sont le pendant clinique des données médico-administratives hébergées au sein du SNDS. L'exploitation directe de leurs données se heurte toutefois à plusieurs difficultés. Tout d'abord, la nature et le périmètre des données collectées dans les EDS sont le reflet des systèmes d'information utilisés par l'établissement. Leur qualité et leur caractère exploitable dépendent directement des systèmes d'information sources, et notamment de leur capacité à héberger correctement des dossiers finement renseignés par les professionnels de santé. **Par ailleurs, la structuration des EDS reste encore très éclatée, sans standards d'interopérabilité partagés et dépendante des initiatives locales et des financements associés.** Les initiatives des dernières années, bien qu'essentielles pour le développement des entrepôts de données de santé, n'ont pas permis de créer ce réseau d'entrepôts de données de santé interopérables entre eux, freinant ainsi la pleine exploitation de ces données.

<sup>59</sup> « Data : les 10 nouveaux lauréats de l'AAP pour l'accompagnement à la constitution d'EDS hospitaliers », *Health & Tech*, 18 décembre 2023.

<sup>60</sup> « La première plateforme interrégionale de données massives hospitalières », site internet du *GCS HUGO*, 2 octobre 2023.

<sup>61</sup> « France 2030 : 16 nouveaux programmes d'excellence soutenus pour accélérer la recherche et l'innovation en santé », site internet du *ministère de l'Enseignement supérieur et de la Recherche*.

### 1.3. DES CHAMPS ENCORE PEU COUVERTS ET DES DONNÉES DIFFICILEMENT EXPLOITABLES

- a. Un périmètre insuffisamment couvert :  
la médecine de ville

**La médecine de ville demeure un champ où les bases de données sont peu présentes et faiblement structurées.** Cela s'explique par le caractère fragmenté de l'exercice de la médecine de ville, essentiellement libérale. Si le Sniiram renseigne sur la consommation de soins des patients en collectant l'activité des praticiens à des fins de remboursement, il ne comporte aucune donnée clinique sur les motifs de la consultation remboursée.

**Par ailleurs, les médecins généralistes ou spécialistes exerçant en ville utilisent une grande variété de logiciels médicaux développés par des éditeurs distincts.** Si le Ségur de la santé numérique a permis d'enclencher une dynamique d'harmonisation des outils à la disposition des médecins de ville, en prévoyant des incitations financières pour ceux qui utilisent des logiciels de gestion de cabinets référencés par les pouvoirs publics<sup>62</sup>, le temps d'adaptation est long.

On observe depuis quelques années des premières initiatives pour structurer les données médicales de la médecine de ville. La Cnam dispose d'une base recueillant les activités des communautés professionnelles territoriales de santé (CPTS), l'outil Rézone CPTS<sup>63</sup> et développe des outils de *data visualisation* sur la prévalence des pathologies et sur la pratique des professionnels de santé libéraux. Le réseau Sentinelles<sup>64</sup>, piloté par Santé publique France, l'Inserm et Sorbonne Université, est quant à lui un réseau de recherche et de veille en soins de premier recours qui

<sup>62</sup> Agence du numérique en santé, *Fiche pratique Ségur du numérique, Médecins de ville*.

<sup>63</sup> Site internet de *Rézone CPTS*.

<sup>64</sup> Site internet du *réseau Sentinelles*.



rassemble des grandes bases de données en médecine générale et en pédiatrie. On peut également citer le lancement du premier entrepôt de données de santé de médecine de ville, le projet P4DP, qui a pour objectif d'être mis en relation à terme avec le SNDS<sup>65</sup>. Ce projet est financé dans le cadre de France 2030 à hauteur de 10 millions d'euros sur 3 ans<sup>66</sup>. À partir des données issues de l'entrepôt, un outil de visualisation des données « augmentées » doit être réalisé afin de produire des rapports de veille épidémiologique et d'étudier les parcours de soins en ville des patients en fonction de leurs caractéristiques<sup>67</sup>.

**Des initiatives privées ont également permis d'apporter aux équipes de recherche des informations structurées sur la médecine de ville :** par exemple, la base de données THIN (*The Health Improvement Network*<sup>68</sup>), développée par le groupe CEGEDIM, agrège les dossiers médicaux informatisés et anonymisés de 69 millions de patients recueillis auprès de médecins généralistes et spécialistes dans 7 pays européens.

**En l'absence d'offre complète et structurée pour les données de la médecine de ville, un champ important du parcours de soins est non couvert, alors même que notre système de santé est amené à se transformer avec le virage ambulatoire<sup>69</sup>.** Ces données médicales demeurent pour beaucoup éclatées dans les logiciels des officines et cabinets médicaux, sous un format et une nomenclature non standardisés.

<sup>65</sup> « Lancement de l'entrepôt national de données de médecine de ville », site internet du [Health Data Hub](#), septembre 2022.

<sup>66</sup> « P4DP, un consortium pour créer le premier entrepôt de données de santé pour la médecine générale », site internet du [Health Data Hub](#), mars 2023.

<sup>67</sup> « P4DP : le premier entrepôt national de données de santé de médecine de ville », site internet du [Health Data Hub](#), novembre 2023.

<sup>68</sup> « Cegedim Health Data lance sa base de données de vie réelle Européenne THIN en Italie », [GlobeNewswire](#), juin 2021.

<sup>69</sup> Institut Montaigne, [Soins ambulatoires et à domicile : un virage essentiel](#), février 2023.

## b. Des données externes trop peu exploitées

**Aux données issues directement du système de santé s'ajoute une masse de données externes pouvant éclairer l'état sanitaire de la population.** L'ensemble des informations relatives au patient (sexe, âge), à son environnement (lieu d'habitation, profession, déplacements) ainsi qu'à ses habitudes de vie (pratique d'une activité physique, nutrition) participent à la compréhension d'une pathologie et à la définition d'un diagnostic et d'un parcours de soins pertinents.

**La source de ces données externes est aujourd'hui de plus en plus diversifiée.** Les données générées par le patient, par exemple par le biais des appareils connectés (applications de santé, montres et balances connectées) constituent des informations très utiles pour les travaux de recherche. De la même manière, l'appariement des données de santé avec des données statistiques d'une autre nature (catégories socio-professionnelles, lieu de résidence, données socio-démographiques), notamment collectées par l'Insee, voire des données collectées sur l'environnement, peut contribuer à nourrir les conclusions d'un projet de recherche ou orienter plus finement une campagne de prévention ou de dépistage, dans une approche globale des enjeux de santé publique. À titre d'exemple, le projet *Green Data for Health* (GD4H)<sup>70</sup>, permet d'accéder à un vaste catalogue de données existantes sur l'environnement.

Devant la multiplicité des bases de données, l'Institut Montaigne a réalisé une première cartographie non-exhaustive permettant de recenser les principales bases de données de santé, la nature des données qu'elles contiennent ainsi que l'institution responsable de leur maintenance.

<sup>70</sup> « *Green Data for Health, Common environmental data space for health* », Ministère de la Transition écologique et de la cohésion des Territoires.

Tableau 1 : des données de natures diverses, structurées dans de nombreuses bases

(liste non exhaustive)

Catégories de données	Bases principales / exemples de bases	Responsable de la base	Niveau de gouvernance
<b>Données médico-administratives</b> <i>Données collectées à des fins de remboursement</i>	<ul style="list-style-type: none"> <li>✓ <b>Système national des données de santé</b> <ul style="list-style-type: none"> <li>• Sniiram</li> <li>• PMSI</li> <li>• BCMD</li> <li>• SI-DEP : data vaccin COVID</li> <li>• Données des MDPH</li> </ul> </li> <li>✓ <b>Base de données publiques des médicaments</b></li> <li>✓ <b>Base statistique Annuelle des Établissements de Santé</b></li> </ul>	<ul style="list-style-type: none"> <li>✓ <b>Cnam</b> <ul style="list-style-type: none"> <li>• Cnam</li> <li>• ATIH</li> <li>• Unité CépiDC (Inserm)</li> <li>• CNSA</li> </ul> </li> <li>✓ <b>ANSM, HAS, UNCAM</b></li> <li>✓ <b>Drees</b></li> </ul>	<ul style="list-style-type: none"> <li>✓ <b>National</b></li> </ul>
<b>Données cliniques</b> <i>Données collectées à des fins de recherche</i>	<ul style="list-style-type: none"> <li>✓ <b>Bases sur des pathologies précises</b> <ul style="list-style-type: none"> <li>• ESMÉ (cancer)</li> <li>• Orphanet (maladies rares)</li> <li>• Hepather (diabète)</li> </ul> </li> <li>✓ <b>Bases issues des cohortes nationales et enquêtes épidémiologiques</b> <ul style="list-style-type: none"> <li>• Cohorte Constances</li> <li>• Base Epicov</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>✓ <b>Instituts de recherche</b> <ul style="list-style-type: none"> <li>• Unicancer</li> <li>• Inserm</li> </ul> </li> <li>✓ <b>Instituts de recherche</b> <ul style="list-style-type: none"> <li>• Inserm</li> <li>• Inserm et Drees</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>✓ <b>National / local</b></li> </ul>
<b>Données cliniques</b> <i>Données collectées pour le soin</i>	<ul style="list-style-type: none"> <li>✓ <b>Entrepôts de données de santé hospitaliers</b></li> <li>✓ <b>Base sur la médecine de ville</b> <ul style="list-style-type: none"> <li>• Des initiatives éparées, en cours de structuration</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>✓ <b>CHU</b></li> <li>✓ <b>NA</b></li> </ul>	<ul style="list-style-type: none"> <li>✓ <b>Local</b></li> <li>✓ <b>NA</b></li> </ul>
<b>Données externes</b>	<ul style="list-style-type: none"> <li>✓ <b>Données environnementales, socio-économiques</b> <ul style="list-style-type: none"> <li>• Des initiatives éparées, en cours de structuration</li> </ul> </li> <li>✓ <b>Données générées par le patient via des applications, télésurveillance</b> <ul style="list-style-type: none"> <li>• Des initiatives éparées, en cours de structuration</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>✓ <b>NA</b></li> </ul>	<ul style="list-style-type: none"> <li>✓ <b>NA</b></li> </ul>

#### 1.4 DES ACTIONS NATIONALES ET EUROPÉENNES AYANT CONTRIBUÉ À LA CONSTRUCTION D'UN CADRE SUR L'UTILISATION SECONDAIRE DE LA DONNÉE DE SANTÉ

a. Une stratégie française en cours de construction

**En France, la prise de conscience du potentiel que constituent les données de santé, et l'intérêt de leur exploitation, a été précoce.** Dès 2016, la loi de modernisation de notre système de santé<sup>71</sup> a initié un premier mouvement d'agrégation des données au niveau national avec le SNDS. Cette loi a également consolidé les bases d'une gouvernance centralisée en créant l'Institut national des données de santé (INDS) et en renforçant ses prérogatives. La loi relative à l'organisation et à la transformation du système de santé (OTSS) de 2019<sup>72</sup> a confirmé cette structuration autour d'un offreur national de données de santé, à la suite du rapport Villani de 2018 qui appelait la France à devenir un leader européen en matière de données de santé. La Plateforme des données de santé, généralement appelée *Health Data Hub* (HDH), a ainsi été créée, en substitution à l'INDS, pour « garantir un accès aisé et unifié, transparent et sécurisé, aux données de santé afin d'améliorer la qualité des soins et l'accompagnement des patients<sup>73</sup> ». **Cette plateforme a un rôle d'accompagnement auprès des porteurs de projets déposant une demande d'accès aux données du SNDS et à d'autres bases de données. Le Health Data Hub est voué à prendre un rôle de guichet unique, en réponse à la complexité de l'écosystème des données de santé.** Il est amené également à traiter les demandes d'exploitation des données de son catalogue. En 2022, 97 projets et

<sup>71</sup> Loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé.

<sup>72</sup> « Le HDH, au service de la recherche sur les données de santé depuis trois ans », site internet du *Health Data Hub*, décembre 2022.

<sup>73</sup> *Ibid.*

192 extractions de données ont été accompagnés par le *Health Data Hub*<sup>74</sup>.

Une dynamique similaire de création de guichet unique a pu être observée dans d'autres pays européens à l'image de la Finlande, qui a mis en place la plateforme unique *FinData* en 2019. *FinData* dispose de bases de données provenant de multiples détenteurs de données nationales tels que les entités publiques du secteur social et de la santé (hôpitaux publics notamment), de *Statistics Finland*, de l'Institut d'assurance sociale (*Kela*), ou encore des données de citoyens inscrits dans le dossier électronique national *Kanta*<sup>75</sup>. Le Royaume-Uni a quant à lui mis en place un *Health Data Research Innovation Gateway*<sup>76</sup>, point d'entrée unique pour les chercheurs qui permet de découvrir et de demander l'accès à l'ensemble des données de santé du Royaume-Uni.

**Dans la continuité de la première feuille de route de 2019 portée par la Délégation du numérique en santé (DNS)<sup>77</sup>, la feuille de route du numérique en santé 2023-2027<sup>78</sup> a été publiée en mai 2023.** Elle affiche explicitement l'ambition de développer l'utilisation des données de santé en inscrivant comme dix-huitième et dernière action « Développer la recherche en santé numérique et en particulier l'utilisation secondaire des données de santé ». **Un des leviers prioritaires inscrits dans cet objectif est la publication pour le premier semestre 2024 d'une stratégie nationale sur la réutilisation et l'ouverture des données de santé.**

<sup>74</sup> Commission des affaires sociales, Sénat, « Données de santé : une réforme encore en cours de chargement », Rapport d'information, juillet 2023.

<sup>75</sup> Site internet de *MyKanta*.

<sup>76</sup> Catalogue du *Health Data Research Innovation Gateway*.

<sup>77</sup> Feuille de route, « Accélérer le virage numérique », 2019.

<sup>78</sup> Feuille de route du numérique en santé 2023-2027, « Mettre le numérique au service de la santé ».

**Afin d'alimenter la stratégie nationale sur la réutilisation et l'ouverture des données de santé, une mission spécifique de préfiguration d'une feuille de route dédiée aux données de santé** a été confiée par le ministre de la Santé et de la Prévention à Jérôme Marchand-Arvier en date du 30 mai 2023 afin de « jeter les bases pour la construction d'une nouvelle feuille de route pour l'utilisation secondaire des données de santé<sup>79</sup> ». Les conclusions de cette mission ont été rendues publiques en janvier 2024.

**La mission<sup>80</sup> souligne notamment le rôle du *Health Data Hub* qui a su s'imposer comme un partenaire clé de l'écosystème des données de santé mais estime que le *Health Data Hub* demeure freiné dans le plein accomplissement de ses missions.** En cause, le choix autour de la solution technique de sa plateforme d'hébergement, qui doit notamment accueillir une copie des données du SNDS pour faciliter son accès aux porteurs de projet. Le *Health Data Hub* a choisi la solution technique *Microsoft Azure*, qui offrait selon lui, au regard des produits du marché européen, les capacités techniques requises (capacités de calcul, stockage, croisement...) pour héberger la copie du SNDS. Tandis que ses précédents avis soulignaient le risque de transfert des données personnelles vers les États-Unis, la CNIL a finalement autorisé *Microsoft* à être l'hébergeur du *Health Data Hub* pour une durée de trois ans dans le cadre du projet EMC2, dans une décision publiée le 31 janvier dernier<sup>81</sup>. Elle a toutefois exprimé des regrets quant au fait qu'aucun des prestataires envisagés ne puisse garantir la protection des données contre l'application des lois de pays tiers.

<sup>79</sup> « Données de santé : la mission de Jérôme Marchand-Arvier dévoilée », *Mind Health*, juin 2023.

<sup>80</sup> Jérôme Marchand-Arvier, Pr Stéphanie Allassonnière, Aymeril Hoang, Dr Anne-Sophie Jannot, « Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire des données de santé », décembre 2023.

<sup>81</sup> Délibération 2023-146 du 21 décembre 2023 autorisant le groupement d'intérêt public « Plateforme des données de santé » à mettre en œuvre un traitement automatisé de données à caractère personnel ayant pour finalité la constitution d'un entrepôt de données dans le domaine de la santé, dénommé « EMC2 ».

Le choix de cette solution d'hébergement non-souveraine cristallise des tensions de l'écosystème autour du *Health Data Hub*. La mission interministérielle recommandait notamment de programmer l'arrêt de l'hébergement sur *Microsoft Azure* et d'identifier une solution de type cloud *SecNumCloud* à horizon de 24 mois.

**b. Une stratégie européenne aux contours encore peu précis**

**Ces initiatives nationales se structurent autour d'un cadre européen qui, en matière de données de santé, est en construction.** Le cadre juridique européen, et notamment l'entrée en application du Règlement général sur la protection des données (RGPD) du 25 mai 2018<sup>82</sup>, a posé un cadre général à la réglementation autour des données de santé. Toute demande d'exploitation de données de santé doit se conformer à cette législation. **Par ailleurs, un espace européen des données de santé est en cours de construction<sup>83</sup>, l'European Health Data Space (EHDS), le règlement ayant été adopté le 13 décembre 2023 par le Parlement européen.** Ce projet concerne à la fois l'usage primaire et secondaire des données de santé. Dans une optique d'utilisation primaire, ce projet doit permettre le partage des données de santé dans le cadre du parcours de soins *via* la connexion des contacts nationaux des pays membres pour l'e-santé (*National Contact Point for eHealth*) et d'assurer le partage transfrontalier des synthèses médicales (*Patient Summary*) et ePrescriptions<sup>84</sup>. Le second volet du projet concerne l'utilisation secondaire des données de santé des citoyens européens. **Il vise à harmoniser les normes, les pratiques, les infrastructures pour l'utilisation des données de santé.** Il s'agira par exemple de donner un accès unique aux bases de données pertinentes en Europe : ceci pourrait représenter

<sup>82</sup> « Le règlement général sur la protection des données (RGPD), mode d'emploi », Ministère de l'Économie des Finances et de la Souveraineté industrielle et numérique, avril 2023.

<sup>83</sup> « Espace européen des données de santé », site internet de la [Commission européenne](#).

<sup>84</sup> Isabelle Zablit-Schmitz, « [Europe de la Santé, quelle place pour le numérique ?](#) », Institut Montaigne, avril 2022.

une simplification majeure car aujourd'hui, les autorisations d'accès sont faites dans chacun des États membres, selon des procédures qui leur sont propres.

Le modèle français sert d'exemple à l'élaboration du futur règlement européen sur l'utilisation des données de santé. Le *Health Data Hub*<sup>85</sup> est à la tête du consortium choisi par la Commission européenne chargé de mettre en place le projet pilote de l'Espace européen des données de santé. Ce consortium réunit seize partenaires, issus d'une dizaine de pays européens. Les dispositions de ce règlement encadreront des éléments structurants de la mise à disposition et de l'exploitation des données de santé, comme la formulation du consentement du patient à la collecte des données. **Son adoption conduira nécessairement la France à adapter son modèle de mise à disposition des données de santé.**

**L'articulation entre le contrôle de ce nouvel espace européen des données de santé et le contrôle exercé par les régulateurs nationaux, comme la CNIL en France, pose encore question.** Un avis du Comité européen de la protection des données et du Contrôleur européen de la protection des données de juillet 2022<sup>86</sup> préconise de confier la compétence exclusive aux autorités de protection des données pour toute question relative à la protection des données de santé personnelles. Il alerte également sur la nécessité d'assurer une cohérence entre le futur règlement mettant en œuvre l'EHDS et le RGPD, voire, pour notre droit national, la loi « Informatique et Libertés » de 1978. Pour le modèle français, il s'agit également de préserver le rôle de la CNIL dans ses prérogatives d'autorisation des projets relatifs à l'utilisation des données de santé, alors qu'un modèle fondé sur une autorisation octroyée par le propriétaire des données de santé est en discussion dans le cadre des travaux sur le

<sup>85</sup> « Lancement d'un projet pilote pour l'Espace européen des données de santé », site internet du [Health Data Hub](#), juillet 2022.

<sup>86</sup> European Data Protection Board, *Avis conjoint EDPB-CEPD 03/2022 sur la proposition de règlement relatif à l'espace européen des données de santé*, adopté le 12 juillet 2022.

règlement européen. **Enfin, dans leur avis, les régulateurs nationaux attirent l'attention sur la nécessité de retenir des solutions d'hébergement des données de santé souveraines et localisées sur le territoire de l'Union européenne, au regard de la sensibilité des données et de leur volume (500 millions de citoyens européens).** Plus largement, la régulation des données de santé à l'échelle européenne s'inscrit dans le contexte tendu de la régulation liée aux nouvelles technologies. La récente adoption de l'*AI Act*<sup>87</sup> par les 27 États membres de l'Union européenne est un exemple symptomatique. Sa validation s'est faite au terme de négociations serrées entre les pays portant une approche souple destinée à favoriser l'essor des innovations (c'est notamment le cas de la France), et ceux possédant une vision régulatrice plus stricte ayant vocation à encadrer le marché de l'IA et ses usages, qui ont finalement fait prédominer leur vision. Les prérogatives de l'*AI Act* devront également s'articuler avec des articles de l'EHDS prévoyant un encadrement spécifique des données de santé par l'intelligence artificielle. La licéité de la réutilisation des données de santé dans le cadre d'algorithmes d'intelligence artificielle est conditionnée à des finalités de santé publique ou de qualité et de sécurité des soins<sup>88</sup>.

## 2 Les cas d'usage de la donnée de santé à des fins d'utilisation secondaire

**L'utilisation secondaire regroupe l'ensemble des cas d'usage ayant d'autres finalités que les soins du patient concerné.** Dans ce domaine, les utilisations des données de santé sont vastes et sources de progrès incommensurables pour la recherche médicale, la prévention et la qualité des soins, mais aussi dans le domaine de la gestion du système de santé et son efficacité. De nombreux bénéfices liés à l'utilisation secondaire des données de santé peuvent être identifiés, comme la **surveillance épidémiologique, l'amélioration des pratiques et de l'accès aux services de santé, l'évaluation des innovations, la recherche et l'amélioration des connaissances ainsi que le pilotage de la politique de santé**<sup>89</sup>.

**Les usages de la donnée de santé sont quasi infinis mais demeurent encore peu connus du grand public, alors même que les bénéficiaires directs d'un meilleur usage de la donnée sont les patients et les professionnels de santé.**

Un rapport du Healthcare Data Institute de 2023<sup>90</sup> identifie trois catégories d'usage à l'utilisation secondaire des données de santé :

- L'utilisation à des  **fins de recherche**  pour étudier la pertinence d'un parcours de soins d'une population spécifique ou encore pour identifier des nouvelles stratégies thérapeutiques pour les patients.
- L'utilisation pour le **développement d'innovations technologiques** tels que les algorithmes d'intelligence artificielle (IA), capables de **transformer les pratiques et l'organisation des soins**. Le développement de tels algorithmes nécessite de mobiliser un

<sup>87</sup> « *AI Act : quelles suites après la ratification par les membres de l'UE ?* », *Health & Tech*, 6 février 2024.

<sup>88</sup> *European Data Protection Board, Avis conjoint EDPB-CEPD 03/2022 sur la proposition de règlement relatif à l'espace européen des données de santé, adopté le 12 juillet 2022.*

<sup>89</sup> *Conseil scientifique consultatif du Health Data Hub, Bénéfices et risques de l'utilisation des données de santé à des fins de recherche, Position paper, janvier 2024.*

<sup>90</sup> *Healthcare Data Institute, « Génération et exploitation des données hospitalières à visée de recherche en santé publique et en épidémiologie », 2023.*

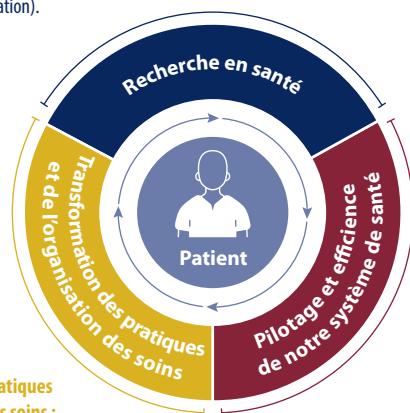
grand nombre de données de santé pour assurer leur pertinence. Ces innovations peuvent aider à la décision pour les professionnels de santé et à une meilleure précision dans le diagnostic. Ils peuvent également prédire les risques d'apparition d'un cancer ou les récives d'une maladie.

- L'utilisation pour **le pilotage de notre système de santé**, à l'image de la gestion de l'épidémie de Covid-19 où la donnée a pu jouer un rôle essentiel dans le suivi des cas contacts ou encore dans celui de la campagne de vaccination. Il existe d'autres modèles de pilotage par la donnée, qui permettent d'identifier les besoins de santé d'une population sur un territoire et d'adapter en fonction les politiques de santé et les parcours de soins territoriaux<sup>91</sup>.

Graphique 1 : des cas d'usage de la donnée de santé

#### Recherche en santé :

- Réalisation des essais cliniques (mobilisation de données en vie réelle, bras de contrôle synthétiques);
- Analyse de la pertinence des parcours de soin (prévision des risques de rechute, réhospitalisation).



#### Transformation des pratiques et de l'organisation des soins :

- Meilleur diagnostic et dépistage des pathologies;
- Développement d'outils de prévention;
- Gestion des établissements de santé et du système de santé.

#### Pilotage et efficacité de notre système de santé :

- Détection des fraudes et analyse des dépenses de santé;
- Mesure de la qualité des soins;
- Évaluation des politiques publiques;
- Suivi de la gestion des épidémies.

Dans le but d'apporter des exemples concrets à des cas d'usage d'utilisation secondaire des données de santé, l'Institut Montaigne a mené plus d'une cinquantaine d'auditions auprès d'acteurs représentatifs de l'écosystème : pouvoirs publics, associations de patients, institutionnels, industriels, universitaires, laboratoires pharmaceutiques. **Nous avons ainsi pu recueillir des exemples inédits témoignant de l'immense potentiel que représentent les données de santé.**

### 2.1. L'UTILISATION DES DONNÉES POUR OPTIMISER LE PILOTAGE DE NOTRE SYSTÈME DE SANTÉ

- Un outil participant à l'efficacité du pilotage du système de soins

**Historiquement, les données de santé en France ont été utilisées pour réguler le système de soins.** Les données du SNDS servent à déterminer le financement des acteurs du système de santé. Le développement de la tarification à l'activité dans le financement des établissements de santé depuis les années 2000 a permis de multiplier le nombre de données récoltées *via* ces établissements. **À l'heure où notre système de soins recherche un fonctionnement à la fois plus efficace mais aussi davantage axé sur la qualité des soins, leur utilisation peut aujourd'hui trouver d'autres déclinaisons pour améliorer la gestion du système de soins.**

**Tout d'abord, les données de santé sont essentielles pour détecter la fraude.** À titre d'exemple, le système de contrôle de l'Assurance maladie (s'appuyant sur les données du SNDS) a permis d'économiser 150 millions d'euros de fraudes au premier semestre 2023<sup>92</sup>.

<sup>92</sup> « Lutte contre les fraudes : résultats à mi-année », site internet de l'Assurance maladie, octobre 2023.

<sup>91</sup> Institut Montaigne, « Santé : faire le pari des territoires », mai 2022.

**Si les données de santé peuvent être utilisées pour piloter les acteurs, elles sont également essentielles pour mesurer la qualité des soins.** Les données de santé peuvent contribuer à construire des indicateurs de qualité des soins (taux de mortalité, taux de réhospitalisation...) qui peuvent évaluer l'activité des établissements ou des professionnels. Ces indicateurs peuvent mesurer l'expérience du patient (soit la manière dont le patient a vécu une expérience de soins), les processus du parcours de soins ou encore les résultats d'une intervention sur la qualité de vie du patient<sup>93</sup>. Au Royaume-Uni, le recueil des résultats rapportés par les patients en termes de qualité de vie est obligatoire pour quatre pathologies cibles : la chirurgie du genou, la chirurgie de la hanche, la chirurgie des varices et la hernie<sup>94</sup>. Bien que cette pratique demeure encore limitée en France, des établissements de santé initient la production d'indicateurs d'évaluation de soins : c'est le cas du réseau Unicancer<sup>95</sup> qui utilise ces données pour créer des indicateurs à destination de leurs professionnels de santé. Ces initiatives s'inscrivent dans l'émergence du concept de valeur en santé (ou *Value-Based Healthcare*) qui propose un système de santé fondé sur la valeur. La valeur est mesurée à partir des résultats qui importent aux patients et du coût pour atteindre ces résultats<sup>96</sup>.

**b. Un élément essentiel des politiques de santé publique**

**Les bases de données de santé peuvent être utilisées à des fins de pharmacovigilance, pour évaluer les bénéfices mais aussi les risques d'un médicament ou d'un traitement.** Les institutions sanitaires (l'Agence nationale de sécurité et du médicament – ANSM notamment) comme les laboratoires pharmaceutiques et les industriels ont très tôt

<sup>93</sup> Institut Montaigne, « *Système de santé : soyez consultés!* », 2019.

<sup>94</sup> Laure Millet, « *Indicateurs de qualité des soins : quels modèles pour la France ?* », Institut Montaigne, 2019.

<sup>95</sup> « *Démarche qualité des Centres de lutte contre le cancer* », site internet de [Unicancer](#).

<sup>96</sup> Institut Montaigne, « *Système de santé : soyez consultés!* », 2019.

développé des systèmes recueillant et exploitant les données de santé à cette fin. Les industriels et laboratoires pharmaceutiques s'appuient sur les données de remboursement et les données cliniques issues de leurs solutions (croisées avec des bases hospitalières, nationales ou issues des études cliniques) pour mesurer les bénéfices de leurs produits.

**La crise sanitaire a également montré combien l'exploitation des données de santé est un élément essentiel de santé publique,** notamment dans le cadre de gestion d'une pandémie. Durant la crise du Covid-19, les données de dépistage, agrégées dans SI-DEP, et de vaccination, collectées dans Vaccin Covid<sup>97</sup>, ont été essentielles pour suivre la circulation du virus, prendre les décisions sanitaires de gestion de crise (diminution ou interdiction des déplacements...) et déterminer les priorités de la campagne vaccinale. Les données de santé sont aussi utilisées tous les ans pour gérer l'épidémie de grippe (date de lancement de la vaccination, adaptation du vaccin...). Ce travail effectué par Santé publique France s'appuie sur des données de surveillance issues de différents réseaux (réseau Sentinelles, base OSCOUR, données des agences régionales de santé, etc.)<sup>98</sup>.

**Enfin, les données de santé offrent des perspectives intéressantes d'amélioration des politiques de santé, via une territorialisation plus affirmée et une adaptation de l'offre de soins aux bassins de populations et à leurs besoins<sup>99</sup>.** L'utilisation des données de santé permettrait une meilleure analyse des besoins de santé d'une population et une meilleure adaptation de l'offre de soins<sup>100</sup>. Aux États-Unis, les établissements de santé participent à des HIE (*Health Information Exchange*), qui sont des réseaux d'échange électronique d'informations de santé, utilisant ces données pour suivre l'état sanitaire de la population dans un territoire

<sup>97</sup> « *Vaccin Covid : un outil numérique pour le suivi de la vaccination* », site internet de [l'Assurance maladie](#), juillet 2023.

<sup>98</sup> « *Fardeau de la grippe en France métropolitaine, bilan des données de surveillance lors des épidémies 2011-12 à 2021-22* », site internet de [Santé publique France](#).

<sup>99</sup> Institut Montaigne, « *Santé : faire le pari des territoires* », mai 2022.

<sup>100</sup> *Ibid.*

donné. À titre d'exemple, le *Delaware Health Information Network* (DHIN)<sup>101</sup> traite plus de 14 millions de documents médicaux chaque année. **La gestion optimisée de ces données de santé permet d'améliorer la coopération entre les professionnels de santé, de créer des parcours de soins plus personnalisés et d'adapter les politiques de santé selon les besoins de la population de l'État.**

## 2.2. L'UTILISATION DES DONNÉES DE SANTÉ POUR TRANSFORMER LES PRATIQUES ET L'ORGANISATION DES SOINS

a. Un outil essentiel pour le diagnostic, les dépistages précoces et la prévention

**Grâce aux données de santé, de nombreuses innovations voient le jour chaque année et représentent des aides précieuses dans les parcours de soins, en permettant d'affiner le diagnostic des professionnels de santé, de réaliser des dépistages précoces, et d'anticiper des récidives**<sup>102</sup>. La digitalisation des dossiers médicaux et l'émergence des outils associés (téléconsultation, e-prescription, *chatbots*) assurent le recueil d'un grand nombre de données représentant un potentiel immense pour la prévention<sup>103</sup>. À titre d'exemple, des chercheurs<sup>104</sup> américains ont récemment mis au point une IA prédisant le risque de développer un cancer du sein cinq ans avant l'apparition des symptômes. Une entreprise française, Therapixel<sup>105</sup>, a également développé un assistant

<sup>101</sup> Site internet du *Delaware Health Information Network*.

<sup>102</sup> À titre d'exemple, le projet DAICAP porté par l'AP-HP et soutenu par le Health Data Hub développe un algorithme d'IA d'aide au diagnostic du cancer de la prostate afin d'améliorer la qualité de l'interprétation de l'IRM.

<sup>103</sup> Laure Millet, « *La prévention en santé, les nouveaux impératifs* », septembre 2023.

<sup>104</sup> Arasu et al., "Comparison of Mammography AI Algorithms with a Clinical Risk Model for 5-year Breast Cancer Risk Prediction: An Observational Study", *Radiology*, juin 2023.

<sup>105</sup> Site internet de *Therapixel*.

numérique au radiologue pour analyser les mammographies et détecter les cancers du sein à un stade précoce. Pour élaborer son algorithme, l'entreprise s'est appuyée sur les données de plus de 640 000 images numérisées de mammographie provenant de 86 000 patientes<sup>106</sup>. Une étude clinique lancée aux États-Unis a ainsi démontré que cette solution permettait aux radiologues de détecter 20 % de cancers de plus que s'ils étaient seuls<sup>107</sup>.

Dans une optique similaire, des **outils de prévention, plus personnalisés et ciblés se développent**, à l'image de *chatbots*, pouvant agir comme des compagnons de santé. Les patients peuvent ainsi réaliser de manière autonome leur bilan de santé et le cas échéant, être orienté vers le professionnel de santé adéquat. L'outil *Mon Bilan Cardio*<sup>108</sup>, permet à partir des données complétées par le patient sur un espace sécurisé, de détecter et d'analyser les facteurs de risques cardiovasculaires. **Des algorithmes d'évaluation des risques identifient les comportements susceptibles d'impacter la santé du patient** : alimentation, activité physique, sommeil. L'outil propose au patient une restitution personnalisée des résultats ainsi que des solutions d'orientation vers le professionnel de santé adéquat.

b. Des innovations technologiques pour l'aide à la gestion des établissements et de notre système de santé

**Les données de santé appliquées peuvent être également exploitées pour améliorer l'efficacité du fonctionnement des établissements de santé.** Des logiciels de *machine learning* se développent pour automatiser et optimiser le codage des séjours hospitaliers. Les hôpitaux renseignent

<sup>106</sup> *ibid.*

<sup>107</sup> « *MammoScreen, l'intelligence artificielle qui détecte les cancers du sein* », Bigmédia bpiFrance, octobre 2022.

<sup>108</sup> « *Malakoff Humanis : Mon Bilan Cardio ou une nouvelle approche de la prévention santé* », Assurance & Banque, 2022.



régulièrement leur activité à travers un ensemble de codes transmis à l'ATIH (qui gère le PMSI) : de cette déclaration dépend la rémunération des établissements de santé. **Les logiciels développés permettent de prédire les diagnostics issus des Dossiers Patients Informatisés (DPI) et de les transformer en codes qui seront facturés à l'Assurance maladie**<sup>109</sup>. Cette action permet d'optimiser le temps pris par des professionnels à coder leurs actes et de le réallouer *in fine* sur du temps médical. Des nouvelles technologies de pointe continuent également à se développer afin de faciliter le traitement des informations contenues au sein des dossiers médicaux à l'image des *Large Language Models* (LLM), des outils générés par l'IA qui permettent de comprendre et générer du langage naturel. Dans le cadre du traitement de données, ils peuvent être utilisés afin de structurer, traiter, synthétiser ou analyser les données issues de dossiers médicaux.

**Ces innovations technologiques peuvent également être appliquées à des bases nationales puis locales dans le cadre d'un pilotage des politiques de santé par la donnée.** À ce titre, le modèle de la responsabilité populationnelle expérimentée par la Fédération hospitalière de France représente une application prometteuse<sup>110</sup>. Ce modèle, développé au Québec<sup>111</sup>, repose sur un ensemble d'acteurs à l'échelle du territoire, allant au-delà des acteurs de santé traditionnels, qui s'engagent à améliorer la santé de leur population en travaillant à des parcours de soins adaptés à ses besoins de santé. Afin d'identifier les besoins de santé d'un territoire, la Fédération hospitalière de France a développé un algorithme sur le PMSI national qui permet de classer les populations en strates de besoins. Ces algorithmes sont ensuite appliqués au niveau local pour définir la stratification des besoins de santé de la population. **Ce pilotage au niveau local par la donnée permet d'optimiser les**

<sup>109</sup> On peut notamment citer les solutions développées par l'entreprise française *Sancare*.

<sup>110</sup> « La FHF en action. Responsabilité populationnelle : tous acteurs de notre santé ! », site internet de la *Fédération hospitalière de France*, avril 2022.

<sup>111</sup> Institut Montaigne, « Santé : faire le pari des territoires », mai 2022.

**ressources d'un territoire et d'adapter l'offre de santé aux besoins de santé locaux**<sup>112</sup>.

### 2.3. L'UTILISATION DES DONNÉES DE SANTÉ POUR LA RECHERCHE

#### a. Une source d'avancées significatives pour la recherche médicale

**L'utilisation secondaire des données de santé offre une nouvelle dimension à la recherche médicale.** En particulier, elle permet de raccourcir de manière significative la durée des essais cliniques en mettant à disposition des chercheurs des cohortes de patients, ou des cohortes de contrôle, dans des délais rapides et de les rendre plus fiables grâce à l'utilisation de données massives et de qualité. À titre d'exemple, la richesse de la plateforme *Medidata*<sup>113</sup>, qui rassemble les données collectées à des fins de recherche et issues de 30 000 essais cliniques<sup>114</sup> menés par différents laboratoires pharmaceutiques dans plusieurs pays, a permis en 2020 d'accélérer significativement l'essai clinique du vaccin Moderna. La plateforme contribue aussi significativement au développement de l'oncologie de précision, qui a facilité l'autorisation historique de la commercialisation du médicament Keytruda par la FDA, le premier médicament permettant de traiter les cancers via leur origine génétique<sup>115</sup>.

**Les données de santé permettent ainsi la mise en place de « bras de contrôle synthétiques », c'est-à-dire d'une cohorte de patients virtuels présentant l'ensemble des propriétés des patients réels.**

<sup>112</sup> Fédération hospitalière de France, « Feuille de route : pour un système d'information et d'organisation populationnel », 2022.

<sup>113</sup> *Medidata*, site internet de Dassault Systèmes.

<sup>114</sup> « *Medidata favorise la diversité dans les essais cliniques avec plus de 30 000 études et 9 millions de participants* », *Silicon.fr*.

<sup>115</sup> *Medidata*, *White paper: Precision Oncology Trials: A look Ahead*, 2018.

Cette cohorte synthétique peut présenter les caractéristiques d'un seul organe ou de toute partie du corps humain, comme le cœur. Outre le gain de temps, ce dispositif permet de réduire l'effet placebo détectable dans le cadre d'un essai clinique en vie réelle. Paris-Saclay Cancer Cluster<sup>116</sup> a réussi à structurer les données en cancérologie de plusieurs établissements hospitaliers (AP-HP, Institut Gustave Roussy) et à identifier les résistances à certains traitements chez les patients. **Ces données permettent aux chercheurs de constituer rapidement des cohortes de patients pertinentes pour leurs projets de recherche.**

Par ailleurs, dans le cadre du développement d'essais cliniques pour les maladies rares, les chercheurs peuvent être confrontés à des problèmes de recrutement de patients candidats, au regard de la taille de la population concernée. **À ce titre la mobilisation de données de vie réelle permet une approche différente de l'approche classique des essais cliniques et de constituer plus rapidement une cohorte de patients témoins sur des bases de données nationales.** Le projet *DeepSarc*, porté par le Centre Léon Bérard propose d'identifier des approches thérapeutiques ciblées pour les patients atteints de sarcomes (tumeurs rares) en croisant les données issues de la base Sarcome avec les données du SNDS<sup>117</sup>. Cette mobilisation des données du SNDS offre l'opportunité aux équipes de travailler sur un plus grand nombre de données par rapport aux essais cliniques traditionnels.

- b. Une contribution indéniable à l'amélioration de la qualité des soins

**L'exploitation secondaire des données de santé peut permettre de mieux définir la pertinence du parcours de soins et d'améliorer les prescriptions de traitement ou de protocole thérapeutiques, en fonction des caractéristiques de chaque patient.** Grâce à l'analyse des

<sup>116</sup> Site internet du [Paris-Saclay Cancer Cluster](#).

<sup>117</sup> « *DeepSarc* », site internet du [Health Data Hub](#).

données médico-administratives du SNDS, ainsi que celles des établissements hospitaliers, il est possible d'observer les éventuelles disparités de prises en charge proposées aux patients et d'élaborer des prévisions des risques de réhospitalisation.

Le suivi et la surveillance du parcours de soins peuvent également être significativement améliorés grâce à l'utilisation des données de santé récoltées à travers des outils de télésurveillance médicale. Des entreprises ont notamment conçu des solutions de télésurveillance qui, à partir des informations récoltées auprès du patient (observance des traitements, suivi du rythme cardiaque, suivi du taux d'insuline), sont en capacité **de prédire les éventuelles ruptures dans la prise en charge ou une potentielle dégradation du patient et peuvent alerter les équipes soignantes en amont.** La solution de télésurveillance *NephroWise*, grâce à l'analyse des constantes du patient (poids, tension) ou encore de ses résultats de biologie, identifie les patients les plus à risque de réhospitalisation et de rejet du greffon<sup>118</sup>.

**Les données de santé peuvent aussi améliorer la personnalisation des traitements et des parcours de soins.** L'exploitation des données permet de développer des dispositifs médicaux de plus en plus adaptés au patient, notamment en développant des modèles capables de reproduire les mécanismes biochimiques d'une partie du corps et ainsi simuler le comportement et la réaction d'un organe où sera implanté le dispositif médical.

<sup>118</sup> *NephroWise*, site internet de [Sêmeia](#).

## Des jumeaux numériques du corps humain créés à partir de la donnée

À partir des données de santé de *Medidata Solutions*<sup>119</sup>, éditeur de logiciels de santé américain acquis en 2019, Dassault Systèmes travaille au développement de « jumeaux numériques » du corps humain. **Le projet consiste à répliquer virtuellement, à partir de données de vie réelle, toutes les caractéristiques du corps humain.** Ce « jumeau virtuel » doit permettre de mieux comprendre le fonctionnement du corps humain, d'anticiper ses réactions face à diverses procédures et de développer des dispositifs médicaux personnalisés et adaptés à chaque patient. Cela sera rendu possible par la combinaison des informations médicales propres à un patient avec une grande quantité de données liées au fonctionnement du corps humain<sup>120</sup>.

À ce jour, Dassault Systèmes a déjà développé un modèle virtuel 3D du cœur humain, fruit du projet « *Living Heart* » (2014)<sup>121</sup>, et du cerveau humain, fruit du projet « *Living Brain* »<sup>122</sup> pour la recherche sur l'épilepsie, dans le cadre du *Human Brain Project*<sup>123</sup> de l'Union européenne (2013-2023). Dassault Systèmes prévoit d'étendre cette technologie à d'autres organes tels que le foie et les poumons, avant de créer un jumeau numérique du corps humain dans son ensemble<sup>124</sup>.

Outre l'intérêt présenté par les jumeaux numériques du point de vue de **l'amélioration du soin et du progrès scientifique, cette technologie est également susceptible de présenter des avantages financiers.** Elle peut réduire la longueur et le coût des recherches médicales et pharmaceutiques. Des essais cliniques virtuels sont en cours aux États-Unis avec la *FDA* pour tester des dispositifs médicaux, réduisant ainsi la nécessité d'un grand nombre de patients pour les essais cliniques finaux<sup>125</sup>.

**Pour finir, de nouvelles catégories de données à l'image des données synthétiques présentent la possibilité de bouleverser les protocoles de recherche ainsi que les démarches juridiques associées.** Les données synthétiques<sup>126</sup> sont générées de manière artificielle (par opposition aux données collectées dans le monde réel), grâce à des méthodes d'IA. **Ces dernières sont ainsi comparables à des acteurs, imitant le comportement et les caractéristiques des données réelles, sans pour autant renvoyer à des informations personnelles identifiables ou vérifiables.** Ainsi, ces données ne sont pas soumises aux mêmes règles que le RGPD et sont particulièrement intéressantes du point de vue de la protection de la vie privée<sup>127</sup>.

<sup>119</sup> *Medidata, Site internet de Dassault Systèmes.*

<sup>120</sup> *Dassault Systèmes (2023, août 31). Virtual twin experiences in life sciences & healthcare.*

<sup>121</sup> *"The Living Heart Project", site internet de Dassault Systèmes.*

<sup>122</sup> *"Living Brain for Targeted Epilepsy Surgery", site internet de Dassault Systèmes.*

<sup>123</sup> *"The Human Brain Project ends: What has been achieved", Humain Brain Project, septembre 2023.*

<sup>124</sup> *Compiegne, J. (2023, 28 février). Dassault Systèmes crée la surprise au CES avec ses jumeaux numériques du corps humain. French Healthcare.*

<sup>125</sup> *ibid.*

<sup>126</sup> *Gourraud, P. (2023, 22 juin). Opinion | Données synthétiques : la clé pour protéger la vie privée. Les Echos.*

<sup>127</sup> *EHR-SaFE: Generating high-fidelity and privacy-preserving synthetic electronic health records. (2022, 21 décembre).*

### 3 Des défis qui subsistent freinant les usages de la donnée secondaire en France

**L'exploitation des données de santé a été identifiée par les pouvoirs publics comme un vecteur puissant de progrès en matière de recherche médicale et d'efficacité de l'organisation du système de santé.** Les initiatives nationales des dernières années ont considérablement renforcé la position de notre pays sur la gouvernance de ses données de santé. Malgré cette volonté politique, force est de constater que la France prend du retard dans le déploiement et la mise en œuvre de ces initiatives, avec un positionnement en 14<sup>e</sup> place des pays de l'OCDE pour ce qui relève de la maturité des usages<sup>128</sup>, de la possibilité d'accéder aux ensembles de données et de la capacité à les coupler dans le secteur de la santé.

**Le retard pris impose à la France de relever quatre grands défis – technique, économique, juridique et citoyen – qui limitent la pleine exploitation des données de santé. Or, ces défis ne pourront être relevés qu'à la condition d'une définition claire d'une stratégie et d'une gouvernance nationale.** Les avancées nationales des dernières années ont eu des difficultés à être mises en œuvre sur le terrain, en raison notamment de la multiplicité des acteurs et de leur éloignement géographique. Par ailleurs, différentes initiatives nationales et locales se superposent aujourd'hui sans qu'une priorité claire ne soit énoncée, ni un ciblage des moyens mis en œuvre.

La refondation de la gouvernance nationale est évoquée au sein du rapport de la mission interministérielle sur les données de santé (2024). **Pour la mission, il s'agit d'abord de penser la constitution d'une instance interministérielle, à travers le renforcement du rôle du comité stratégique des données de santé, capable de clarifier l'écosystème lié**

<sup>128</sup> Jillian Oderkirk, "Survey Results: National Health Data Infrastructure and Governance", *OECD Health Working Papers*, n°127, 2021.

**aux données de santé.** D'autre part, le positionnement du *Health Data Hub* comme coordinateur et fédérateur des différentes instances et son affirmation par rapport aux autres acteurs sont présentés comme des leviers majeurs.

#### 3.1. UN DÉFI TECHNIQUE : DES BASES DE DONNÉES DE SANTÉ À STANDARDISER ET À RENDRE INTEROPÉRABLES

##### a. Un enjeu d'interopérabilité entre les données de santé et les différentes bases

Les données issues des systèmes d'information en santé, de par leur nombre, leur format et l'usage initial prévu lors de leur collecte, demeurent hétérogènes. Elles diffèrent ainsi par leur type, leur finalité, leur niveau de qualité et d'exhaustivité. **Par ailleurs, l'usage la donnée de santé à des fins secondaires est dépendant du format dans lequel elle a été collectée dans le cadre de son utilisation primaire.**

La numérisation croissante impose aux professionnels de santé de communiquer de manière dématérialisée et d'échanger des informations issues de leurs logiciels médicaux. Néanmoins, *« pour que cette communication dématérialisée soit possible, il faut que tous les acteurs et surtout, tous leurs systèmes d'information parlent le même langage<sup>129</sup> »*. **Ce langage commun est couramment désigné comme l'interopérabilité des systèmes d'information en santé.** L'interopérabilité offre ainsi la possibilité d'échanger des données de différentes origines (logiciels médicaux, format, etc.)<sup>130</sup>. **L'interopérabilité dans le cadre de l'utilisation primaire concerne l'échange de données individuelles** (dossier patient par dossier patient)

<sup>129</sup> « Vous reprendrez bien un peu d'interopérabilité », site internet de l'Agence du numérique en santé.

<sup>130</sup> Marcel Goldberg, Marie Zins, « Le Health Data Hub (suite), Pourquoi ? Comment ? », *Med. Sci.*, Volume 37, Number 3.

tandis que lors **d'une utilisation secondaire, cela implique l'échange de bases de données complètes** (données en masse).

**Un des premiers types d'interopérabilité est l'interopérabilité sémantique, qui renvoie à la capacité des systèmes d'information à échanger l'information textuelle** : chaque donnée est ainsi reliée à un vocabulaire partagé et contrôlé. **Ce vocabulaire partagé est généralement défini au niveau international et est couramment désigné comme terminologie de santé.** Au sein des logiciels médicaux, les données saisies par le professionnel de santé sont codifiées et classées dans la norme terminologique.

Plusieurs terminologies distinctes existent en fonction des domaines et peuvent être citées (*liste non exhaustive*) :

- la Classification internationale des maladies 10 (CIM 10), utilisée majoritairement par les hôpitaux et qui est la norme utilisée pour l'alimentation du PMSI ;
- la Classification internationale des Soins Primaires (CISP), utilisée dans le cadre des soins primaires et de la médecine de ville<sup>131</sup> ;
- la LOINC, utilisée dans le cadre des analyses des laboratoires et de biologie<sup>132</sup> ;
- la SNOMED-CT, terminologie multi-domaines internationale adoptée récemment par la France<sup>133</sup>.

**Pour pouvoir être échangées entre les différents systèmes d'information, les données de santé, stockées dans des logiciels médicaux distincts, doivent aussi répondre à des obligations d'interopérabilité technique avec des structures et des formats compatibles.** À ce titre, le standard HL7-FHIR (*Fast Healthcare Interoperable Resources*) constitue un standard international adopté dans de nombreux pays<sup>134</sup> et est de plus

<sup>131</sup> « Terminologie – CISP2 », site internet de l'Agence du numérique en santé.

<sup>132</sup> « Terminologie – LOINC International », site internet de l'Agence du numérique en santé.

<sup>133</sup> « SNOMED-CT », site internet de l'Agence du numérique en santé.

<sup>134</sup> OCDE, « Digital Health a a glance », *Health at a Glance 2023: OECD Indicators*, 2023.

en plus répandu en France. Il assure l'interopérabilité entre les échanges de données issues de différents logiciels dans le cadre du soin.

Lancé en 2021, le programme du Ségur du numérique en santé a investi le champ de l'interopérabilité avec pour objectif de « *généraliser le partage fluide et sécurisé des données de santé entre professionnels et usagers*<sup>135</sup> ». Le Ségur du numérique a ainsi bénéficié d'un investissement inédit de 2 milliards d'euros jusqu'en 2026 et a notamment permis de définir des référentiels communs entre les différents logiciels métiers. Le Ségur du numérique en santé, piloté conjointement par la Délégation au numérique en santé (DNS) et l'Agence du numérique en santé (ANS), avait pour priorité principale l'interopérabilité à des fins d'usage primaire. **À ce titre, ce programme n'intègre pas de stratégie autour de l'utilisation secondaire des données de santé.** Néanmoins, il convient de saluer l'effort d'harmonisation en cours pour l'utilisation primaire, car la mise en place de ces standards communs pourra bénéficier à terme aux usages secondaires, les données primaires étant en amont mieux structurées.

Pour autant, dans le cadre de l'utilisation secondaire, l'échange des données ne concerne plus l'échange d'un dossier patient entre des professionnels de santé mais bien un échange de données en masse. **Des standards distincts d'interopérabilité doivent ainsi être appliqués afin d'assurer le bon échange entre les bases**, à l'image du modèle OMOP (*Observational Medical Outcomes Partnership*). Ce type de modèle standardise différentes sources de données dans un même format, à des fins de stockage et d'analyse de la donnée. Ce processus implique une correspondance vers un ensemble de tables de données, avec des colonnes communes<sup>136</sup>. Le modèle de données OMOP définit également les nomenclatures à utiliser, soit la liste des codes utilisables dans chaque cellule de données et leur signification. Le *Panorama de la santé 2023* de l'OCDE précise notamment que ce modèle a la faculté de normaliser les

<sup>135</sup> « Le Ségur du numérique en santé », site internet de l'Agence du numérique en santé.

<sup>136</sup> HAS, *Entrepôts de données de santé hospitaliers en France, Quel potentiel pour la Haute Autorité de santé ?*, 2022.

sources de données pour une analyse plus globale. Il indique également que les standards relatifs à l'OMOP devraient fonctionner avec les autres standards d'interopérabilité technique « *de manière à ce que les données puissent être collectées une seule fois à des fins primaires et utilisées à de nombreuses reprises à des fins secondaires*<sup>137</sup> ».

Tableau 2 : des niveaux d'interopérabilité distincts

Utilisation	Utilisation primaire Échange de données individuelles		Utilisation secondaire Échange de données en masse
	Interopérabilité sémantique	Interopérabilité technique	Interopérabilité entre bases de données
Catégorie d'interopérabilité	<p><b>Usages :</b> échange d'information textuelle, construction d'un vocabulaire entre les SI.</p> <p><b>Standards existants :</b> nombreuses terminologies (CIM, CISP, SNOMED-CT).</p>	<p><b>Usages :</b> échange d'information entre des SI distincts.</p> <p><b>Standards existants :</b> FHIR.</p>	<p><b>Usages :</b> construction de bases de données issues du soin, appariement entre des bases distinctes.</p> <p><b>Standards existants :</b> OMOP.</p>

**La question de l'interopérabilité est notamment cruciale pour les entrepôts de données de santé hospitaliers** qui montent en puissance dans les établissements de santé, malheureusement sans standardisation des données ni uniformisation des formats et solutions techniques retenus. Un rapport de la Haute Autorité de santé déplore ainsi que « *malgré la diffusion du modèle OMOP pour les études multicentres, aucun modèle commun de données ne fait actuellement consensus. Les transformations de données depuis les SI sources vers les jeux de données d'étude sont peu documentées publiquement*<sup>138</sup> ». Les EDS présentent un

enjeu majeur pour la mise à disposition de données cliniques mais ceux récemment créés ou en cours de création demeurent assez hétérogènes. **Des standards propres à chaque EDS existent sans aucune obligation de standardisation.** Un groupe de travail « Ensemble des variables communes à tous les entrepôts » a rendu publique en septembre 2023 une proposition de socle commun de données et variables de santé à collecter pour les entrepôts de données hospitaliers<sup>139</sup>. Ce socle, composé de 51 items tels que des données cliniques, d'admissions, médicamenteuses ou les diagnostics, a pour objectif d'établir un standard pour les données, en matière de collecte, mise en qualité et standardisation (définition de spécifications techniques et de terminologies).

**L'enjeu est d'assurer un minimum de cohérence et d'interopérabilité entre les bases de données des hôpitaux.** À titre d'exemple, le *Danish Health Data Authority* au Danemark est chargé de développer des standards pour uniformiser les données collectées<sup>140</sup> venant alimenter les registres nationaux (*Danish Clinical Registries*). **La mise en place de standards communs à la constitution des bases de données issues du soin permettrait à terme de constituer des réseaux territoriaux d'entrepôts de données de santé.** Les Hôpitaux universitaires du Grand Ouest (GCS HUGO) ont notamment mis en place une plateforme interrégionale des données de santé capable d'intégrer et d'analyser les données issues des différents CHU du réseau (Nantes, Rennes, Brest, Angers, Tours). Pour pouvoir mettre en place ce réseau territorial de données de santé, les divers entrepôts développés par les différents CHU ont été construits selon une même architecture technique et logicielle afin d'assurer l'interopérabilité des données. Cette plateforme est hébergée dans un environnement sécurisé au sein du CHU de Nantes.

<sup>137</sup> *Panorama de la santé 2023, OCDE.*

<sup>138</sup> HAS, *Entrepôts de données de santé hospitaliers en France, Quel potentiel pour la Haute Autorité de santé ?*, 2022.

<sup>139</sup> Note à l'attention du Comité Stratégique du Système National des Données de Santé, « Définition d'un socle de données commun aux entrepôts de données de santé hospitaliers », 25 septembre 2023.

<sup>140</sup> "National health registers", site internet du *Danish Health Data Authority*.

## Proposition 1

Mettre en place des règles de standardisation communes aux bases de données de santé nationales (à l'image du modèle OMOP) afin d'avoir un format commun pour faciliter le partage.

- La définition de ces standards communs doit être mise en priorité autour des entrepôts de données de santé hospitaliers afin de constituer des réseaux territoriaux de données de santé.
- La définition de ces règles de standardisation doit être réalisée en concertation avec la Délégation du numérique en santé, l'Agence au numérique en santé et le *Health Data Hub* selon une répartition claire des rôles :
  - la Délégation au numérique en santé (DNS) aura pour mission de choisir le format à respecter à l'échelle nationale (sémantique, classification, nomenclature, transcription informatique etc.) ;
  - l'Agence du numérique en santé (ANS), opérateur de la DNS, apportera son appui technique pour étudier la faisabilité et la maturité des bases actuelles pour basculer vers un standard commun et construire des scénarios de transition vers des modèles communs entre les différents gestionnaires des bases de données de santé ;
  - le *Health Data Hub* sera l'opérateur à même d'identifier les besoins du terrain et d'accompagner les acteurs dans la transition vers un modèle de standardisation commun.
- Une étude de faisabilité menée préalablement sera indispensable pour évaluer le temps et le coût de la mise en œuvre.

- b. La médecine de ville : l'enjeu de la collecte et de la constitution des bases de données

**L'enjeu de l'interopérabilité est encore plus saillant au sein de la médecine de ville.** Les professionnels de santé exerçant en libéral utilisent des logiciels médicaux très variés dont les standards de bases de données sont amenés à converger dans le cadre du Ségur mais demeurent encore extrêmement liés à leurs systèmes d'information. À la différence de l'hôpital, où le niveau de rémunération dépend de la justesse et la finesse du codage, il n'existe aucune incitation pour les médecins généralistes et spécialistes exerçant en libéral à rentrer les données de leurs consultations dans un format homogène. En plus de l'absence d'incitation financière, ces professionnels de santé n'ont aucune possibilité de visualiser l'intégralité de leurs données ni de se comparer avec les professionnels qui les entourent. **Ces données de santé ne sont ainsi pas regroupées dans une base de données commune et demeurent éparpillées entre les différents éditeurs de logiciels privés.**

Au Royaume-Uni, l'initiative *General Practice Data for Planning and Research data collection* portée par le *NHS Digital* a pour **objectif de réunir les 300 bases de données collectant les données des médecins généralistes au sein d'une plateforme nationale unique agrégée**<sup>141</sup>. Les données collectées de manière systématique par le *NHS Digital* renseignent ainsi sur les diagnostics des patients, leurs résultats de tests médicaux, leur santé physique, mentale et sexuelle ainsi que sur leur sexe, ethnie et orientation sexuelle. Le *NHS Digital* ne collecte pas les noms, adresses et champs libres présents au sein des dossiers médicaux des médecins<sup>142</sup>.

Par ailleurs, la médecine de ville utilise principalement la terminologie de la CISP (Classification internationale des soins primaires) à la différence des hôpitaux qui utilisent quant à eux majoritairement la CIM 10.

<sup>141</sup> "About the GDPR programme", site internet du *NHS Digital*.

<sup>142</sup> *ibid.*

L'interopérabilité et la traçabilité entre ces différentes bases de données demeurent pourtant essentielles, dans la perspective de reconstruire l'intégralité du parcours de soins d'un patient. **La construction récente de réseaux d'entrepôts de données de santé en ville**, à l'image du projet P4DP ou encore l'EDS CONTACT de SOS Médecins France, **constituent des premières expérimentations qui permettront de définir les critères d'interopérabilité entre soins primaires et secondaires**. À titre d'exemple, l'EDS CONTACT piloté par SOS Médecins France a pour objectif de structurer les données de soins des 1 300 médecins de la Fédération issues des visites à domicile<sup>143</sup>.

D'autres pays, comme les États-Unis, ont investi dans le développement de standards à destination des soins ambulatoires. Développé par l'ONC à partir de 2020, le *United States Core Data for Interoperability*<sup>144</sup> (USCDI) est un jeu standardisé de classes et d'éléments de données de santé développé pour faciliter une meilleure interopérabilité et des échanges d'informations de santé au niveau national. À partir du 1<sup>er</sup> janvier 2023, les logiciels des prestataires de soins devaient être conformes aux standards USCDI pour être reconnus comme systèmes électroniques de dossiers médicaux. **La mise en place de ces standards permet la mise en œuvre d'un dossier médical unifié sur l'ensemble du territoire**<sup>145</sup>.

<sup>143</sup> « CONTACT, l'entrepôt de données de santé piloté par SOS Médecins France », site internet de [SOS Médecins France](#).

<sup>144</sup> USCDI, "United States Core Data for Interoperability", Draft Version 4 (January 2023).

<sup>145</sup> La solution technique Healthshare permet notamment de collecter des données diverses et de les harmoniser (alignement des nomenclatures, rapprochement des identités, suppressions des doublons, etc.).

## Proposition 2

**Constituer un réseau d'entrepôts de données de santé dédié à la médecine de ville, en commençant par définir un socle minimum de données du dossier patient à collecter lors des consultations de ville. Cette liste devra être intégrée aux logiciels métier en cours de mise à niveau dans le cadre du Ségur du numérique.**

- **Définir, en collaboration avec les syndicats de médecins de ville, un socle commun de données** (poids, résultats de tests médicaux, diagnostics, sexe) **à récolter de manière systématique par les médecins de ville**. À l'image des investissements mis dans le Ségur du numérique, les éditeurs de logiciel seront incités à mettre en conformité leurs solutions afin d'intégrer ce socle et d'assurer la collecte et la standardisation des données saisies.
- La mise en œuvre devra mobiliser les acteurs suivants :
  - **Délégation au numérique en santé / Agence du numérique en santé** : en concertation avec les syndicats de médecine de ville, ces deux institutions seront en charge de définir un socle minimum de données de santé cliniques collectées par les professionnels de santé de ville.
  - **Éditeurs de logiciels** : les éditeurs de logiciels devront être incités à modifier leurs solutions pour faciliter le travail de saisie des médecins de ville et collecter ces données dans un format standardisé répondant aux critères d'interopérabilité.
  - **Groupements Régionaux d'Appui au Développement de la e-Santé (GRADEs)** : placés au niveau régional, sous l'autorité des Agences régionales de santé (ARS), ces groupements auront pour objectif d'accompagner



les médecins de ville dans la prise en main de ces nouveaux outils et ainsi de faciliter leur adhésion au partage de données.

c. Un retraitement souvent nécessaire pour la constitution des bases de données de santé

**Les données de santé, bien souvent stockées sous forme « brute », sans traitement, demeurent difficiles à exploiter. La donnée de santé doit être de qualité, fiabilisée et exploitable en masse.** Pour ce faire, il est nécessaire que les informations contenues dans les données soient standardisées, c'est-à-dire inscrites dans une nomenclature commune permettant de comparer des éléments de même nature : un diagnostic médical, des effets secondaires, la prescription d'un traitement.

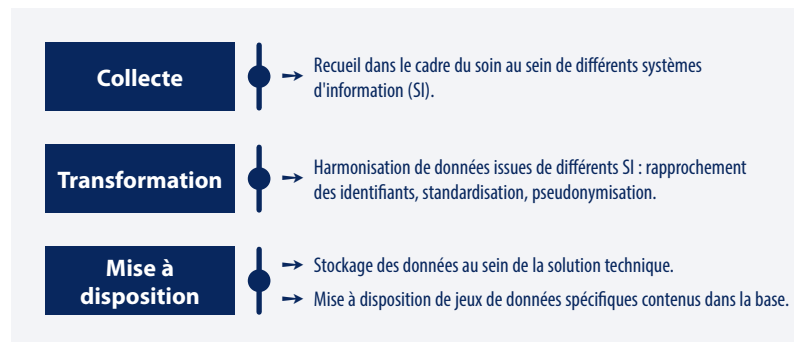
À titre d'exemple, un rapport de la Haute Autorité de santé de 2022<sup>146</sup> distingue trois phases clés pour structurer les données d'un entrepôt de données de santé hospitalier dans le cadre de la réutilisation secondaire :

- **la collecte** à partir des différents systèmes d'information qui permet de « copier » ces données, générées initialement dans le cadre du soin, dans des environnements techniques distincts du soin pour ensuite les charger dans l'EDS ;
- **la transformation** pour « intégrer, harmoniser et mettre en qualité ». Cette étape de transformation et d'agrégation essentielle comporte l'intégration des données, la déduplication des identifiants (notamment des patients, codes de séjour etc.), la standardisation en utilisant un modèle de données commun pour harmoniser les données issues de logiciels différents et la pseudonymisation (suppression des éléments directement identifiants des données) ;

<sup>146</sup> HAS, « Entrepôts de données de santé hospitaliers en France, Quel potentiel pour la Haute Autorité de santé ? », 2022.

- **la mise à disposition des jeux de données**, que ce soient des sous-ensembles de données ou des visualisations des données stockées dans l'entrepôt.

Graphique 2 : les étapes de la construction d'une base de données



**La donnée de santé en France est faiblement standardisée.** Si le codage de données médico-administratives (les actes de soins), notamment dans les hôpitaux, obéit à des nomenclatures communes (comme la CIM 10), de nombreuses données produites à l'hôpital de nature médicale (comptes-rendus de consultations ou de réunions de l'équipe soignante) ne sont pas répertoriées dans le PMSI et, si elles sont collectées, sont de nature textuelle, c'est-à-dire qu'elles ne sont pas standardisées et donc difficilement exploitables sans retraitement. **On estime que 80% des informations de santé<sup>147</sup> ne sont disponibles que dans des documents non-structurés** (lettres d'adressage, comptes-rendus médicaux, résultats d'analyse, ordonnances...).

**Le travail de traitement de la donnée textuelle et sa standardisation reviennent donc à l'acteur qui cherche à exploiter les données de santé**, que ce soit un organisme de recherche, un industriel ou un

<sup>147</sup> « L'IA au service de la reconnaissance contextuelle des documents médicaux », Hospitalia, 15 mai 2023.

établissement hospitalier. **De nombreuses innovations technologiques, mobilisant l'intelligence artificielle, se développent afin de travailler la donnée textuelle.** C'est le cas des *Large Language Models* (LLM), outils générés par l'IA et aptes à comprendre le langage humain. Ils sont utilisés pour la structuration, le traitement, la synthèse ou l'analyse des données. Unicancer a également mis au point son propre moteur de recherche, « ConSoRe »<sup>148</sup>, grâce à une intelligence artificielle, pour retrouver des informations structurées ou textuelles dans les dossiers des patients.

En parallèle, des initiatives des établissements et des logiciels développés par des acteurs privés proposent des solutions susceptibles de **standardiser et structurer la donnée textuelle issue des comptes-rendus médicaux.** Ces outils demeurent coûteux et complexes à élaborer face à la masse importante de données de natures très différentes qui demande un traitement lourd. Certaines *start-ups* en santé, comme Codoc, proposent des outils pour répondre à cette problématique de traitement. L'entreprise a collecté 1,7 million de documents par mois, hébergés dans les EDS des établissements parties prenantes au projet et développé un moteur de recherche multimodal permettant de lier des données textuelles et des données structurées.

**Le travail de standardisation des données est coûteux, chronophage et demande d'importants moyens humains pour les acteurs exploitant les données de santé.** Les profils requis demeurent encore trop rares : il est nécessaire de combiner à la fois des compétences médicales et techniques, pouvant allier des connaissances en santé et en informatique. Dans le cadre des entrepôts de données de santé hospitaliers, les équipes projet nécessitent d'être pluridisciplinaires et regroupent, outre des chercheurs, des *data analysts*, des bio-informaticiens, des techniciens de tous niveaux. Par ailleurs, les salaires proposés au sein des établissements de santé et/ou instituts de recherches restent encore peu attractifs pour des *data analysts* ou *data managers* spécialisés.

<sup>148</sup> SIH Solutions, « ConSoRe : tracer l'histoire de la maladie en structurant les données médicales », décembre 2018.

Ainsi, de telles équipes font encore trop défaut dans la construction des EDS et au sein même des établissements de santé. **Ces derniers, bien que dotés de directions de systèmes d'information, sont encore trop peu nombreux à avoir pris un tournant stratégique et à avoir dédié une direction entière à la stratégie autour des données de santé.** Lorsqu'elles existent, ces directions sont rarement placées à un niveau stratégique de la gouvernance.

### La direction des données de l'Institut Curie

Centre pionnier de lutte contre le cancer, l'Institut Curie a été notamment précurseur *via* la création d'une direction consacrée aux données de santé en 2017. Cette équipe composée de 15 personnes accompagne 80 équipes de recherches hébergées de l'Institut Curie.

L'objectif de la direction des données est de tirer parti des données de l'Institut Curie produites dans le cadre du soin pour la recherche. L'Institut Curie dispose de plus de **15 millions de documents médicaux, images radiologiques, lames numérisées d'anatomo-pathologies et données dites « omiques » issues de près de 580 000 dossiers patients**<sup>149</sup>. La direction des données assure ainsi un retraitement de ces données pour permettre leur réutilisation à des fins de recherche. Elle souhaite également tirer le plein potentiel des données de santé en développant des partenariats avec les acteurs académiques et industriels de santé<sup>150</sup>.

<sup>149</sup> VivaTech23 : « faire preuve d'intelligence numérique pour utiliser demain l'IA » (J. Guérin, Curie), Health & Tech, juin 2023.

<sup>150</sup> Institut Curie : Julien Guérin prend la tête de la direction des données, site internet de l'Institut Curie, décembre 2022.

L'intégration d'équipes dédiées à la stratégie des données représente un élément essentiel, dans la valorisation de l'ensemble des données collectées par un établissements de santé à des finalités de soin. Ces équipes sont notamment chargées de définir les usages prioritaires de la donnée ainsi que de choisir les innovations technologiques prioritaires susceptibles d'appuyer leur stratégie. Récemment, le CHU de Grenoble Alpes a ainsi annoncé la création d'une direction *Data* pour 2024<sup>151</sup>.

### Proposition 3

**Intégrer, au sein des établissements de santé, des *Data Pôles* chargés de définir une stratégie des données de santé.** Coordonnées par un *Chief Data Officer* et placées au niveau de la direction générale, ces équipes seront chargées de définir les priorités des usages de la donnée et de retraiter les données cliniques afin de constituer des bases de données fiables.

- Les *Data Pôles* et le *Chief Data Officer* auront pour missions principales de :
  - définir les usages prioritaires des données de santé de l'établissement et de préparer les usages des données par l'IA afin d'optimiser le traitement des données textuelles ;
  - prioriser les partenariats éventuels de l'établissement pour le traitement et l'usage de données.
- Ces équipes spécifiques pourraient être créées en priorité au sein des CHU pilotant des entrepôts de données de santé hospitaliers déjà importants. Leur financement serait conditionné à la mise en place de tarifications et

contrats types permettant de rémunérer leurs services (cf prochaine proposition).

### 3.2. UN DÉFI ÉCONOMIQUE : L'INSUFFISANTE VALORISATION ÉCONOMIQUE DE LA DONNÉE ET DU TRAVAIL DE CONSTITUTION DES BASES DE DONNÉES

a. L'utilisation secondaire de la donnée :  
une collecte, un traitement et un hébergement coûteux

Le potentiel de création de valeur lié à une pleine exploitation des données de santé est estimé à 7,3 milliards d'euros par an pour la France<sup>152</sup>. Pour autant, leur valorisation est encore faible, à l'image des efforts que peuvent déployer les différents acteurs pour en améliorer la qualité et la mettre à disposition des organismes pertinents. Ces opérations ont un coût non négligeable, dont le financement n'est pas encore intégré à la valeur même de la donnée. **La vaste majorité des acteurs auditionnés, du secteur privé comme du secteur public, partage ce constat et déplore l'absence d'un modèle économique de la donnée de santé.**

La mise en place de programmes nationaux ou locaux utilisant la donnée de santé impliquent des investissements significatifs. **Les outils techniques de collecte, de traitement ou de standardisation de la donnée sont des outils coûteux dont il est difficile de mesurer le retour sur investissement.** Pour autant, d'après l'OCDE, il est estimé qu'un investissement d'un dollar dans la santé numérique a un potentiel de rendement de trois dollars, au regard des bénéfices supposés sur l'amélioration des résultats en matière de santé et la réduction du gaspillage généré par la redondance des actes<sup>153</sup> administratifs. La majorité de l'enveloppe de

<sup>151</sup> « CHU de Grenoble : une direction *Data* pour optimiser l'utilisation des données », *Health & Tech*, 15 janvier 2024.

<sup>152</sup> *Veltys et Health Data Hub*, « Valorisation économique de l'usage secondaire des données de santé », juillet 2022.

2 milliards d'euros du Ségur du numérique en santé était consacrée à la structuration des données et à la mise en place de référentiels à des fins primaires. Ce dernier point constitue un investissement essentiel dans la mesure où **la mise à niveau de la collecte des données dans le cadre du soin aura des conséquences et répercussions sur la maturité des données à traiter à des fins secondaires**. D'autres investissements significatifs ont été réalisés dans le cadre du plan d'investissement de France 2030, par le biais duquel la santé numérique avait été dotée de 718,4 millions d'euros<sup>154</sup>. À ce titre, **un appel à projets a été lancé en juillet 2022 pour renforcer des entrepôts de données de santé hospitaliers : une enveloppe de 75 millions d'euros** (dont 40 millions d'euros sont issus des fonds France 2030 et 35 millions d'euros sont issus de l'Ondam<sup>155</sup>) a été allouée pour financer en deux vagues les projets lauréats.

Les six lauréats de la première vague ont ainsi pu bénéficier de 40 millions d'euros sur près de 3 ans<sup>156</sup> (soit 2 millions par an et par entrepôt). Pour autant, le groupe de travail « Financement des bases de données de santé publique » du comité stratégique des données de santé estime dans une note rendue publique en avril 2023<sup>157</sup> que « *l'appel à projets EDS France 2030 constitue un financement d'amorçage considérable mais n'offre cependant pas de perspective pérenne pour les établissements de santé* ». **Ce groupe de travail évalue le coût de développement d'un entrepôt de données de santé à 2 millions d'euros par an sur 5 ans, puis à 3 millions d'euros par an le coût de son fonctionnement nominal**. Dans un premier temps, il serait ainsi nécessaire de fournir une dotation de 60 à 90 millions d'euros par an pour financer 30 EDS (soit 1 par CHU).

<sup>153</sup> OCDE, *Panorama de la santé*, 2023.

<sup>154</sup> « France 2030 : six premiers lauréats pour constituer et consolider des entrepôts de données de santé hospitaliers », site du [Gouvernement](#), avril 2023.

<sup>155</sup> « Data : les 10 nouveaux lauréats de l'AAP pour l'accompagnement à la constitution d'EDS hospitaliers », *Health and Tech*, 18 décembre 2023.

<sup>156</sup> *Ibid.*

<sup>157</sup> Note à l'attention du Comité Stratégique du Système National des Données de Santé, financement des entrepôts de données de santé (EDS) au sein des CHU, mars 2023.

## L'entrepôt des données de santé des Vétérans aux États-Unis

La *Veterans Health Administration*<sup>158</sup> est l'unité du ministère des Anciens Combattants des États-Unis qui met en œuvre le programme de santé du ministère, via un système de sécurité sociale nationalisé.

En juin 2023, le système de soins du ministère des Anciens Combattants comptait plus de 9 millions d'inscrits pour 1 323 établissements de santé<sup>159</sup>. Les dossiers patients constitués et recueillis par les différents acteurs sont enregistrés dans le système *VistA* (*Veterans Health Information Systems and Technology Architecture*), adopté par le ministère dès les années 1980<sup>160</sup>. Le système *VistA* couvre plus de 180 applications cliniques, financières et administratives intégrées à un unique entrepôt de données. **Les coûts de maintenance et d'évolution du système *VistA* ont été estimés par le Government Accountability Office à près de 3 milliards de dollars entre 2015 et 2017**<sup>161</sup>.

En 2018, le ministère des Anciens Combattants a lancé un programme de modernisation de ses dossiers patients informatisés (*Electronic Health Record Modernization program*<sup>162</sup>), visant à remplacer les 133 instances *VistA* par une nouvelle solution commerciale *Millenium*<sup>163</sup>. Résultant d'un contrat conclu entre le ministère

<sup>158</sup> Site internet du [Veterans Health Administration](#).

<sup>159</sup> Site internet du ministère des Anciens Combattants, *Statistics at a Glance*, juin 2023.

<sup>160</sup> Site internet du [ministère des Anciens Combattants](#).

<sup>161</sup> Site internet du [Government Accountability Office](#).

<sup>162</sup> Site internet du ministère des Anciens Combattants, *Frequently Asked Question*.

<sup>163</sup> [Cerner Millenium](#).

et Cerner de 10 milliards de dollars pour 10 ans<sup>164</sup>, le programme de modernisation vise à renforcer l'interopérabilité des données du ministère des Anciens Combattants et celles du ministère de la Défense et de la Garde Côtière.

La constitution de bases de données de santé fiables et pérennes représente un coût conséquent dans le temps long. **Pour autant, les acteurs qui constituent les bases de données peinent à avoir de la visibilité sur les financements octroyés, qui demeurent par ailleurs limités dans le temps.** D'autre part, ces financements restent insuffisants pour couvrir les coûts liés à la collecte, à la standardisation et à la fiabilisation des données.

**Aujourd'hui, l'absence de modèle de financement pérenne pour la collecte et le partage des données de santé donne la prévalence au « réflexe propriétaire » pour ceux qui constituent des bases de données.** La mise en place de financements structurels pour la collecte des données de santé pourrait donc permettre, au-delà de garantir leur pérennisation, le développement de leur potentiel.

- b. Un système de redevance à mettre en place pour les producteurs des bases de données

**Il existe peu ou pas d'incitations financières pour rétribuer le producteur de la donnée ou pour inciter les détenteurs de la donnée à la partager.** Au contraire, on constate même une tendance à limiter l'usage de la donnée aux établissements qui la produisent, freinant ainsi les initiatives de construction de bases de données partenariales.

<sup>164</sup> «VA Signs Contract with Cerner for an Electronic Health Record System», Site internet du [ministère des Anciens Combattants](#).

**Comment élaborer un système de valorisation de la donnée de santé?** La vente des données de santé est interdite mais il est possible pour un producteur de données de facturer des prestations de service. **Aujourd'hui, il n'existe pas de contrats types et de modèles de tarification permettant de rétribuer le producteur de la donnée pour son travail de stockage, de constitution de la base et de mise en qualité de la donnée.** Des partenariats rémunérés sont bel et bien réalisés, mais l'absence d'un cadre national freine la visibilité financière tant du côté de l'exploitant de la donnée que du producteur. Le manque de lisibilité sur une doctrine nationale est ainsi source de freins à l'innovation.

Une réflexion portée par le groupe de travail sur le financement du comité stratégique des données de santé indique que **le financement des entrepôts de données de santé hospitaliers doit être un financement mixte<sup>165</sup>, incluant une part fixe** (coûts de cadrage, de mise en place et autres coûts incompressibles de fonctionnement tels que le maintien en conditions opérationnelles etc.) **et une part variable** définie selon des indicateurs d'activité (nombre de partenariats ou projets réalisés; nombre de publications; nombre de bases partagées dans le catalogue du *Health Data Hub* etc.) et des indicateurs de performance (délai d'accès aux données, niveau de qualité des données, etc.).

Des travaux sont en cours menés par ce même groupe afin d'élaborer des tarifs unifiés liés à la mise à disposition des données de santé<sup>166</sup>. **Cette grille de tarification doit permettre notamment pour les établissements de santé disposant d'entrepôts de données de santé de refacturer les coûts de mise à disposition des données de l'entrepôt et de développer une stratégie de partenariat public / privé** (notamment avec des industriels dans le cadre d'essais cliniques ou d'évaluation suite à la mise sur le marché de leurs produits de santé).

<sup>165</sup> Note à l'attention du Comité Stratégique du Système National des Données de Santé, financement des entrepôts de données de santé (EDS) au sein des CHU, mars 2023.

<sup>166</sup> « Une grille de tarification de l'accès aux données des entrepôts hospitaliers proposée à la rentrée (Stéphanie Combes) », Tic Pharma, juillet 2023.

**Ces tarifs unifiés existent déjà à l'étranger où les institutions nationales facturent des coûts de mise à dispositions de jeux de données de santé aux porteurs de projets.** En Finlande, l'institution nationale *FinData* qui est le point d'accès unique pour les porteurs de projets pour la mise à disposition de données, applique ainsi des **frais**<sup>167</sup> **couvrant ceux de demande de permis, ceux liés à l'extraction et la livraison des données, au temps de traitement et de pseudonymisation des données ainsi que ceux relatifs au maintien en condition opérationnelle de leur plateforme.** Les frais appliqués peuvent varier de 250€ dans le cadre d'une thèse à 3 000€ (pour une demande qui nécessite d'extraire des données auprès de 15 entités ou plus).

Un fonctionnement similaire s'applique au Danemark, où le *Research Services* du *Danish Health Authority* facture ses services auprès des utilisateurs. Les coûts permettent de couvrir le traitement de la demande, l'extraction des données ou encore les conseils aux porteurs de projet. La facturation est réalisée sur une base mensuelle au taux horaire en vigueur, selon le nombre d'heures consacrées par projet<sup>168</sup>. Comme précisé sur le site du *Danish Health Authority*, les coûts appliqués par le *Research Services* couvrent uniquement les coûts relatifs à la mise à disposition des données et les coûts de fonctionnement du service. On retrouve le même fonctionnement au Royaume-Uni avec le *NHS Digital*. Les données de santé ne sont pas facturées en tant que telles mais des frais sont appliqués pour couvrir le coût du traitement et des services proposés. **Ces frais varient selon la demande, le volume de données demandées et la fréquence de mise à disposition**<sup>169</sup>. À titre d'exemple, le *NHS Digital* estime ainsi le coût à 28 300 livres sterling pour 3 ans pour une organisation non-commerciale sur des données hospitalières.

<sup>167</sup> "Pricing", site internet du *FinData* : <https://findata.fi/en/pricing/>.

<sup>168</sup> "About Research Services", site internet du *Sundhedsdata-styrelsen*.

<sup>169</sup> "Data Access Request Service (DARS) charges from 2020/2021", *NHS Digital*.

## Proposition 4

**Faire émerger un modèle économique de la donnée en mettant en place des contrats et tarifications types entre les producteurs de données (CHU, instituts de recherche) et les utilisateurs (chercheurs, industriels, associations de patients) pour permettre aux producteurs des bases de données de santé de facturer les services liés à leur mise à disposition (collecte, traitement, stockage) et rémunérer leur travail de constitution.**

- La mise en place de tels contrats types permettra de renforcer les partenariats publics-privés entre producteurs de bases de données et utilisateurs finaux, qu'ils soient industriels ou chercheurs.
- Une fois ces contrats et tarifications définis, les producteurs de bases de données pourront conclure des partenariats de deux manières distinctes :
  - de manière B2B, entre **un producteur de base de données et l'utilisateur**, ces contrats et tarifications serviront de référence pour assurer une transparence et une homogénéité nationale ;
  - **par l'intermédiaire du *Health Data Hub*** afin que ce dernier centralise les demandes d'accès aux données et assure un service de contractualisation.

### 3.3. UN DÉFI JURIDIQUE : UN CADRE RÉGLEMENTAIRE PROTECTEUR MAIS COMPLEXE À APPRÉHENDER

**Les données de santé recouvrent un périmètre large.** On retrouve les données attachées à une personne physique, celles obtenues lors d'examens médicaux ou de dépistages et celles relatives à la prise d'un traitement, mais également les données relatives à la taille et au poids, ainsi que le croisement de différentes données de mesure avec des données comportementales, pour établir des diagnostics ou des prédictions médicales.

**Le caractère très personnel des données de santé implique un niveau de sensibilité et de protection soutenu afin de les réutiliser.** À ce titre, de nombreux textes législatifs et réglementaires encadrent l'utilisation des données de santé et protègent les droits des personnes concernées. Les réglementations européenne et française se complètent. Au niveau européen, le Règlement général de protection des données de 2018 (RGPD)<sup>170</sup> a vocation à s'appliquer à l'ensemble des pays européens même si chaque État garde une marge de manœuvre sur l'application de certains articles. **Le RGPD classe les données de santé dans la catégorie des données « sensibles », qui font l'objet d'un cadre juridique plus protecteur que les autres données personnelles (article 9 du RGPD).** Le cadre juridique impose également une finalité de traitement à tout projet d'utilisation des données de santé. Ces dernières ne peuvent être traitées que dans un but déterminé et légitime.

Avant la mise en place du RGPD, la France disposait déjà d'un cadre très protecteur des données de santé avec la loi Informatique et Libertés (1978)<sup>171</sup> ainsi que le Code de la santé publique<sup>172</sup> ou encore le Code de l'action sociale et des familles<sup>173</sup>.

<sup>170</sup> Règlement du Parlement européen et du Conseil relatif 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 abrogeant la directive 95/46/CE.

<sup>171</sup> Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

a. Une procédure complexe pour accéder aux données de santé

**L'accès aux données de santé est fortement réglementé : en fonction de la catégorie de la recherche et du type de traitement demandé, des procédures distinctes peuvent être appliquées.** D'après le *Health Data Hub*, il existe près d'une dizaine de procédures d'accès aux données<sup>174</sup>. Les recherches peuvent ainsi être classifiées en deux catégories<sup>175</sup> :

- **Des recherches impliquant la personne humaine (RIPH)**, à savoir des recherches où une collecte de données supplémentaire par rapport à celle déjà effectuée dans le cadre du soin est nécessaire. Il existe trois catégories de RIPH selon les risques encourus par la personne. Pour cette catégorie de recherche, l'avis d'un comité de protection des personnes (CPP) est systématiquement requis.
- **Des recherches n'impliquant pas la personne humaine (RNIPH)**, soit des recherches conduites à partir de données déjà collectées dans le cadre du soin ou pour une recherche antérieure.

**Pour chacune de ces deux catégories, deux types de procédures existent : une procédure classique et des procédures simplifiées. La procédure classique nécessite une autorisation de la Commission nationale de l'informatique et des libertés (CNIL) après un avis du comité d'expertise compétent.** Ce comité d'expertise diffère selon la catégorisation de la recherche :

- Dans le cadre d'une recherche impliquant la personne humaine (RIPH), l'avis du comité des protections des personnes (CPP) doit être rendu dans un délai de 45 jours.

<sup>172</sup> Loi n° 2002-303 du 4 mars 2002 relative aux droits des maladies et à la qualité du système de santé.

<sup>173</sup> Code de l'action sociale et des familles.

<sup>174</sup> « Quelle procédure d'accès aux données pour votre projet de recherche ? », site internet du *Health Data Hub*, juin 2021.

<sup>175</sup> *Health Data Hub*, « Guide pédagogique : procédures d'accès aux données de santé en France », 2021.

- Dans le cadre d’une recherche n’impliquant pas la personne humaine (RNIPH), le Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de santé (CESREES)<sup>176</sup>, doit se prononcer sur l’intérêt des projets de recherche dans un délai d’un mois.

Suite à l’avis du comité d’expert, le dossier est examiné par la CNIL<sup>177</sup> qui émet son avis dans un délai de 2 mois, renouvelable. La commission examine la méthodologie retenue, la nécessité du recours à des données à caractère personnel et leur pertinence par rapport à la finalité du traitement. **Ce régime d’autorisation impliquant plusieurs instances d’autorisation contribue à complexifier le cadre réglementaire**<sup>178</sup>.

**Les procédures simplifiées ne nécessitent pas une autorisation de la CNIL mais impliquent que les porteurs de projet s’assurent que leur recherche remplit les conditions fixées par les textes.** À ce titre, de nombreux référentiels et méthodologies de référence ont été développés par la CNIL pour accompagner les porteurs de projets et simplifier leurs démarches. **Certains projets peuvent se soustraire d’une autorisation expresse de la CNIL s’ils déclarent leur conformité à un référentiel ou à une méthodologie de référence.** Cette procédure permet au porteur de projet de raccourcir le délai d’examen du dossier. Il existe aujourd’hui trois référentiels et huit méthodologies de référence (voir Annexe 1). Deux nouvelles méthodologies de références ont été récemment publiées afin de faciliter l’accès des acteurs aux bases de données du SNDS<sup>179</sup>, dans le cas où leur recherche, étude ou évaluation poursuit une mission d’intérêt public ou un intérêt légitime. Ces deux méthodologies de référence concernent un nombre important de données nécessitant un accord du

<sup>176</sup> « Calendrier du CESREES », site internet du Health Data Hub.

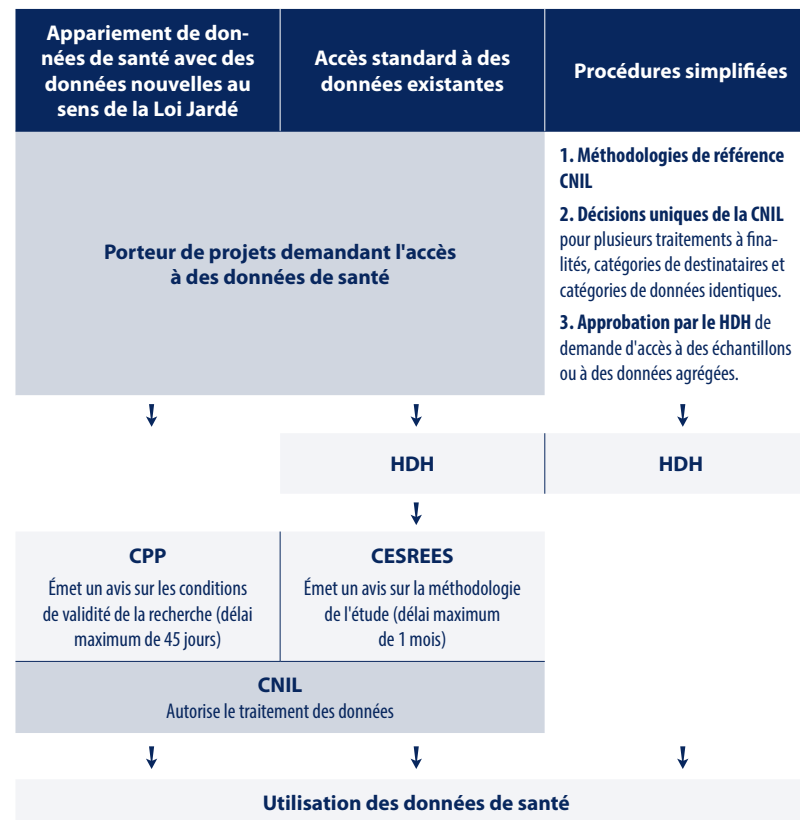
<sup>177</sup> « Demande d’autorisation d’une recherche en santé : les informations à fournir et les critères d’octroi », site internet de la CNIL, janvier 2023.

<sup>178</sup> Jérôme Marchand-Arvier, Pr Stéphanie Allasonnière, Aymeril Hoang, D<sup>r</sup> Anne-Sophie Jannot, « Fédérer les acteurs de l’écosystème pour libérer l’utilisation secondaire des données de santé », décembre 2023. p. 59-60.

<sup>179</sup> « Recherche en santé : la CNIL adopte deux méthodologies de référence permettant l’accès à la base principale du SNDS », site internet de la CNIL, octobre 2023.

CESREES au préalable. D’autres procédures simplifiées existent : certains organismes ont des **accès permanents aux données du SNDS**<sup>180</sup>, notamment les acteurs institutionnels du monde de la santé.

Tableau 3 : schématisation du circuit des demandes d’accès aux données de santé



Source : rapport annuel du Health Data Hub 2022.

<sup>180</sup> Guide accès permanents, Documentation du SNDS et SNDS OMOP.



**Ces procédures s'appliquent aux projets fondés sur le traitement de données pseudonymisées**<sup>181</sup>, soient des données pour lesquelles les caractéristiques personnelles directes de la personne sont remplacées par un identifiant. Cette action n'est pas irréversible, car il reste possible de réidentifier la personne grâce à la conservation d'une table de correspondance qui permet de rapprocher l'identité de la personne à son identifiant dans la base de données. **La donnée anonyme correspond quant à elle à une donnée à laquelle on a retiré tous les caractères d'identification de la personne ou de la pathologie**<sup>182</sup>.

b. Des délais d'instruction des dossiers jugés trop longs

**Le délai d'instruction d'un dossier de demande d'exploitation de données de santé apparaît comme un frein notable pour les porteurs de projet.** Parfois, ils contribuent à mettre en échec les projets déjà financés ou à rendre difficiles des coopérations avec des équipes de chercheurs internationales. Or les délais d'instruction des organes de contrôle, par ailleurs encadrés réglementairement, paraissent comparables à ceux observés dans les autres pays européens<sup>183</sup>. À titre d'exemple, **le délai moyen de traitement d'un dossier par la CNIL en 2022 était de 66 jours pour un projet de recherche, et de 58 jours pour un projet hors recherche** (entrepôt de données de santé, accès précoce...).

**Pourtant, le service dédié à l'accompagnement juridique des données de santé de la CNIL a dû faire face, ces dernières années, à une**

<sup>181</sup> « Recherche scientifique (hors santé) : enjeux et avantages de l'anonymisation et de la pseudonymisation », site internet de la CNIL, janvier 2022.

<sup>182</sup> « L'anonymisation des données, un traitement clé pour l'open data », site internet de la CNIL, octobre 2019.

<sup>183</sup> D'après certains acteurs industriels auditionnés par la mission d'information du Sénat, tel Sanofi, les délais de réponse du Cesrees et de la Cnil sont comparables à ceux requis pour l'accès à des bases de données européennes hors de France, mais beaucoup plus longs que pour les sources de données nord-américaines, dont le modèle d'accès, fondé sur le principe de licences annuelles ou par projet, peut ne prendre que trois à quatre semaines.

**augmentation continue des demandes d'exploitation des données de santé** : composé de 11 personnes en 2022, il a traité 362 demandes d'autorisation en 2019 et 455 en 2022, soit une augmentation de 26%<sup>184</sup>. Son activité de conseil a suivi cette tendance à la hausse et les demandes formelles d'accompagnement sont passées de 115 en 2019 à 154 en 2022, soit une hausse de 34%.

**Toutefois, pour les projets qui nécessitent une extraction du SNDS, après l'autorisation de la CNIL, le délai de mise à disposition des données SNDS par la Cnam semble rallonger significativement les délais. Le délai est de 10 à 12 mois en moyenne (+ 56 % du délai médian entre 2020 et 2022), ce qui porte la totalité de la procédure à un délai moyen de 18 mois**<sup>185</sup>. Les retards pris par la Cnam tiennent à la complexité des demandes qui lui sont adressées, qui exigent de plus en plus des appariements de données, des tirages d'échantillons et des extractions lourdes. Ils s'expliquent également par la forte augmentation des demandes qui lui sont adressées.

À titre de comparaison, l'institution finlandaise *FinData* met à disposition des porteurs de projets les données dans une plateforme indépendante et sécurisée nommée *Kapseli*. Le délai d'obtention des données varie entre 1 à 3 mois. Dans son rapport de juillet 2023, la Commission des Affaires sociales du Sénat estimait ainsi prioritaire de « prévoir des dispositifs transitoires d'appui à la Cnam mais aussi d'engager le transfert de la copie du SNDS au Health Data Hub sur la solution qui sera politiquement assumée<sup>186</sup> ». Le rapport de la mission interministérielle va encore plus loin en proposant la sortie de *Microsoft Azure* d'ici deux ans et l'identification rapide d'une solution d'hébergement transitoire.

<sup>184</sup> Chiffres communiqués par le service santé de la CNIL.

<sup>185</sup> Commission des affaires sociales, Sénat, « Données de santé : une réforme encore en cours de chargement », Rapport d'information, juillet 2023.

<sup>186</sup> Commission des affaires sociales, Sénat, « Données de santé : une réforme encore en cours de chargement », Rapport d'information, juillet 2023.

c. Un cadre réglementaire difficile  
à appréhender pour les acteurs

**Ces règles et procédures de mise à disposition des données de santé font l'objet d'une appropriation partielle par les acteurs.** D'une part, elles ont été fréquemment modifiées : la procédure de demande d'accès à des données de santé hébergées dans le SNDS ainsi que les interlocuteurs auxquels les porteurs de projet doivent s'adresser ont été modifiés dans la loi OTSS de 2019<sup>187</sup>. **D'autre part, les divers éléments à apporter pour constituer un dossier de demande de traitement de données restent complexes à assembler de manière à permettre une instruction efficace et brève.** Les porteurs de projet en amont de chaque instruction doivent ainsi constituer des dossiers complexes<sup>188</sup> contenant notamment l'identification de la juste procédure applicable à leur projet comme la réalisation d'une analyse d'impact relative à la protection des données (évaluant le risque de la recherche) ou encore la documentation visant à prouver la conformité de leur recherche à la procédure choisie.

**Ces référentiels sont majoritairement connus des porteurs de projet mais leur niveau de maturité et d'expertise juridique reste encore très hétérogène** en fonction des structures. Depuis la mise en place du RGPD, chaque structure doit désigner un délégué à la protection des données (DPO)<sup>189</sup>. Il est chargé d'assurer la mise en conformité au RGPD des traitements de données personnelles et notamment de veiller à ce que les données personnelles soient collectées, utilisées, stockées et transférées de manière appropriée et conforme à la réglementation sur la protection des données personnelles.

<sup>187</sup> La loi de 2019 a remplacé l'INDS (Institut national des données de santé) par la Plateforme des données de santé dans la fonction de secrétariat pour le dépôt des demandes.

<sup>188</sup> Health Data Hub, « Guide pédagogique : procédures d'accès aux données de santé en France », 2021.

<sup>189</sup> Section 4, articles 37 à 39 du RGPD.

Cette fonction encore nouvelle peine à pleinement s'imposer au sein des institutions et des entreprises. Par ailleurs, l'instabilité du cadre réglementaire de ces dernières années a rendu son appropriation par ces acteurs encore trop partielle, freinant le bon déroulement des projets. **Il apparaît ainsi prioritaire que les porteurs de projet, les professionnels juridiques qui les accompagnent ainsi que les professionnels de santé montent en compétences sur les démarches réglementaires.**

**Le Health Data Hub constitue un guichet unique, à même d'accompagner les porteurs de projet dans leurs démarches réglementaires**<sup>190</sup>. Il travaille également en étroite collaboration avec la CNIL afin d'élaborer des procédures simplifiées pour l'accès aux données. Ce travail d'accompagnement doit se poursuivre, en coordination avec la CNIL, pour permettre aux acteurs une bonne appropriation du cadre réglementaire. **Des formations dédiées sont à développer pour permettre une montée en compétences de tous les acteurs et assurer à terme leur autonomie.** Des premières formations, développées par le Health Data Hub, existent déjà et permettent notamment de présenter de manière pédagogique les différentes procédures d'accès aux données de santé<sup>191</sup>.

**Ces formations existantes doivent aller encore plus loin et se développer en coopération avec les acteurs de terrain afin que ces derniers puissent présenter des cas concrets qu'ils ont pu rencontrer et partager des solutions pratiques.** À terme, cette communauté, animée par le Health Data Hub et la CNIL pourrait contribuer à la construction de cette offre de formation. **La constitution d'un club « DPO Santé » animé par ces deux institutions pourrait permettre d'adapter cette offre de formations aux besoins du terrain et ainsi, d'adapter certaines procédures réglementaires répondant à ces cas pratiques.**

<sup>190</sup> « Faciliter l'usage des données de santé », site internet du Health Data Hub.

<sup>191</sup> REGL-001 : Présentation des procédures d'accès aux données personnelles de santé, Formations du Health Data Hub, site internet du Health Data Hub.

### 3.4. UN DÉFI CITOYEN : UNE NÉCESSAIRE SENSIBILISATION DES CITOYENS ET DES PROFESSIONNELS DE SANTÉ À L'UTILISATION SECONDAIRE

a. Des citoyens informés de l'utilisation des données de santé mais désireux de la construction d'un cadre de confiance renforcé

**Les études menées sur la perception des citoyens à l'égard de l'utilisation des données de santé font émerger le besoin de renforcer le cadre de confiance et la transparence autour de l'utilisation secondaire de ces dernières.** L'une des principales inquiétudes formulées est celle de la finalité de ce partage et des éventuels abus qui pourraient y être associés.

Une récente enquête menée en 2021 par le *Health Data Hub* et l'association de patients France Assos Santé<sup>192</sup> révèle par exemple que **77% des Français sont désireux d'être mieux informés et de renforcer leurs connaissances sur l'utilisation secondaire des données de santé.** Par ailleurs, seulement 50% des Français connaissent leurs droits en matière de données personnelles. **Les finalités et les bénéfices du traitement de données de santé ne sont ainsi pas spontanément identifiés par les citoyens.**

**Le manque de perception du cadre existant autour des données de santé freine la pleine adhésion des citoyens à la réutilisation de leurs données de santé.** Une étude menée par la Fondation Roche a révélé qu'un Français sur deux n'est pas prêt à partager ses données de santé<sup>193</sup>. Par ailleurs, si 89%<sup>194</sup> des Français déclarent faire confiance à leur médecin traitant ou médecins spécialistes pour garantir la sécurité de leurs

<sup>192</sup> Auprès d'un panel de 1 000 personnes représentatif de la population française.

<sup>193</sup> Fondation Roche, « Données de santé : réconcilier libre-choix et progrès médical », octobre 2019.

<sup>194</sup> Fondation Roche, Rapport de l'Observatoire de l'accès au numérique en santé, juin 2021.

données personnelles, ils sont en revanche un sur quatre à déclarer ne pas faire confiance aux services publics pour garantir cette sécurité. L'acculturation des citoyens à la culture des données de santé, voire leur responsabilisation en la matière, passe ainsi par la participation des patients dans la gouvernance des bases de données de santé. À titre d'exemple, au sein de la gouvernance des entrepôts de données de santé, les représentants des usagers sont systématiquement associés<sup>195</sup>.

Par ailleurs, on assiste à une multiplication des cyberattaques à l'encontre des acteurs de santé : après celle contre l'hôpital de Corbeil-Essonnes fin août 2022 et contre celui de Versailles début décembre 2022, la CNIL a révélé en février 2024 une fuite de données parmi les plus importantes observées en France, touchant 33 millions de personnes<sup>196</sup>. La cyberattaque, qui a ciblé deux opérateurs de tiers payants pour le compte de complémentaires santé, concernerait des données telles que l'état civil, la date de naissance et le numéro de sécurité sociale, susceptibles de servir de base à des attaques de hameçonnage. Ces attaques contribuent à créer un cadre de défiance de la population française vis-à-vis de l'utilisation potentielle de leurs données de santé : selon un rapport du ministère de la Santé paru le 5 février dernier, 86% des Français considèrent que leurs données de santé sont des informations sensibles et 78% craignent notamment qu'elles soient utilisées à des fins commerciales<sup>197</sup>. **Dans un tel contexte, il apparaît prioritaire d'offrir aux citoyens les garanties en termes de sécurité de stockage de leurs données.**

**L'existence d'un cadre réglementaire d'information innovant et en développement, même s'il est perfectible, semble essentiel pour garantir, d'une part, le droit des personnes concernées et construire, d'autre part, un climat de confiance autour de la donnée de santé.**

<sup>195</sup> « La gouvernance des données », site internet de l'*EDS de l'AP-HP*.

<sup>196</sup> Violation de données de deux opérateurs de tiers payant : la CNIL ouvre une enquête et rappelle aux assurés les précautions à prendre, site internet de la CNIL, 7 février 2024.

<sup>197</sup> Ministère du Travail, de la Santé et des Solidarités, *Les Français et le numérique en santé*, février 2024.

Cette préoccupation doit être d'autant plus grande que le traitement et l'utilisation des données de santé échouent à gagner la confiance de nombreux citoyens.

**b.** L'information et le consentement des patients :  
une obligation juridique, garante de la construction  
d'un cadre de confiance

La loi « informatiques et libertés » prévoit une obligation d'information aux patients auxquelles sont rattachées les données de santé faisant l'objet d'un traitement. Le RGPD a renforcé cette obligation et prévoit dans son article 17 un droit à l'effacement permettant à la personne concernée d'obtenir la suppression de ses données personnelles dans certains cas, par exemple lorsque les données ne sont plus nécessaires aux finalités énoncées ou que la personne concernée retire son consentement.

**L'obligation d'information à l'égard de la personne à laquelle les données de santé se rattachent est de deux ordres :**

- **Une information individuelle** des patients auprès desquels ont été recueillies les données à caractère personnel : il s'agit de **recueillir le consentement** du patient ou sa **non-opposition** ;
- **Une information générale** des patients fréquentant des établissements à la transmission de données à caractère personnel : il s'agit d'**informer** le patient, qui conserve néanmoins un droit d'opposition. À l'image du don d'organes, si le patient ne s'oppose pas de manière explicite, ses données sont, de fait, incluses dans la recherche.

**Dans le cadre des procédures simplifiées, et notamment des méthodologies de référence MR003 et MR004, seule l'information générale des patients est nécessaire avant la collecte et l'utilisation des données**<sup>198</sup>. Ces deux catégories de recherche concernent un grand nombre

<sup>198</sup> Health Data Hub, « Guide pédagogique : procédures d'accès aux données de santé en France », 2021.

des recherches : à titre d'exemple, en 2022, le service santé de la CNIL a traité 390 dossiers de MR003 et 652 dossiers de MR004 sur un total de 1 855 dossiers de méthodologies de référence, soit 56 % des dossiers<sup>199</sup>.

Cette mise en place de l'information générale est en mesure de faciliter la mise en œuvre des projets de recherche à la différence du recueil du consentement. **En France, pour l'information patient, le cadre de l'opt-out s'applique : si le patient ne s'oppose pas de manière explicite à la réutilisation de ses données, ces dernières peuvent être utilisées et incluses dans la recherche**<sup>200</sup>. Lors de la création d'un entrepôt de données de santé<sup>201</sup>, dans le cadre de l'application du référentiel ou sur autorisation<sup>202</sup> de la CNIL, les patients sont informés du transfert de leurs données au sein de l'entrepôt. Sans opposition explicite du patient, les données peuvent ainsi intégrer l'entrepôt. La personne conserve néanmoins le droit de s'opposer à tout moment et de demander le retrait de ses données de l'entrepôt. Ce droit d'opt-out a également été introduit au Royaume-Uni en 2018<sup>203</sup>. Il est ainsi possible pour un patient âgé de plus de 13 ans de changer le statut de son consentement *via* son espace personnel virtuel afin d'empêcher que ses données de santé soient utilisées à des fins de recherche médicale<sup>204</sup>.

L'information patient doit néanmoins répondre aux exigences du RGPD : elle doit être « *délivrée de façon concise, transparente, compréhensible et aisément accessible. Elle doit pouvoir être abordable par le grand public* »<sup>205</sup>.

<sup>199</sup> Chiffres communiqués par le service santé de la CNIL.

<sup>200</sup> Nesrine Benyahia, « *Cinq ans après le RGPD, quelles protections pour les données de santé* », Institut Montaigne, mai 2023.

<sup>201</sup> « *La CNIL adopte un référentiel sur les entrepôts de données de santé* », site internet de la [CNIL](#), octobre 2021.

<sup>202</sup> CNIL, *Référentiel relatif aux traitements de données à caractère personnel mis en œuvre à des fins de création d'entrepôts de données dans le domaine de la santé*.

<sup>203</sup> National Data Guardian for Health and Care, *Review of Data Security, Consent and Opt-Outs*.

<sup>204</sup> « *Understanding the national data opt-out* », site internet du [NHS Digital](#).

<sup>205</sup> « *Conformité RGPD : comment informer les personnes et assurer la transparence ?* », site internet de la [CNIL](#), juillet 2019.

Le support d'information est libre : l'information peut être délivrée à l'oral, par écrit ou par tout autre moyen. **Néanmoins, les porteurs de projet et établissements se heurtent bien souvent à une réalité pratique et peinent à identifier le bon canal d'information afin de respecter le droit d'information du patient et lui communiquer des retours sur la recherche à laquelle il aura été associé.** Cela semble d'autant plus nécessaire que, d'après une étude de la Fondation Roche, les citoyens sont 98 % à réclamer un retour d'information en cas de partages de données<sup>206</sup>.

**De plus en plus d'acteurs, y compris des *start-ups* en santé, imaginent ainsi des solutions qui allègent la charge technique de l'information des patients et assurent la délivrance de l'information.** Certains ont mis en place des « portails d'information » numériques permettant l'information générale des patients, les renseignant sur les différentes recherches en cours et leur permettant d'exercer leurs droits d'opposition. Néanmoins, ce type de portail demeure encore trop confidentiel, peu connu par les professionnels de santé et *in fine* peu diffusé aux patients. **Par ailleurs, ces portails sont éclatés entre les différentes institutions et ne permettent pas au patient d'avoir une vision globale, à l'échelle du pays, sur les études susceptibles d'utiliser ses données.**

**L'information et le recueil du consentement font aussi l'objet de solutions numériques de plus en plus fiables et sophistiquées.** Par exemple, le portail « D'Arno » créé par Codoc recense les études utilisant les données de santé d'un patient et lui permet de s'opposer à leur utilisation<sup>207</sup>. Les données peuvent alors être détruites. La connexion à ce portail peut s'effectuer par le biais de France Connect. L'entreprise DrData a quant à elle créée l'outil *Isalid* qui s'assure que l'identification du patient soit correcte et que le processus d'information du patient soit tracé.

<sup>206</sup> Fondation Roche, « Données de santé : réconcilier libre-choix et progrès médical », octobre 2019.

<sup>207</sup> Site internet de [Codoc](#).

## DrData et la solution Isalid

La solution *Isalid*, développée par la *start-up* française, DrData<sup>208</sup> permet **d'informer, recueillir et tracer le consentement des patients** quant à l'utilisation de leurs données de santé, en permettant aux professionnels de santé d'envoyer à leur base de patients des notices d'information et des demandes de consentement. La plateforme donne également au patient, s'il le souhaite, la possibilité de retirer son consentement au traitement des données.

**La solution *Isalid* s'assure que l'identification du patient est correcte, que le processus de consentement éclairé est respecté** et vérifie la qualité des informations (suppression des personnes décédées, etc.). Elle garantit ainsi **une simplicité et une rapidité de l'information et du recueil du consentement des patients** grâce à des *workflows* automatisés.

c. Des pistes pour créer un cadre de confiance capable d'alimenter et maintenir une culture de la donnée

**Pour répondre aux réticences des citoyens, la valorisation des cas d'usage pour lesquels les données de santé sont exploitées constitue un levier incontournable. Cela participe du renforcement de l'aculturation à la donnée de santé, tant auprès des citoyens que des professionnels de santé.** Dans cette optique, le *Health Data Hub* s'est doté d'une direction citoyenne, chargée de mener des études, d'animer des formations, débats et ateliers participatifs pour sensibiliser le grand public aux données de santé. Elle a également pour objectif d'animer

<sup>208</sup> Site internet de [DrData](#).

des consultations et des concertations. Ce type d'initiative, essentielle, demeure encore peu connue du grand public.

**La communication à destination des patients et des professionnels de santé doit contribuer à envisager l'usage de la donnée au-delà de l'échelle individuelle, dans le sens de l'intérêt collectif.** Il est important de faire comprendre aux citoyens que l'utilisation optimale des données de santé repose sur un traitement de masse d'informations, et non sur la manipulation de leurs seules données personnelles. **Une étude montre ainsi que lorsque les citoyens connaissent la finalité de traitement des données de santé et identifient le bénéfice collectif, leur adhésion au partage des données est plus importante**<sup>209</sup>. Dans cette perspective, un rapport du Conseil scientifique consultatif du *Health Data Hub* d'octobre 2023 rapporte les bénéfices liés à l'utilisation secondaire des données de santé en les comparant aux risques associés, notamment ceux liés à la révélation des données ou encore à la surveillance des professionnels de santé. L'ensemble de ces risques sont évalués de bas à quasiment nul par le comité scientifique consultatif<sup>210</sup>.

**Enfin, la construction d'un cadre de confiance passe par une plus grande appropriation des patients comme professionnels de santé du cadre juridique et par le développement d'outils permettant à chacun d'exercer son droit d'opposition et d'information.** Si les portails cités précédemment constituent des pistes sans aucun doute pertinentes, ils doivent être plus centralisés et veiller à demeurer simples et transparents pour répondre pleinement aux exigences d'information et de libre consentement.

<sup>209</sup> *Health Data Hub, Citoyens et données de santé : ce qu'ils en pensent et jusqu'où sont-ils prêts à les partager, 2020.*

<sup>210</sup> *Conseil scientifique consultatif du Health Data Hub, Bénéfices et risques de l'utilisation des données de santé à des fins de recherche, Position paper, janvier 2024.*

Des premiers instruments plus ambitieux se développent, à l'image d'un **portail d'information permettant de sensibiliser les citoyens français aux droits relatifs à leurs données de santé.** La mission interministérielle recommande, dans la continuité de cette initiative, un **portail national de transparence, centralisé et géré par le Health Data Hub.** Il aurait vocation à recenser les traitements sur les données de santé et simplifier les conditions d'exercice des droits d'opposition des patients sur la réutilisation de leurs données. **Mon espace santé<sup>211</sup> constitue une piste intéressante pour héberger ce type de portail,** à la condition que son utilisation devienne systématique et familière pour les professionnels de santé comme pour les patients.

Dans ce climat d'incertitude sur l'usage de la donnée, **les professionnels de santé eux-mêmes ont un rôle central à jouer.** Au croisement entre la recherche et les patients, ils sont la clé de voûte du système de l'utilisation secondaire de la donnée. Il s'agira donc non seulement de les **rendre créateurs de données réutilisables** mais aussi de **les engager sur cette utilisation secondaire.** L'index de confiance Ipsos 2023 démontre que les médecins sont en haut du classement des professions dignes de confiance en France<sup>212</sup>. Avec 64 % des personnes interrogées ayant confiance dans les médecins, ces derniers doivent être pris en compte tant dans la stratégie à mettre en œuvre pour répondre aux inquiétudes des citoyens que dans son élaboration. **L'enjeu est double : convaincre les professionnels de santé et leur donner les clés de compréhension nécessaires pour qu'ils puissent à leur tour répondre aux interrogations que leurs patients pourraient avoir.**

<sup>211</sup> *Mon espace santé a été créé par la loi de 2019 OTSS. Cet espace numérique sécurisé et personnalisé rassemble les données administratives du titulaire du compte, son dossier médical personnel et ses données de remboursement. Son utilisation reste encore minoritaire puisque seulement 11 % des espaces numériques ont été activés. On peut citer également le Dossier Pharmaceutique, créé dans le cadre de la loi HPST de 2009 qui permet aux pharmaciens de consulter l'historique des médicaments prescrits aux patients afin de prévenir des risques liés au cumul de certains médicaments. La CNIL estimait qu'en 2016, 34 millions de dossiers pharmaceutiques étaient actifs.*

<sup>212</sup> *IPSOS Global Trustworthiness Index 2023, Who does the world trust?, octobre 2023.*

## Proposition 5

**Favoriser la communication et l'accès à l'information pour construire une culture citoyenne, confiante et sensible aux bénéfices du partage de la donnée de santé.**

- **Communication vers les professionnels de santé** : de nouveaux dispositifs de communication, portés par l'Assurance maladie, le Ministère de la santé ou le *Health Data Hub*, devraient permettre de **visualiser les bénéfices génériques d'un partage de la donnée** (agrégation de données par territoire, spécificités de la patientèle, comparaison avec d'autres spécialistes, nombre de professionnels de santé, etc.).
- **Communication citoyenne** : de nouveaux dispositifs de communication envers les citoyens devront mettre en avant les usages de la donnée de santé et permettre de partager les résultats de recherches nationales. **Un « Bilan Data » pourrait être proposé tous les ans via *Mon espace santé* pour illustrer la mobilisation des données de santé au niveau national et les principaux résultats de recherche.** Cette communication doit également donner des garanties en termes de sécurité de stockage des données de santé. Par ailleurs, il apparaît essentiel de mettre en avant les droits des citoyens et présenter le cadre réglementaire associé à l'utilisation secondaire des données de santé. **Cette nouvelle culture de la confiance ne pourra faire l'impasse de la mise en place d'un registre national d'opposition, permettant aux patients d'exercer leur droit d'opposition de manière simple.** Ce registre national pourrait être accessible directement via *Mon espace santé* afin de délivrer de manière éthique et transparente l'information.

## Conclusion

Les données de santé coproduites par les professionnels de santé et les patients contiennent des informations précieuses pour la recherche, l'innovation technologique et le pilotage de notre système de santé. Depuis quelques années, notre pays a réalisé des avancées significatives dans la structuration de ses données de santé : création du SNDS, mise en place du *Health Data Hub*, soutien aux entrepôts de données de santé hospitaliers. Le sujet est désormais omniprésent et porté par un fort écosystème d'acteurs. **Pour autant, force est de constater que les initiatives nationales des dernières années peinent à se déployer de manière effective et que notre pays prend du retard dans la mobilisation, l'utilisation et la valorisation de ses données de santé alors qu'il était encore pionnier il y a une décennie.** Les usages et les bénéfices de la donnée de santé demeurent encore trop peu connus de nos citoyens comme de nos professionnels de santé.

**L'année 2024 représente un tournant afin que notre système de santé prenne pleinement le pari des données de santé.** Cette note a pour ambition de contribuer à la stratégie nationale sur la réutilisation et l'ouverture des données de santé, prévue pour le premier semestre. Elle identifie quatre défis à relever pour que son déploiement soit effectif : des défis d'ordre technique, économique, juridique et citoyen et formule cinq propositions opérationnelles à même de relever ces défis. Le plein potentiel des données de santé ne sera en effet atteint qu'à la condition d'une meilleure interopérabilité entre les bases de données afin de faciliter leur mobilisation et leur appariement. Par ailleurs, sans la constitution de bases de données dédiées à la médecine de ville, un pan entier du parcours de soins demeure non-couvert. La création de *Data Pôles* chargés de définir une stratégie des données au sein des établissements de santé constitue un impératif afin de créer des réseaux territoriaux de données de santé. Néanmoins, la constitution de ces bases est dépendante de la création d'un cadre économique, afin d'inciter les producteurs au partage des données. Pour finir, il apparaît prioritaire d'introduire une véritable culture de la donnée chez nos professionnels de santé comme nos citoyens afin de les sensibiliser à leurs usages.

## Méthodologies de références

Méthodologie de référence (MR)	Description
<b>MR-001</b> : recherches dans le domaine de la santé avec recueil du consentement.	La MR-001 encadre les traitements comprenant des données de santé et présentant un caractère d'intérêt public, réalisés dans le cadre de recherches nécessitant le recueil du consentement de la personne concernée. La MR-001 concerne les recherches interventionnelles, les essais cliniques de médicaments et les recherches nécessitant la réalisation d'un examen des caractéristiques génétiques.
<b>MR-002</b> : études non-interventionnelles de performances concernant les dispositifs médicaux de diagnostic <i>in-vitro</i> .	La MR-002 concerne les études non interventionnelles de performances menées sur les dispositifs médicaux <i>in vitro</i> en vue de leur mise sur le marché.
<b>MR-003</b> : recherches dans le domaine de la santé sans recueil du consentement.	La MR-003 encadre les traitements comprenant des données de santé et présentant un caractère d'intérêt public, réalisés dans le cadre de recherches impliquant la personne humaine pour lesquelles les personnes concernées ont été informées et ne s'opposent pas à participer. La MR-003 concerne des recherches non interventionnelles et des essais cliniques de médicaments par grappe. Les patients doivent individuellement être informés.
<b>MR-004</b> : recherches n'impliquant pas la personne humaine, études et évaluations dans le domaine de la santé.	La MR-004 encadre les traitements de données à caractère personnel à des fins d'étude, évaluation ou recherche n'impliquant pas la personne humaine. Il s'agit surtout des études portant sur la réutilisation de données.
<b>MR-005</b> : études nécessitant l'accès aux données du PMSI et/ou des RPU par les établissements de santé et les fédérations hospitalières.	La MR-005 encadre l'accès par des établissements de santé et des fédérations hospitalières aux données du PMSI et aux RPU (Résumé de passage aux urgences), mises à disposition sur la plateforme sécurisée de l'Agence technique de l'information sur l'hospitalisation (ATIH). Les responsables de traitement sont tenus de documenter les projets menés dans le registre des activités de traitement. Les appariements à d'autres données sont interdits. Les traitements doivent être enregistrés auprès d'un répertoire public tenu par l'INDS.

Méthodologie de référence (MR)	Description
<b>MR-006</b> : études nécessitant l'accès aux données du PMSI par les industriels de santé.	La MR-006 encadre l'accès par les industriels de santé aux données du PMSI de l'ATIH (Agence technique de l'information sur l'hospitalisation). Les projets menés doivent être consignés dans le registre des activités de traitement. Aucun appariement avec d'autres données personnelles n'est permis. Les traitements doivent être enregistrés auprès d'un répertoire public tenu par l'INDS.  Les industriels doivent avoir recours à un bureau d'études ou un laboratoire de recherche ayant réalisé un engagement de conformité auprès de la CNIL. Ils doivent également réaliser un audit indépendant tous les 3 ans sur l'utilisation des données et le respect de l'interdiction des finalités interdites.
<b>MR-007</b> : recherches, études ou évaluations nécessitant l'accès aux données de la base principale du SNDS par les organismes agissant dans le cadre de leur mission d'intérêt public.	La MR-007 encadre l'accès par des organismes agissant dans le cadre de leur mission d'intérêt public aux données de la base principale du SNDS. Pour bénéficier de la MR-007, les responsables de traitement doivent respecter ses conditions et obtenir un avis favorable du CESREES (Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé).
<b>MR-008</b> : recherches, études ou évaluations nécessitant l'accès aux données de la base principale du SNDS par les organismes agissant dans le cadre de leurs intérêts légitimes.	La MR-008 encadre l'accès par les organismes agissant dans le cadre de leurs intérêts légitimes aux données de la base principale du SNDS.



## Fiches pays

## 1. FINLANDE

Généralités	
Autorité en charge de délivrer les permis.	<p><b>Finnish Social and Health Data Permit Authority (FinData).</b></p> <p>Une seule institution, <i>FinData</i>, <b>accorde les permis d'accès aux données, compile et traite les données</b> de diverses sources et en héberge une partie. <i>FinData</i> fait office de « <i>one-stop stop</i> » : un citoyen ou un chercheur n'a plus qu'un <b>seul interlocuteur</b> centralisateur pour obtenir des données de sources diverses.</p> <p><i>FinData</i> possède des données en propre (75 % de leurs données proviennent des registres de l'Institut Finlandais de la santé et du bien-être, l'Institut d'assurance sociale (Kela) et Statistics Finland.</p>
Plateforme nationale regroupant les bases de données.	<i>FinData</i> , créée en 2019. La plateforme est opérationnelle depuis janvier 2020.
Gouvernance / Paysage des différentes institutions de santé.	<p><i>FinData</i> est placé sous l'égide du ministère des affaires sociales et de la santé et dispose d'un directeur nommé par le ministère ainsi que d'un groupe de pilotage nommé par le ministère. <i>FinData</i> fonctionne en collaboration avec l'Institut national de la santé et du bien-être (<i>National Institute for Health and Welfare</i>), mais de façon autonome.</p> <p>Le médiateur parlementaire (Ombudsman) et le médiateur pour la protection des données supervisent les activités de <i>FinData</i>.</p>
Principales bases de données existantes.	<p>Les bases de données accessibles via <i>FinData</i> sont nombreuses et proviennent de multiples détenteurs de données nationales, notamment :</p> <ul style="list-style-type: none"> <li>• des entités publiques du secteur social et de la santé (des hôpitaux publics par exemple) ;</li> <li>• des registres d'organismes privés de services sociaux et de soins de santé (des laboratoires par exemple) ;</li> <li>• des clients enregistrés dans les services Kanta<sup>213</sup> (plateforme personnelle de données de santé).</li> </ul>

<sup>213</sup> Site internet de My Kanta : <https://www.kanta.fi/en/my-kanta-pages>.

Généralités	
Parcours et délais d'accès aux données pour les chercheurs.	<p>Lorsque <i>FinData</i> reçoit une demande de permis de la part d'un chercheur, elle analyse la demande et vérifie la disponibilité des données demandées ainsi que les coûts affiliés à leur obtention puis à leur traitement. Une fois ces informations réunies, <i>FinData</i> octroie le permis d'accès.</p> <p><i>FinData</i> compile et pseudonymise les données auxquelles aura accès le chercheur ou le citoyen. <i>FinData</i> met ensuite ces données à disposition dans la plateforme indépendante et sécurisée nommée Kapseli.</p> <p><b>Le délai d'obtention d'un permis varie entre 1-3 mois.</b></p>
Coûts associés à la demande de permis.	<p>Une demande de permis s'accompagne de frais qui comprennent :</p> <ul style="list-style-type: none"> <li>• frais de demande de permis ;</li> <li>• coûts encourus pour l'extraction et la livraison des données ;</li> <li>• heures de travail liées au traitement et à la pseudonymisation des données ;</li> <li>• frais liés à l'entretien de Kapseli.</li> </ul> <p>Les frais pour l'obtention d'un permis varient entre 250 € (dans le cadre d'une thèse) et 3 000 € (pour une demande qui nécessite d'aller extraire des données auprès de 15 entités ou plus).</p>

## 2. DANEMARK

Généralités	
Autorité en charge de délivrer les permis.	<p>Le <b>Danish Health Data Authority</b>, créé en 2015, collecte, stocke et gère les données de santé. À ce titre, elle est l'<b>autorité responsable des registres nationaux</b> (<i>Danish Clinical Registries</i>) qui regroupent des données relatives à certaines pathologies clés ou au système de santé danois et <b>développe des standards</b> pour uniformiser les données qu'elle collectionne.</p> <p>Le <i>Danish Health Data Authority</i> est l'organe chargé de délivrer des permis aux chercheurs qui font une demande d'accès à ces données.</p>
Plateforme nationale regroupant les bases de données.	Le <i>Research Health Data Gateway</i> (en cours de construction – voir ci-après pour plus d'informations).

Généralités	
Principales bases de données existantes.	<p>Il y a plusieurs entités, en plus du <i>Danish Health Data Authority</i>, qui possèdent des bases de données, notamment :</p> <ul style="list-style-type: none"> <li>• <i>Statistics Denmark</i>, autorité gouvernementale chargée de collecter et de publier les statistiques sur la société danoise.</li> <li>• <i>The Danish Clinical Registries</i>.</li> <li>• <i>National Genome Center</i>, qui a pour mission principale de développer de meilleurs diagnostics et des traitements plus ciblés pour les patients à l'aide du séquençage du génome entier (WGS).</li> <li>• <i>Danish National Biobank</i>, créée en 1982, elle possède plus de 25 millions d'échantillons biologiques (environ 1 million d'échantillons supplémentaires chaque année) issues des bases de données biologiques régionales.</li> <li>• <i>Regional biobanks</i>, centres locaux où sont enregistrées les données concernant les échantillons biologiques des patients de la zone géographique alentour.</li> </ul>
Parcours et délais d'accès aux données pour les chercheurs.	<p>Pour accéder aux données de santé, les chercheurs doivent faire une demande auprès du <i>Research Services</i> du <i>Danish Health Data Authority</i> (DHDA). Si les critères sont remplis, le <i>Research Services</i> du DHDA donnera accès à une plateforme virtuelle indépendante et sécurisée (<i>The Secure Research Platform</i>) et le chercheur pourra alors obtenir les données (préalablement pseudonymisées) nécessaires à sa recherche. Les données des registres nationaux de santé ne peuvent être distribuées que par l'intermédiaire de cette plateforme.</p> <p>Le délai d'obtention des données est d'environ 30 jours ouvrés.</p>
Coûts associés à la demande de permis.	<p>Une facturation des services du <i>Research Services</i> du <i>Danish Health Data Authority</i> est demandée aux utilisateurs afin de financer les coûts liés au traitement des demandes ainsi que pour les conseils et l'assistance quant à l'utilisation des données une fois transmises au chercheur.</p> <p>Les coûts couvrent :</p> <ul style="list-style-type: none"> <li>• le traitement de la demande ;</li> <li>• l'octroi de l'accès aux données sur la <i>Secure Research Platform</i> ;</li> <li>• l'extraction des données ;</li> <li>• les conseils et l'assistance pour les demandes adressées au <i>Research Services</i> après la livraison des données ;</li> <li>• la mise à jour des données et la modification de l'accès aux données ;</li> <li>• les frais liés à la gestion de la <i>Secure Research Platform</i>.</li> </ul> <p>La facturation se fait sur une base mensuelle au taux horaire en vigueur. La facture indique le nombre d'heures consacrées à chaque projet.</p>

### 3. ROYAUME-UNI

Généralités	
Autorité en charge de délivrer les permis.	<p><b>Health Data Research UK</b>, créé en 2018, est l'institut national de données sur la santé qui concerne le Royaume-Uni. Il encadre le processus et délivre les autorisations pour les projets de la NHS.</p> <p>Le <i>Health Data Research UK</i> a accès à un large éventail de données de santé provenant du NHS, d'universités, d'instituts de recherche et d'organisations caritatives, d'entreprises privées et de technologies portables et mobiles.</p>
Plateforme(s) nationale(s) regroupant les bases de données.	<p>À l'initiative du <i>Health Data Research UK</i>, le <i>Health Data Research Gateway</i>, créé en 2020, est le point d'entrée pour les chercheurs afin de découvrir et de demander l'accès aux ensembles de données de santé du Royaume-Uni. Le <i>Gateway</i> répertorie des informations sur chaque ensemble de données (telles que la description, la taille de la population contenue dans l'ensemble de données et la base juridique de l'accès).</p> <p>Le <i>NHS Digital</i> possède une plateforme qui agit comme un « entrepôt de données » et qui y donne accès (sur demande) de façon anonymisée : le <b>Secondary Use Services</b>.</p>
Principales bases de données existantes.	<p>Le <i>Gateway</i> répertorie plus de 750 ensembles de données<sup>214</sup>, auxquels les utilisateurs peuvent demander l'accès en en faisant la demande.</p> <p>Le <i>NHS Digital</i> possède aussi plusieurs collections de données<sup>215</sup> couvrant de nombreux aspects de la santé et des soins sociaux, collectées auprès d'un large éventail de services du NHS, d'autorités locales et d'organisations du secteur indépendant.</p>
Parcours et délais d'accès aux données pour les chercheurs.	<p>Deux parcours d'accès aux données distincts pour les chercheurs :</p> <ul style="list-style-type: none"> <li>• <b>par le Health Data Research Gateway</b> : il agit comme un portail permettant la découverte d'ensembles de données auxquels les chercheurs peuvent demander l'accès.</li> <li>• <b>par le SUS</b> : le <i>Secondary Uses Service</i> (SUS) est géré et développé par le <i>NHS Digital</i>. Le SUS est dépositaire de données. Il est en mesure de partager des données aux utilisateurs agréés. Toute organisation ou particulier qui souhaite utiliser des données doit au préalable faire une demande auprès du <i>Data Access Request Service</i> (DARS). Lorsque la demande DARS est approuvée, les données seront mises à disposition via le <i>Data Access Environment</i> (DAE). Une nouvelle initiative du NHS a été développée en 2023 : le <i>Secure Data Environments</i> (SDE) <i>Service</i><sup>216</sup>. Modifiant l'accès aux données de santé du <i>NHS England</i>, il s'agit d'un <b>moyen par défaut</b> d'accès aux données à des fins de recherche et d'analyse.</li> </ul>

<sup>214</sup> Site internet du [Health Data Research Innovation Gateway](#).

<sup>215</sup> "Data collections", NHS Digital.

<sup>216</sup> "The NHS England Secure Data Environment", NHS Digital.

Généralités	
Facturation / coûts associés.	<p><i>NHS Digital</i> est financé par des fonds publics. Les données ne sont pas facturées mais des frais sont appliqués pour couvrir le coût du traitement et des services proposés. Les frais sont calculés sur la base des éléments suivants : le type de demande ; le volume de données demandées, y compris les liens entre les données ; la fréquence de diffusion des données.</p> <p>Une fois la demande de données soumise par un chercheur, le <i>NHS Digital</i> transmet plus de détails concernant le coût associé à la demande de données. Des exemples de coûts par projet sont transmis par le <i>NHS Digital</i><sup>217</sup>.</p>

L'auteur de cette note tient à remercier **Pauline Hodille**, rapporteure de cette note, pour son aide précieuse et ses conseils durant l'élaboration de ce travail ainsi que le Professeur **Pierre-Antoine Gourraud** et **Ali Saïb** pour leur relecture attentive. Une pensée particulière est portée à l'ensemble de l'équipe de l'Institut Montaigne ayant permis sa réalisation, notamment **Polly Lefèvre**, **Margaux Tellier-Poulain**, **Manon Palies** et **Lou Vincent**.

Enfin, ce travail n'aurait pas été possible sans l'expertise des personnes auditionnées ou consultées durant la réalisation de ce travail :

- **Audrey Acloque**, *DPO*, Unicancer
- **Jean-Baptiste Angeloglou**, *General Manager*, PSIH
- **Jean-Marc Aubert**, vice-président Healthcare, Iqvia
- **Ségolène Ayme**, présidente Conseil Scientifique Consultatif, Health Data Hub
- **Sophie Bauer**, présidente, Syndicat des Médecins Libéraux
- **Caroline Beaudon**, *Strategic Consulting Manager*, Roche Diagnostics
- **Benoît Bélanger**, directeur des opérations, DrData
- **Nesrine Benyahia**, docteur en droit, présidente et fondatrice de DrData, membre associé de l'Institut Droit et Santé (Université Paris Cité-Inserm)
- **Pierre-Henri Bertoye**, directeur Prospective Réglementaire, référentiels et formation, Unicancer
- **Antoine Bloch**, directeur général France, bioMérieux
- **Laurent Borella**, directeur santé, Malakoff Humanis
- **Stéphanie du Boucher**, responsable affaires publiques, Roche Pharma
- **Pierre-Yves Brossard**, *Healthcare Strategist*, Doctolib
- **Stéphane Camon**, directeur du pilotage économique et de la data, Allianz
- **Alice de Chalendar**, adjointe du *leader* d'écosystème Ma Santé, Allianz

<sup>217</sup> "Data Access Request Service (DARS) charges from 2020/2021", *NHS Digital*.

- **Remy Choquet**, *Head of Evidence Generation*, Medical Data platforms and Data sharing, Roche Pharma
- **Hugo Crochet**, directeur des SI et des Data, Centre Léon Bérard, co-pilote du projet DeepSarc
- **Thomas Croisier**, *Partner*, en charge de l'écosystème Santé pour l'Europe, Kearney
- **Arthur Dauphin**, chargé de mission numérique en santé, France Assos Santé
- **Arthur Delapalme**, président et directeur général, Codoc
- **Cécile Delval**, directrice *Real World and Clinical Evidence*, Air Liquide Healthcare
- **Mélodie Deneuve**, directrice de la communication, Mutualité Française
- **Pierre Dubreuil**, directeur de projets, Délégation au numérique en santé
- **Françoise Ducimetière**, cheffe de projet, DeepSarc, Centre Léon Bérard
- **Thomas Duval**, *CEO*, Sancare
- **Nicolas Eiferman**, directeur général France Bénélux, Intersystems
- **Anne-Sophie Etzol**, responsable des relations institutionnelles, Inserm
- **Céline Falco**, coordinatrice CONTACT (EDS SOS Médecins France), présidente Fondation April
- **Vincent Figureau**, responsable des partenariats et du mécénat, Mutualité Française
- **Pierre Fillard**, directeur technique et scientifique, fondateur, Therapixel
- **Christophe Gauer**, directeur de cabinet du DG – Directeur des Affaires Publiques, Allianz
- **Julien Guérin**, directeur des données, Institut Curie
- **Caroline Guillot**, directrice citoyenne, Health Data Hub
- **Hélène Guimiot-Breaud**, cheffe du service de la santé, CNIL
- **Alice Herreye**, directrice des Affaires Juridiques, Roche Pharma
- **Mariuca de Hillerin**, responsable affaires publiques, Arkhn

- **Thierry Houselstein**, directeur médical, MACSF
- **Harold Juillet**, directeur général, Oracle Health
- **Aki Kuivalainen**, président fondateur, Predicell
- **Christophe Lapiere**, directeur des systèmes d'information en santé, Mutualité Française
- **Franck Le Ouay**, président fondateur, Lifen
- **Martin Le Roux-Cauché**, directeur Affaires publiques Santé, Air Liquide
- **Patrick Lethenet**, *Senior VP*, R&D – IT Solutions & System Development, bioMérieux
- **Franck Lethimonnier**, directeur Institut Thématique Technologie pour la Santé, Inserm
- **Stéphane Locret**, directeur recherche & enseignement, Ramsay Santé
- **Thomas London**, directeur associé, Mc Kinsey & Company
- **Bertrand Lukacs**, vice-président Conseil Scientifique Consultatif Health Data Hub
- **Emeric Lemaire**, *CEO*, Arkhn
- **Carine Malcus**, *Senior Director*, Global clinical and regulatory affairs support office, and global bioethics compliance, bioMérieux
- **Antoine Malone**, responsable Pôle prospective, Europe et International, Fédération hospitalière de France
- **Anne-Laure Martin**, directrice Data & Partenariat, Unicancer
- **Sébastien Massart**, *Chief Corporate Strategist*, Dassault Systèmes
- **Jean-Christophe Masseron**, président, SOS Médecins France
- **Loïck Menvielle**, directeur de la chaire Management in Innovative Health, EDHEC Business School
- **Etienne Minvielle**, directeur du Centre de Recherche en gestion, Polytechnique, Institut Gustave Roussy
- **Arnaud Morigault**, associé, Eurogroup Consulting
- **Pierre Moustial**, président fondateur, Lauxera Partners
- **Dinh-Phong Nguyen**, responsable de la veille et de la prospective, Agence de l'innovation en santé
- **Sophie Ollivier**, *R&D Chief Data Officer*, Servier

- **Guillemette Pardoux**, responsable communication, France Cohortes
- **Aymeric Perchant**, directeur de projets, Délégation au Numérique en Santé (DNS)
- **Olivier Perche**, responsable Affaires Publiques, Roche Diagnostics
- **David Petauton**, directeur de la communication, Agence du numérique en santé
- **Dominique Pon**, directeur général, La Poste Santé & Autonomie
- **Christian Portafax**, responsable du service des données et système d'information de santé (Direction Santé), Mutualité Française
- **Yves Raisin**, *Global Data Privacy Officer*, bioMérieux
- **Gérard Raymond**, président, France Assos Santé
- **Raphaële Renard-Penna**, professeur des Universités – praticien hospitalier, radiologue, Faculté Sorbonne Université, Hôpital Pitié Salpêtrière APHP, Projet DAICAP HDH
- **Grégoire Rey**, directeur, France Cohortes
- **Matthieu Robain**, directeur scientifique, Direction Data, Unicancer
- **Ulysse Rodts**, cofondateur, Kanopy Med
- **Raphaël Sigogne**, responsable data, Paris Saclay Cancer Cluster
- **Daniel Szeftel**, président co-fondateur, Sêmeia
- **Éric Tanneau**, secrétaire général, Syndicat des Médecins Libéraux
- **Antoine Tesnière**, directeur général, PariSanté Campus
- **Isabelle Tongio**, *Director – Publics and Government Affairs*, bioMérieux
- **Vincent Trely**, président fondateur, Association pour la Sécurité des SI de Santé (APSSIS)
- **Pierre Vaysse**, *Leader écosystème Ma Santé* et directeur technique Collectives, Allianz
- **Damien Vergé**, directeur de la stratégie, des études et de la statistique, Cnam

Les opinions exprimées dans ce rapport n'engagent ni les personnes précédemment citées ni les institutions qu'elles représentent.

- **ANS** : Agence Nationale du Numérique en Santé
- **ANSES** : Agence Nationale de Sécurité Sanitaire Alimentaire Nationale
- **ANSM** : Agence Nationale de Sécurité du Médicament et des Produits de Santé
- **AP-HP** : Assistance Publique – Hôpitaux de Paris
- **ARS** : Agence Régionale de Santé
- **ATIH** : Agence Technique de l'Information sur l'Hospitalisation
- **BCMD** : Base de données sur les Causes Médicales de Décès
- **CESREES** : Comité Éthique et Scientifique pour les Recherches, les Études et les Évaluations dans le Domaine de Santé
- **CHR** : Centre Hospitalier Régional
- **CHU** : Centre Hospitalier Universitaire
- **CIM 10** : Classification Internationale des Maladies 10
- **CISP** : Classification Internationale des Soins Primaires
- **CNAM** : Caisse Nationale d'Assurance Maladie
- **CNIL** : Commission Nationale de l'Informatique et des Libertés
- **CNSA** : Caisse Nationale de Solidarité pour l'Autonomie
- **CPP** : Comité de Protection des Personnes
- **CPTS** : Communauté professionnelle territoriale de santé
- **DNS** : Délégation au Numérique en Santé
- **DPI** : Dossier Patient Informatisé
- **DPO** : Délégué à la protection des données
- **DREES** : Direction de la Recherche, des Études, de l'Évaluation et des Statistiques
- **EDS** : Entrepôt de Données de Santé
- **EHDS** : *European Health Data Space*
- **FHF** : Fédération Hospitalière de France
- **FHIR** : *Fast Healthcare Interoperable Resources*
- **GCS** : Groupement de Coopération Sanitaire
- **GRADeS** : Groupement Régional d'Appui au Développement de la e-Santé (GRADeS)

- **HAS** : Haute Autorité de Santé
- **HDH** : *Health Data Hub*
- **HIE** : *Health Information Exchange*
- **IA** : Intelligence Artificielle
- **IDS** : Institut des Données de Santé
- **INCa** : Institut National du Cancer
- **INDS** : Institut National des Données de Santé
- **INSERM** : Institut National de la Santé et de la Recherche Médicale
- **LOINC** : *Logical Observation identifiers Names & Codes*
- **MDPH** : Maison Départementale des Personnes Handicapées
- **MES** : Mon Espace Santé
- **MR** : Méthodologie de Référence
- **NHS** : *National Health Service*
- **OCDE** : Organisation de Coopération et de Développement Économiques
- **OMOP** : *Observational Medical Outcomes Partnership*
- **ONC** : *Office of the National Coordinator for health information technology*
- **PDS** : Plateforme des Données de Santé
- **PMSI** : Programme de Médicalisation des Systèmes d'Information
- **RGPD** : Règlement Général sur la Protection des Données
- **RIPH** : Recherches impliquant la personne humaine
- **RNIPH** : Recherches n'impliquant pas la personne humaine
- **SNDS** : Système National des Données de Santé
- **SNIIRAM** : Système National d'Information Inter-Régimes de l'Assurance Maladie

ABB France	Crédit Agricole	KPMG S.A.	Rokos Capital
AbbVie	D'Angelin & Co.Ltd	Kyndryl	Management
Accenture	Dassault Systèmes	La Banque Postale	Rothschild & Co
Accuracy	Deloitte	La Compagnie	RTE
Adeo	De Pardieu Brocas	Fruitière	Safran
ADIT	Maffei	Linedata Services	Sanofi
Air Liquide	ECL Group	Lloyds Europe	SAP France
Airbus	Edenred	L'Oréal	Schneider Electric
Allen & Overy	EDF	Loxam	ServiceNow
Allianz	EDHEC Business	LVMH - Moët-Hennessy	Servier
Amazon	School	- Louis Vuitton	SGS
Amber Capital	Ekimetrics France	M.Charraire	SIER Constructeur
Amundi	Engie	MACSF	SNCF
Antidox	EQT	Mazars	SNCF Réseau
Antin Infrastructure	ESL & Network	Média-Participations	Sodexo
Partners	Ethique &	Mediobanca	SPVIE
ArchiMed	Développement	Mercer	SUEZ
Ardian	Eurogroup Consulting	Meridiam	Tecnet Participations
Arqus	FGS Global Europe	Microsoft France	SARL
AstraZeneca	Getlink	Mitsubishi France	Teneo
August Debouzy	Gide Loyrette Nouel	S.A.S	The Boston Consulting
AXA	Google	Moelis & Company	Group
Bain & Company	Groupama	Moody's France	Tilder
France	Groupe Bel	Morgan Stanley	Tofane
Baker & McKenzie	Groupe M6	Natixis	TotalEnergies
BearingPoint	Groupe Orange	Natural Grass	Unicancer
Bessé	Hameur et Cie	Naval Group	Veolia
BNP Paribas	Henner	Nestlé	Verian
Bolloré	Hitachi Energy France	OCIRP	Verlingue
Bouygues	Howden France	ODDO BHF	VINCI
Bristol Myers Squibb	HSBC Continental	Oliver Wyman	Vivendi
Brousse Vergez	Europe	Ondra Partners	Wakam
Brunswick	IBM France	Onet	Wavestone
Capgemini	IFPASS	Optigestion	Wendel
Capital Group	Incyte Biosciences	Orano	White & Case
CAREIT	France	PAI Partners	Willis Towers Watson
Carrefour	Inkarn	Pelham Media	France
Casino	Institut Mérieux	Pergamon	Zurich
Chubb	International SOS	Polytane	
CIS	Interparfums	Prodware	
Cisco Systems France	Intuitive Surgical	Publicis	
Clariane	Ionis Education Group	PwC France & Maghreb	
Clifford Chance	iQo	Raise	
CNP Assurances	ISR	RATP	
Cohen Amir-Aslani	Jeantet Associés	RELX Group	
Compagnie Plastic	Jolt Capital	Renault	
Omnium	Katalyse	Ricol Lasteyrie	
Conseil supérieur du notariat	Kea & Partners	Rivolier	
	Kearney	Roche	

  
 Institut Montaigne  
 59 rue La Boétie, 75008 Paris  
 Tél. +33 (0)1 53 89 05 60  
[institutmontaigne.org](http://institutmontaigne.org)

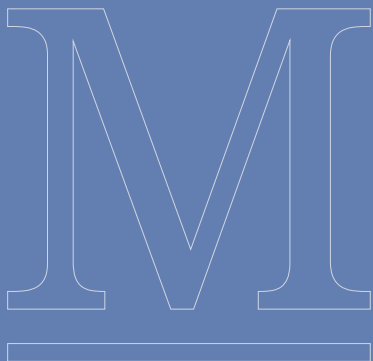
Imprimé en France  
 Dépôt légal : février 2024  
 ISSN : 1771-6756

Les professionnels de santé et les patients coproduisent chaque jour de nombreuses données qui viennent alimenter les dossiers médicaux, les bases de remboursement de l'Assurance maladie ou encore les systèmes d'informations de nos établissements de santé. Ces données sont collectées initialement dans le cadre du soin : on parle alors d'utilisation primaire de la donnée de santé.

Il est possible de donner une seconde vie à la donnée de santé à travers son utilisation secondaire qui ouvre un immense champ d'innovation et de progrès pour notre système de santé.

Accélération de la recherche, dépistages précoces, aide au diagnostic, santé prédictive, amélioration de l'accès aux soins, personnalisation de la prise en charge des patients, évaluation des politiques publiques : le potentiel des données de santé promet des avancées spectaculaires pour améliorer la santé de tous.

Dans le cadre d'une nouvelle stratégie nationale de l'usage des données de santé, cette note fait ressortir quatre grands défis, d'ordre technique, économique, juridique et citoyen, et propose cinq recommandations concrètes pour les relever. Libérons le potentiel des données de santé pour permettre à notre système de santé de se transformer, d'innover et d'adopter un virage résolument préventif.



10 €

ISSN : 1771-6756

NAC2402-01