



**RÉPUBLIQUE
FRANÇAISE**

*Liberté
Égalité
Fraternité*

Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire des données de santé



Tome 2 : Annexes

Jérôme MARCHAND-ARVIER

Pr Stéphanie ALLASSONNIERE

Aymeril HOANG

Dr Anne-Sophie JANNOT

Rapporteurs

Emilie FAUCHIER-MAGNAN (IGAS), Julien FRADEL (Conseil d'État)

5 décembre 2023

SOMMAIRE

ANNEXE 1 : PANORAMA DES ENTREPOTS DE DONNEES DE SANTE	4
ANNEXE 2 : LE CADRE JURIDIQUE DE L'ACCES AUX DONNEES DE SANTE A DES FINS DE RECHERCHE	19
ANNEXE 3 : COMPARAISONS INTERNATIONALES EN MATIERE D'ACCES AUX DONNEES DE SANTE	31
ANNEXE 4 : LE HEALTH DATA HUB : MISSIONS ET REALISATIONS	35

ANNEXE 1 : Panorama des entrepôts de données de santé

Table des matières

ANNEXE 1 : PANORAMA DES ENTREPOTS DE DONNEES DE SANTE	4
1 UN CADRE JURIDIQUE PARTICULIEREMENT EXIGEANT, DONT LE PERIMETRE RESTE A CLARIFIER AUPRES DES ACTEURS	5
2 UNE DYNAMIQUE DE CREATION D'ENTREPOTS DE DONNEES PAR DES ACTEURS VARIES, PUBLICS COMME PRIVES	6
3 LES ENTREPOTS DE DONNEES DE SANTE HOSPITALIERS : UN POTENTIEL IMPORTANT QUI DOIT ETRE CONFORTE PAR UNE MEILLEURE INTEROPERABILITE	8
3.1 DES DONNEES MEDICALES RICHES MAIS HETEROGENES	8
3.2 DES USAGES MULTIPLES	10
3.3 DE PREMIERES INITIATIVES POUR RENFORCER L'INTEROPERABILITE ENTRE LES ENTREPOTS	13
3.4 UNE FRAGILITE : UN MODELE DE FINANCEMENT NON PERENNE	14
3.4.1 <i>Des recettes encore difficiles à estimer dans un écosystème en développement et dans un contexte juridique évolutif</i>	<i>14</i>
3.4.2 <i>De récents financements des entrepôts de données de santé par appels à projets</i>	<i>16</i>
3.4.3 <i>Des coûts des entrepôts récemment objectivés</i>	<i>17</i>

En réponse à la lettre de mission qui l'invitait « à « cartographier la dynamique actuellement à l'œuvre de grands entrepôts de données de santé », la mission a plus particulièrement approfondi la thématique des entrepôts de données de santé. Elle s'est appuyée sur des données transmises par la commission nationale de l'informatique et des libertés (CNIL), sur un questionnaire transmis aux agences régionales de santé, sur des entretiens avec des porteurs d'entrepôts de données de santé, mais aussi sur une riche documentation (notamment un récent rapport de la haute autorité de santé¹ ainsi que les travaux du comité stratégique des données de santé).

Ce panorama des entrepôts de données de santé illustre le foisonnement d'initiatives à l'œuvre dans cet écosystème en cours de maturation.

¹ « Entrepôts de données de santé hospitaliers en France - Quel potentiel pour la Haute Autorité de santé », Haute autorité de santé, 2022

1 Un cadre juridique particulièrement exigeant, dont le périmètre reste à clarifier auprès des acteurs

Selon la haute autorité de santé, « les entrepôts de données de santé (EDS) désignent la mise en commun des données d'un ou plusieurs systèmes d'information médicaux, sous un format homogène pour des réutilisations à des fins de pilotage, de recherche ou dans le cadre des soins ». Cette définition de nature opérationnelle caractérise les entrepôts de données de santé comme une agrégation structurée de données issues de systèmes d'informations médicaux, à des fins d'utilisation secondaire des données.

Dans son référentiel dédié aux entrepôts de données de santé adopté en 2021, la CNIL a esquissé une approche juridique plus extensive de la notion d'entrepôts de données de santé. Les entrepôts de données de santé y sont en effet définis par leurs finalités, parmi lesquelles figure la réalisation de traitements à des fins de recherche, d'étude ou d'évaluation dans le domaine de la santé. Une lecture stricte du texte laisse ainsi penser que l'ensemble des traitements pour réutiliser des données à des fins de recherche, d'étude ou d'évaluation doit se conformer au référentiel, à l'exception des traitements qui ont obtenu une autorisation explicite des personnes concernées.

Les remontées sur les entrepôts de données auprès de la mission attestent ainsi d'une certaine confusion sur la notion d'entrepôt de données de santé. Certains projets signalés à la mission relèvent ainsi d'outils de pilotage gérés par les agences ou autres opérateurs.

La CNIL précise toutefois par ailleurs qu'un traitement de données est qualifié d'entrepôts de données de santé à partir d'un faisceau d'indices (voir détail dans l'encadré ci-après)

Définition d'un entrepôt de données de santé selon la CNIL

Un traitement de données de santé peut être **qualifié d'entrepôt grâce à un faisceau d'indices**. Les éléments suivants peuvent, entre autres, être pris en compte pour la qualification du traitement en « entrepôt » :

- Origine variée des données (données relatives à la prise en charge médicale du patient, données socio-démographiques, données issues de précédentes recherches, données cliniques associées à des échantillons biologiques etc.) ;
- Base de données généraliste (par ex : dossiers médicaux d'un établissement) ou thématique (par ex : pathologie ou groupe de pathologies ciblées) ;
- Durée de conservation sur une période prolongée (au moins 10 ans) ;
- Base de données le plus souvent alimentée au fil de l'eau selon des critères prédéterminés (par ex : dossier médical des personnes suivies) ;
- Données destinées à la réalisation de traitements ultérieurs.

Les entrepôts de données sont créés principalement pour **collecter et disposer de données à large échelle** : données relatives à la prise en charge médicale du patient, données socio-démographiques, regroupement de données issues de précédentes recherches, registre de pathologies, etc. Ces données **sont ensuite réutilisées**, selon les entrepôts, à des fins de recherche, d'étude, ou encore par exemple pour la production d'indicateurs.

Les registres entrent en général dans la définition des entrepôts de données de santé. En revanche, les entrepôts de données sont à **distinguer des projets de recherche** dont les finalités sont clairement définies dans un protocole et qui répondent à une question scientifique précisément identifiée en amont du traitement. C'est notamment le cas des cohortes, qui constituent des projets de recherche, puisqu'elles sont alimentées par des enquêtes et questionnaires administrés régulièrement aux participants, voire des suivis médicaux.

Cette clarification du périmètre des entrepôts de données de santé mériterait d'être explicitée dans le référentiel de la CNIL, pour limiter les confusions. Certains acteurs rencontrés par la mission considéraient en effet que les cohortes relevaient du champ du référentiel. Mais, même avec cette clarification, on peut estimer que la définition actuelle des entrepôts de données de santé est trop large, et qu'il conviendrait, en s'inspirant de la définition donnée par la HAS, de réserver la qualification d'entrepôt de données de santé au regroupement de données, au sein d'un établissement de santé, issues d'un ou plusieurs systèmes d'information médicaux, pour des réutilisations à des fins de pilotage, de recherche ou dans le cadre des soins.

Par ailleurs, si la réalisation d'un référentiel par la CNIL a été appréciée par les responsables d'entrepôts de données de santé rencontrés, **son niveau d'exigence est toutefois jugé très élevé, d'autant qu'il se cumule avec d'autres réglementations**, par exemple celle des hébergeurs de données de santé ou celui du SNDS, ce qui rend le cadre juridique et technique très complexe.

S'agissant des registres, le respect de ce référentiel paraît particulièrement ambitieux et suppose de profondes évolutions (tant en termes de gouvernance des bases de données que de sécurité), alors même que le respect du RGPD constitue d'ores et déjà un défi important pour nombre de ces structures.

Ainsi, **un bilan de l'application du référentiel auprès des acteurs concernés mériterait d'être dressé, deux ans après son entrée en vigueur, pour ajuster le cas échéant le cadre juridique** (dans le respect du RGPD), ou son périmètre (l'exclusion d'un certain nombre de registres mériterait d'être étudiée), en distinguant les exigences minimales (probablement en-deçà du référentiel actuel) et les « meilleures pratiques ».

2 Une dynamique de création d'entrepôts de données par des acteurs variés, publics comme privés

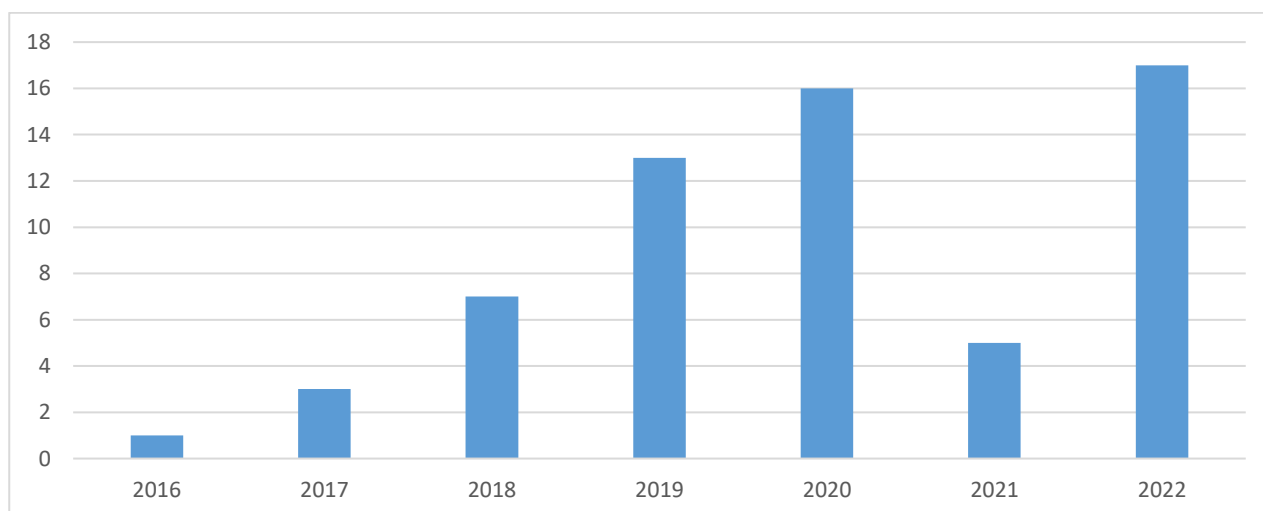
Comme en attestent les données transmises par la CNIL, de nombreux entrepôts de données de santé ont été créés ces dernières années, par des acteurs variés, privés comme publics.

Les données exploitées par la mission concernent les entrepôts de données de santé autorisés par la CNIL². Il ne s'agit pas d'un recensement exhaustif, car certains entrepôts de données de santé ne nécessitent aucune formalité auprès de la CNIL³. Il permet toutefois d'apprécier la dynamique à l'œuvre, la variété des objets des entrepôts autorisés et la nature des acteurs porteurs. Ces éléments devraient être complétés par une cartographie des entrepôts de données de santé en cours de réalisation par la CNIL.

La CNIL a autorisé 63 entrepôts de données entre 2017 et mi 2023. Le nombre total d'entrepôts est toutefois inférieur à 63, certains porteurs de projets ayant sollicité plusieurs autorisations, avec des demandes de modification d'une autorisation initiale. **Une faible proportion des autorisations concerne un entrepôt chaîné avec le SNDS** (9 autorisations relevant de 3 acteurs, l'INCA, Clinytix et Agoria).

Le graphique ci-après illustre la hausse progressive du nombre de dépôts d'autorisation (à l'exception de la période COVID). Cette dynamique, en cours et à venir, a été encouragée par l'appel à projet pour le financement des EDS lancée par l'Etat, dotée de 75 millions d'euros et qui soutient des établissements de santé ou des groupements d'établissements pour créer ou renforcer leur entrepôt (voir partie 3.4.2 *infra*).

Tableau 1 : Nombre d'autorisations d'EDS accordées par la CNIL, par année de dépôt



Source : Données CNIL exploitées par la mission

Par ailleurs, les demandes d'autorisation concernent des **entrepôts de nature et de proportions très variées** : nombreux entrepôts hospitaliers, généralistes, ou spécialisés sur certaines pathologies (cardiologie, VIH, diabète, certaines tumeurs...) mais aussi des registres ou des entrepôts destinés à produire des indicateurs de pilotage.

² Les données publiées par la CNIL recensent 36 déclarations de conformité avec le référentiel des entrepôts de données de santé, dont 5 émanent d'établissements de santé publics. Un certain nombre de déclarations relèvent d'erreurs d'interprétations (par exemple des entrepôts de stockage de matériel Industriel), c'est pourquoi ces données n'ont pas été intégrées dans l'analyse de la mission.

³ Il s'agit en particulier des traitements de données à caractère personnel mis en œuvre lorsque la personne a donné son consentement explicite à cette fin.

La diversité des projets se reflète dans la **variété des porteurs de projets** :

- **Les entrepôts de données de santé autorisés émanent en majorité (55 %) des établissements de santé.** Mi 2023, 27 entrepôts de données de santé hospitaliers étaient autorisés par la CNIL⁴. La quasi-totalité des CHU a ainsi développé un entrepôt de données, mais le développement d'EDS concerne également des établissements non lucratifs (principalement dans le domaine de la cancérologie), ainsi que, plus récemment, depuis 2021, par des établissements lucratifs.
- **Près de 30 % des autorisations d'entrepôts de données de santé émane d'entreprises privées** le plus souvent spécialisées dans l'analyse de données, et relevant parfois de regroupements d'acteurs (Clinytix, Agoria, IQVIA, Openhealth, Compugroup,..).
- Enfin, le reste des autorisations relève d'acteurs très variés : sociétés savantes, réseaux de soins, structures académiques (par ex INSERM), ARS ou groupement régional d'appui à la e-santé, fédération d'acteurs de santé sur un territoire, association médico-sociale, etc.

Tableau 2 : Nombre d'autorisations selon la nature du responsable de traitement

Etablissements de santé public	24
Etablissements de santé privés non lucratif	9
Etablissements de santé privés lucratif	2
Total établissements de santé	35
Entreprises privées	18
Sociétés savantes	3
Autres (fédération, réseau, association, structure académique...)	8
Total général	64*

Source : Données CNIL exploitées par la mission

* Un EDS, porté conjointement par un établissement de santé et une entreprise, est comptabilisé deux fois.

3 Les entrepôts de données de santé hospitaliers : un potentiel important qui doit être conforté par une meilleure interopérabilité

3.1 Des données médicales riches mais hétérogènes

Au sein du paysage français des bases de données en santé, les entrepôts de données de santé hospitaliers présentent deux avantages.

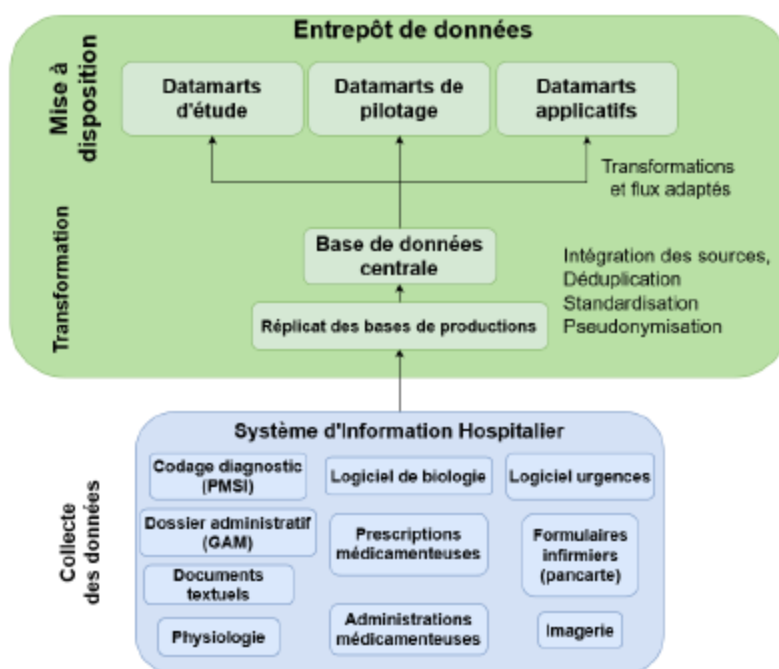
⁴ La donnée diffère des 35 autorisations présentées dans le tableau ci-dessus, car on a retiré les autorisations multiples pour un même entrepôt.

En premier lieu, **ils intègrent de nombreuses données médicales absentes du SNDS**. Ainsi, d'après le rapport de la haute autorité de santé⁵, tous les EDS intègrent les données administratives des patients, les diagnostics et procédures codées, la biologie et les données textuelles. Les données sur le circuit du médicament sont également très souvent intégrées. C'est plus rarement le cas pour des types de données complexes et volumineuses comme la réanimation, l'imagerie ou la génomique.

Une proportion importante de ces données est peu structurée, figurant dans des données de texte, ce qui suscite de nombreux projets pour exploiter de manière plus ou moins automatique ces données de texte.

En second lieu, **les données des entrepôts de données de santé sont produites en routine**, à partir des systèmes d'information hospitaliers (voir graphique ci-dessous), après une étape de pseudonymisation. Les EDS recouvrent ainsi des données médicales concernant un grand nombre de patients, largement supérieur aux volumes de patients des cohortes et des registres. Ces derniers disposent toutefois de données médicales d'une qualité et d'une homogénéité généralement supérieure aux données hospitalières, mais ciblées (sur une population, une pathologie et dans une zone géographique déterminée...).

Schéma 1 : Les trois étapes de structuration des données d'un entrepôt de données



Source : Haute autorité de santé

Dans la grande majorité des cas, les entrepôts sont actuellement hébergés en interne par les établissements hospitaliers, même si des réflexions pour un hébergement sur des services en nuage (cloud) existent, notamment pour disposer de capacités évolutives et pouvoir répondre à des projets de données nécessitant de fortes capacités de traitement.

⁵ Ce rapport s'appuie sur une enquête menée auprès de 17 CHU et 5 autres établissements hospitaliers.

3.2 Des usages multiples

Trois finalités des EDS hospitaliers ont été distinguées par la haute autorité de santé : **la recherche, le pilotage et le soin.**

La recherche scientifique constitue le principal usage aujourd'hui mis en avant par les entrepôts de données de santé. Selon l'enquête de la HAS⁶, les études portent d'abord sur la caractérisation de population (25 %), suivie du développement de processus d'aide à la décision (24 %), de l'étude de facteurs de risque (18 %) et de l'étude des effets de traitement (16 %). Quatre spécialités seraient les plus représentées : la santé publique, la réanimation, la médecine interne et les pathologies cardio-vasculaires.

Les travaux menés sous l'égide du comité stratégique des données de santé identifient de nombreux cas d'usage pour la recherche au sein des EDSH :

- Mieux détecter les symptômes d'une pathologie ;
- Faire un suivi prédictif de l'évolution d'une maladie ;
- Formuler des propositions thérapeutiques plus personnalisées ;
- Optimiser une prise en charge à l'aide un jumeau numérique ;
- Mettre en évidence des signaux d'alerte d'émergence de pathologies ;
- Suivre en population générale, ou encore dans des sous-populations spécifiques non toujours incluses dans les essais ou les cohortes, les impacts et les effets des prises en charge
- En cible, devenir un socle incontournable à la création de grandes cohortes (et devenir une composante à part entière des essais cliniques dits "du futur" (ex. essais adaptatifs, multi-bras et multi-étapes, etc.).

Les illustrations présentées dans l'encadré ci-après soulignent **la variété des projets de recherche comme l'intérêt des appariements des données hospitalières avec d'autres sources de données, en particulier le SNDS.** Par nature, les données des entrepôts de données de santé ne couvrent que le périmètre hospitalier, adapté à certains types d'études qui ne nécessitent pas de suivi longitudinal des patients. Mais ce périmètre demeure insuffisant pour de nombreuses études, notamment pour des pathologies en partie prises en charge en ville.

Illustrations de projets de recherche à partir des données des entrepôts de données de santé hospitaliers

1/ Développement d'un algorithme de prédiction des épidémies de gripes en France (EDS Nantes).

Les données de l'EDS ont été combinées avec d'autres sources de données issues d'internet et de santé publique France afin de développer un algorithme de prédiction des épidémies de gripes en France ;

2/ Etude sur les interactions médicamenteuses et les ruptures de traitement, à l'échelle des six centres de données cliniques d'HUGO, dans le cadre du Health data hub (projet HUGOSHARE sur 420 000 patients).

L'analyse des données des parcours de soins ville-hôpital sur 4 années, chaînées avec le SNDS, doit permettre de prévenir les accidents iatrogéniques en ville et à l'hôpital, d'améliorer les parcours de soins des patients les plus fragiles et de développer un complément aux outils d'aide à la prescription pour les professionnels de santé.

⁶ Analyse menée à partir de 248 études collectées sur les portails d'études en cours de 9 CHU et d'un 1 CH

3/ Construction et validation d'un score prédictif afin d'identifier en préopératoire le profil des patients à risque d'échec pour renforcer leur suivi ou les récuser (projet de l'entrepôt d'ELSAN, avec le CHU Rouen Normandie et l'INRAE sur une cohorte de 35 000 personnes).

4/ Impact des facteurs de risque métabolique et prédiction du risque de complication hépatique dans les maladies chroniques du foie (projet TATOOINE porté par le groupement HUGO).

L'étude portant sur cinq centres hospitaliers universitaires (base de données de 390 000 patients) vise à définir les facteurs de risque métaboliques de cirrhose et de cancer du foie, pour favoriser la prévention de ces pathologies

5/ Développement d'un outil automatique de diagnostic des cancers de la prostate à partir de l'analyse des données d'imagerie en utilisant des technologies d'intelligence artificielle (AP-HP, en partenariat avec l'INRIA).

6/ Projet porté sur la performance des scores pronostiques pour prédire la mortalité au sein d'une cohorte multicentrique de patients hospitalisés pour Covid-19 à l'AP-HP. Les résultats issus de cette recherche ont fait l'objet d'une publication⁷ dans la revue Intensive Care Medicine

7/ Projet pour mieux caractériser les trajectoires de soins des patients âgés atteints de cancer, leurs profils cliniques et biologiques et leurs devenir associés, grâce à un appariement entre les données d'une cohorte de l'AP-HP et les données de son EDS

Les entrepôts de données de santé concourent aussi à l'amélioration du pilotage de l'établissement. Ils peuvent ainsi permettre la production de tableaux de bord et d'indicateurs, comme la récente période de crise l'a illustré. Les EDS peuvent également automatiser la production d'indicateurs qualité. A titre d'exemple, depuis 2023, un outil développé par l'EDS de l'AP-HP a permis la quasi-automatisation de la production d'un indicateur sur les infections nosocomiales, limitant grandement la vérification manuelle des dossiers, auparavant systématique. Les entrepôts pourraient ainsi venir soutenir la production en routine de d'indicateurs nécessaires pour les financements à la qualité dans le futur.

Enfin, **les entrepôts de données de santé pourraient contribuer plus significativement à l'amélioration de la qualité des soins**, avec des retours réguliers vers les professionnels de santé sur leurs pratiques, le développement d'outils transverses (permettant une vue unifiée de l'historique des données d'un patient, avec une transversalité interspécialités). Ainsi, à l'AP-HP, un outil d'aide au repérage des infections du site opératoire a été mis en place à partir des données de son entrepôt ; un projet de suivi des patients en attente de transplantation est également en cours. Un retour plus important vers les professionnels les inciterait par ailleurs à améliorer la qualité de l'information médicale renseignée et serait de nature à augmenter l'implication des professionnels de santé dans les projets d'entrepôts de données de santé. **Ceci peut appeler à atténuer la distinction aujourd'hui stricte entre utilisation primaire et secondaire des données de santé.**

L'exemple de l'entrepôt de l'AP-HP illustre cette imbrication entre utilisation primaire et secondaire.

⁷ Lombardi, Y., Azoyan, L., Szychowiak, P. et al. External validation of prognostic scores for COVID-19: a multicenter cohort study of patients hospitalized in Greater Paris University Hospitals. Intensive Care Med 47, 1426–1439 (2021). <https://doi.org/10.1007/s00134-021-06524-w>

L'entrepôt de données de santé de l'AP-HP

Initié dès 2014, l'EDS de l'AP-HP a été autorisé en 2017, en poursuivant deux grandes finalités :

- Réalisation de projets de recherche et innovation (évaluation de solutions innovantes en émergence),
- Activités de pilotage (dont amélioration des prises en charge et qualité des soins).

Il regroupe les données de santé de plus de 14 millions de patients admis dans les 39 établissements du groupe. Il mobilise plus de 100 ETP (hors prestataires et expertises apportées par d'autres directions), avec des expertises variées (data scientists, médecins experts en informatique médicale, architectes, data managers, experts des données, ingénieurs de la donnée, experts en cybersécurité, experts en protection des données, etc.). Il s'appuie sur une plateforme de calcul maintenue par l'AP-HP. L'hébergement de l'EDS est assuré sur les data centers de l'AP-HP.

Depuis sa mise en place en 2017, 282 projets de recherche ont été soumis à l'avis du comité scientifique et éthique de l'EDS AP-HP, donnant lieu à 66 publications (de nombreux projets étant encore en cours).

Par ailleurs, 63 collaborations ont conduit à la signature d'un contrat ou sont sur le point de le faire. Parmi elles, 12 sont relatives à des partenariats avec des institutions privées (industrie pharmaceutique, start-up fabricant d'équipement d'imagerie, etc...).

Le nombre des projets de recherche a considérablement augmenté depuis 2017, en lien notamment avec le recrutement d'ingénieurs de données et de data scientists grâce à des fonds dédiés à la gestion de la pandémie COVID-19 ou à des domaines de pathologie particuliers (par exemple cancer). Comme pour d'autres entrepôts, la capacité à recruter des personnels qualifiés est considérée comme un défi important, soit pour des questions salariales (notamment pour les experts en mégadonnées- data scientists), soit car certains profils comme les ingénieurs de données manquent.

Un quart des projets vise à développer des algorithmes d'apprentissage automatique d'aide au diagnostic, des algorithmes prédictifs du pronostic ou de la réponse au traitement, en particulier sur le domaine de l'imagerie médicale, qui représente plus de 40 % des projets d'IA/ML. Les données sont également utilisées de manière croissante pour l'évaluation de la balance bénéfice-risque de thérapies innovantes (médicaments, dispositifs médicaux ou aides décisionnelles embarquant de l'intelligence artificielle).

Le nombre de projets portant sur le pilotage de l'activité hospitalière augmente aussi fortement, en réponse à une demande croissante de restitution de ces données sous forme de tableaux de bord. 47 rapports et tableaux de bord mobilisant des données de l'EDS sont disponibles.

Enfin, l'EDS de l'AP-HP a été à l'initiative de plusieurs projets de développement réalisés en open source afin de mettre en qualité les données, faciliter leur exploration et conduire des analyses. L'AP-HP a en particulier une forte activité autour de l'analyse des textes, entre autres à l'aide d'algorithmes d'intelligence artificielle qui ont été adaptés aux enjeux des données cliniques francophones.

3.3 De premières initiatives pour renforcer l'interopérabilité entre les entrepôts

Les entrepôts de données hospitaliers souffrent tout d'abord de la **faiblesse de la documentation des systèmes d'information hospitaliers** par les éditeurs de logiciel. Ainsi, l'enquête de la HAS souligne que moins de la moitié des entrepôts ont mis en place une documentation accessible au sein du CHU sur les flux de données, la signification ainsi que le bon usage des données qualifiées. Les structures de données sont diverses et peu lisibles pour les acteurs externes, ce qui limite la réutilisation des données.

Par ailleurs, les données sont de qualité très hétérogène, avec **des standards de données encore peu utilisés ou hétérogènes** selon les centres. Certains établissements utilisent les modèles OMOP, quand d'autres utilisent un modèle commun de données nommé eHOP. De même, en cancérologie, les CLCC utilisent la norme OSIRIS, quand l'AP-HP utilise la norme FHIR, ce qui freine l'interopérabilité.

Certaines démarches sont initiées au niveau des établissements pour renforcer la documentation et la qualité des données. Ainsi, à l'AP-HP, une équipe « Expertise données » (2 et bientôt 3 ETP) vise à mettre en place un data catalogue, ainsi qu'une gouvernance pour préciser l'organisation et les processus pour s'assurer de l'exploitabilité des données.

Deux types d'initiatives ont émergé à une échelle plus large pour harmoniser les données des entrepôts de données de santé et renforcer la maturité des EDS comme leur interopérabilité.

En premier lieu, le **développement de réseaux d'entrepôts à l'échelle régionale ou interrégionale** (voire nationale, pour la cancérologie) est une tendance marquée, afin de développer des solutions techniques communes, mutualiser les expertises et harmoniser les données. Cette approche a été encouragée par l'appel à projet pour le financement des EDS (voir partie 3.4.2 *infra*). La dynamique la plus avancée concerne le Grand Ouest (voir encadré ci-après), mais des rapprochements émergent dans d'autres régions, comme en Grand Est (projet PAISaGE incluant 7 établissements), ou en Rhône-Alpes (projet HOURS associant 4 établissements). Dans le domaine de la cancérologie, la plateforme ESME, développée par UNICANCER, regroupe les données des centres de lutte contre le cancer.

La structuration interrégionale des entrepôts de données du Grand ouest

Depuis 2016, le groupement de coopération sanitaire Hugo a lancé un programme pluriannuel pour structurer et organiser l'exploitation des données en santé sur le territoire du Grand Ouest. Un réseau interrégional a été créé pour structurer les centres de données cliniques créés au sein des 6 établissements du GCS (les CHU de Rennes, Brest, Angers, Nantes, Tours et l'institut de cancérologie de l'ouest), avec un accompagnement financier, et technique et scientifique de projets multicentriques. Ce rapprochement se traduit par l'utilisation de technologies communes et la définition de standards de données pour favoriser l'interopérabilité. Il s'appuie aussi sur la mutualisation de certaines ressources et expertises, et l'adoption de principes communs de gouvernance et de valorisation. En 2020, une plateforme de données commune (Ouest Data hub) a également été créée pour l'intégration de données pour les projets multicentriques et le chaînage des données. Celle-ci utilise l'infrastructure du CHU de Nantes. Le réseau est en capacité de chaîner (de manière décentralisée) ses données avec les données du SNDS.

La mobilisation des données chaînées au SNDS a permis la réalisation d'une étude en vie réelle sur les statines dans le champ de la pharmacovigilance, mesurant la prévalence des interactions médicamenteuses et des effets indésirables liés aux prescriptions au cours du parcours de soin du patient en ville et à l'hôpital. Cette étude a donné lieu à une publication⁸.

En second lieu, **des travaux ont été engagés à l'échelle nationale, sous l'égide du comité stratégique des données de santé, pour harmoniser les approches des entrepôts de données de santé.** Ainsi, des groupes de travail ont été initiés concernant la gouvernance des entrepôts, l'établissement d'un socle de données, le financement des entrepôts de données de santé ainsi que les principes de tarification (réflexion sur une grille commune). Ces travaux partenariaux constituent une avancée, comme en attestent les premiers résultats en matière de financement présentés *infra*, même s'ils ne vont pas encore jusqu'à l'adoption de normes communes, notamment en termes d'interopérabilité.

3.4 Une fragilité : un modèle de financement non pérenne

3.4.1 Des recettes encore difficiles à estimer dans un écosystème en développement et dans un contexte juridique évolutif

Si l'ensemble des entrepôts de données sont engagés dans des dynamiques partenariales, les partenariats sont encore émergents ou en cours de structuration et les recettes des entrepôts de données de santé restent difficiles à estimer.

⁸ Publication : Bannay, Aurélie, Mathilde Bories, Pascal Le Corre, Christine Riou, Pierre Lemordant, Pascal Van Hille, Emmanuel Chazard, Xavier Dode, Marc Cuggia, et Guillaume Bouzillé. « Leveraging National Claims and Hospital Big Data: Cohort Study on a Statin-Drug Interaction Use Case ». JMIR Medical Informatics 9, no 12 (13 décembre 2021) : e29286. <https://doi.org/10.2196/29286>.

L'ensemble des entrepôts de santé rencontrés par la mission est engagé dans des dynamiques partenariales, ce qui s'explique notamment par l'absence de financement public spécifique jusqu'à récemment. De nombreux partenariats portaient sur des projets d'exploitation des données des entrepôts, ce qui permet un co-financement des projets. A titre d'exemple, l'ensemble des projets de recherche envisagé par l'entrepôt de données ELSAN associe des acteurs extérieurs (CHU de Rouen, l'institut PRAIRIE, institut phocéen de néphrologie, etc).

Les partenariats vont parfois jusqu'à la mise à disposition de compétences de l'établissement hospitalier pour qualifier les données. Ils peuvent aussi avoir une visée technique, à l'instar de la collaboration étroite entre l'AP-HP et l'INRIA pour mettre à profit de l'AP-HP les compétences d'intégration technique de l'INRIA, avec la création d'un laboratoire commun (laboratoire Bernoulli) consacré aux sciences et technologies du numérique dans le champ de la santé. La mission du laboratoire est d'accélérer la recherche et l'innovation autour de l'exploitation de données et de développer de nouveaux outils qui faciliteront la pratique ou la recherche clinique. Il abrite deux chaires en intelligence artificielle.

Enfin, certaines plateformes de données de santé sont conçues en elles-mêmes dans des logiques collaboratives. La plateforme de cancérologie de l'INCA, co-financée par la filière d'intelligence artificielle et cancers (FIAC), constitue un premier exemple d'une telle collaboration, avec un investissement de 9 millions d'euros par les industriels (financement abondé de 8 millions d'euros par la BPI). La FIAC facilite l'accès aux données de santé pour les industriels de santé, qui devaient en retour mettre à disposition leurs données issues de recherches. Le consortium Agoria⁹ constitue un second exemple de rapprochement entre acteurs (privés), afin de constituer une infrastructure de données de santé destinée à héberger des données appariées avec le SNDS, avec l'ambition d'héberger un système fils du SNDS.

Si les partenariats sont nombreux, il existe plusieurs limites au financement par ce biais des entrepôts. Plus globalement, des attentes excessives sont probablement placées dans la valorisation financière des données de santé des entrepôts, au moins tant que ces données restent insuffisamment structurées et d'une qualité très hétérogène. Même pour des bases de données matures, **les financements issus de la valorisation de l'exploitation des données sont très loin de couvrir l'ensemble des coûts.** A titre d'exemple, la banque nationale des maladies rares (BNDMR) a perçu 1,5 million d'euros grâce à une valorisation active de ses données depuis deux ans, pour un coût complet estimé entre 30 et 40 millions d'euros depuis dix ans.

Par ailleurs, **la capacité des entrepôts de données à valoriser leurs données est contrainte par le cadre juridique européen.** Le règlement européen sur la gouvernance européenne des données, entré en vigueur en juin 2022 et applicable depuis septembre 2023, encadre les redevances pouvant être perçues par des producteurs de données, qui doivent être fondées sur les coûts liés au traitement des demandes de réutilisation des données¹⁰ (article 6). Cette approche d'une tarification au coût marginal va à l'encontre d'une tarification fondée sur les coûts complets. Toutefois, aucune jurisprudence n'est encore venue préciser la portée de cette définition. De plus,

⁹ Le consortium, dont le chef de file est Docaposte, compte aussi parmi ses membres Astrazeneca, Impact Healthcare et Takeda. Il n'a pas de personnalité morale à ce jour.

¹⁰ Coûts de reproduction, de fourniture et de diffusion des données ; coûts liés à l'anonymisation ; coûts liés à la maintenance, à l'assistance aux réutilisateurs, etc.

les négociations en cours sur le règlement d'espace européen des données de santé pourraient conduire à des assouplissements – bienvenus - de cette règle pour les données de santé (cf. annexe 2 sur le cadre juridique).

3.4.2 De récents financements des entrepôts de données de santé par appels à projets

Concernant les financements publics, un financement de 75 millions d'euros a été octroyé dans le cadre de l'appel à projet « Accompagnement et soutien à la constitution d'entrepôts de données de santé hospitaliers », piloté par la délégation du numérique en santé. Une partie du financement est issue du programme d'investissement France 2030 (35 M€), le reste est financé par l'objectif national de dépenses l'assurance maladie (ONDAM). Ce financement, qui permet de couvrir jusqu'à 50% des coûts engagés¹¹ jusqu'en 2026, a été très largement salué, car il a permis de conforter le développement des entrepôts auparavant autofinancés (ou financés par des partenariats). Il a également contribué à enclencher une dynamique d'homogénéisation et de mutualisation entre entrepôts (voir détail dans l'encadré ci-dessous). Sa mise en œuvre suscite toutefois des tensions, car il est attendu des établissements de santé une mise à disposition de tout ou partie des données via l'inscription au catalogue de la plateforme des données de santé (Health data hub) (voir partie 1.3.2 du rapport).

Un second appel à projet a été lancé en octobre 2023 (appel à projet DATAE) pour le financement de cas d'usages des entrepôts de données de santé, avec un financement de 3,5 millions d'euros par an, pour trois ans, piloté par la direction générale de l'offre de soins. Celui-ci vise les entrepôts de données de santé matures. Le financement est issu de l'ONDAM.

L'appel à projet « Accompagnement et soutien à la constitution d'entrepôts de données de santé hospitaliers »

Le cahier des charges de l'appel à projet précise qu'il vise les établissements de santé (à partir d'une certaine taille) ou groupements de coopération sanitaire, avec une valorisation des démarches portées par plusieurs établissements de santé.

Il encourage au partage des données des entrepôts, car l'inscription au catalogue de la plateforme des données de santé de tout ou partie de des données est attendue. Il valorise aussi l'adoption de normes communes et de standardisations, et les démarches de mutualisation des moyens entre entrepôts.

Une harmonisation des règles de gouvernance et d'accès aux données est également recherchée, avec une ouverture des comités scientifiques et éthiques à des membres extérieurs à l'entrepôt de données de santé, une clarification des délais d'accès aux données, etc. Les lauréats s'engagent aussi à participer aux travaux du Comité stratégique des données de santé pour définir des grilles de tarification homogènes.

¹¹ L'enveloppe initiale prévue pour ce dispositif de 50 millions d'euros de France 2030 a été abondée, début mars 2023, de 25 millions d'€ supplémentaire, ce qui porte l'enveloppe totale de l'AAP à 75 millions d'euros. Deux vagues successives d'appels à projets ont été organisées : six lauréats ont été annoncés en mars 2023, et les lauréats de la seconde vague devraient prochainement être annoncés.

3.4.3 Des coûts des entrepôts récemment objectivés

Une analyse approfondie des coûts de développement et de fonctionnement des entrepôts de données de santé hospitaliers a été menée en 2023 par le comité stratégique des données de santé, en s'appuyant sur l'expérience de sept entrepôts de données de santé portés par des CHU de taille variable. Une évaluation des coûts a été menée pour chacune des phases de constitution des entrepôts :

- La phase de cadrage qui correspond à l'élaboration d'une stratégie et de l'offre de services, des flux de données, du cadrage réglementaire, de la gouvernance et des modalités d'accès aux données, du choix de la solution technologique, etc. Cette phase d'un an environ représente un coût évalué à 720 000€ en moyenne (de 300 000€ à 950 000€- voir détail dans les tableaux ci-dessous) ;
- Le coût de développement (« Build ») de l'entrepôt couvre la construction de l'infrastructure, la mise en œuvre des premiers flux de données, le déploiement de l'offre de services, la mise en place de la gouvernance, du cadre juridique ainsi que les frais de pilotage. La durée de cette phase est de deux à trois ans selon les entrepôts. Son coût est évalué en moyenne à 3,7 M€, avec des écarts notables (de 1,4 M€ à 6 M€) ;
- Enfin, les coûts d'exploitation (« Run »), qui recouvrent notamment la mise en qualité des données, l'intégration de nouveaux flux de données et le financement des coûts de consommation liés aux projets et aux prestations de sécurité sont évalués à 3,3 M€ en moyenne (de 2,5 M€ à 4,8 M€).

Au total, le groupe de travail a estimé en moyenne à 2M€ par an le coût de développement d'un EDS sur 5 ans puis à 3M€ par an le coût de financement de son fonctionnement nominal.

Tableau 3 : Evaluation des coûts de construction et d'exploitation de six entrepôts de données de santé, en euros

	Lille	Brest	Reims	Rouen	AP-HP	GCS HUGO	Grenoble	Moyenne	Ecart type
Coût du cadrage	597 000	729 200	758 800	735 800	958 000	958 000	294 800	718 800	227 685
Coût d'investissement (« Build »)	2 289 400	2 879 200	4 241 200	1 470 400	6 523 600	6 079 600	2 182 000	3 666 486	1 994 124
Coût d'exploitation (« Run »)	2 471 200	2 898 780	4 006 000	3 059 427	4 752 700	3 344 100	2 674 993	3 315 314	806 791

Source : Annexe de la note du groupe de travail « Financement pérenne des entrepôts de données de santé »

Tableau 4 : Montée en charge des coûts sur cinq ans par établissement

	Année 1	Année 2	Année 3	Année 4	Année 5	TOTAL
Lille	597 000 €	763 133 €	1 257 373 €	2 245 853 €	2 471 200 €	7 334 560 €
Brest	729 200 €	959 733 €	1 539 489 €	2 699 001 €	2 898 780 €	8 826 204 €
Reims	758 800 €	1 413 733 €	2 214 933 €	3 817 333 €	4 006 000 €	12 210 800 €
Rouen	735 800 €	735 200 €	1 041 143 €	1 653 028 €	1 529 713 €	5 694 884 €
APHP	958 000 €	2 174 533 €	3 125 073 €	5 026 153 €	4 752 700 €	16 036 460 €
GCS HUGO	958 000 €	2 026 533 €	2 695 353 €	4 032 993 €	3 344 100 €	13 056 980 €
Grenoble	294 800 €	1 091 000 €	1 625 999 €	2 695 996 €	2 674 993 €	8 382 788 €

Source : Annexe de la note du groupe de travail « Financement pérenne des entrepôts de données de santé »

En conclusion, si l'appel à projet pour le financement des entrepôts de données de santé hospitaliers dans le cadre du programme France 2030 a constitué une avancée importante, un financement pérenne dédié, compte tenu de coûts fixes bien identifiés, devrait être mis en place. Les entrepôts financés devront s'engager en contrepartie au respect des normes établies par le comité stratégique des données de santé (interopérabilité mais aussi en matière de fonctionnement des comités éthiques et scientifiques).

ANNEXE 2 : Le cadre juridique de l'accès aux données de santé à des fins de recherche

Table des matières

ANNEXE 2 : LE CADRE JURIDIQUE DE L'ACCES AUX DONNEES DE SANTE A DES FINS DE RECHERCHE	4
1 UN CADRE JURIDIQUE DES TRAITEMENTS DE DONNEES DE SANTE FIXE PAR LE RGPD ET LA LOI « INFORMATIQUE ET LIBERTES ».....	20
2 UN PARCOURS D'ACCES REGLEMENTAIRE RELATIVEMENT COMPLEXE EN FRANCE.....	20
2.1 UN PROCESSUS D'ACCES AUX DONNEES DE SANTE QUI FAIT INTERVENIR DE NOMBREUX ACTEURS.....	20
2.2 LES METHODOLOGIES DE REFERENCE	21
2.3 UN AVIS DU CESREES SUR LA METHODOLOGIE DU PROJET	23
2.4 UN CADRE SPECIFIQUE POUR L'ACCES AUX DONNEES DU SNDS QUI PREVOIT NOTAMMENT DES ACCES PERMANENTS	24
3 ... APPEL A EVOLUER DANS LA PERSPECTIVE DU FUTUR REGLEMENT EUROPEEN RELATIF A L'ESPACE EUROPEEN DES DONNEES DE SANTE.....	26
3.1 LA CONSECRATION D'UNE OBLIGATION POUR TOUT DETENTEUR DE DONNEES DE SANTE DE LES METTRE A DISPOSITION A DES FINS D'UTILISATION SECONDAIRE.....	26
3.2 UNE DEFINITION EXTENSIVE DES FINALITES DE L'UTILISATION SECONDAIRE DES DONNEES DE SANTE .	26
3.3 UN CIRCUIT D'ACCES AUX DONNEES ARTICULE AUTOUR DU « <i>HEALTH DATA ACCESS BODY</i> » (HDAB) .	27
3.4 LE PRINCIPE D'UNE REDEVANCE COMME CONTREPARTIE DE LA MISE A DISPOSITION DES DONNEES DE SANTE.....	29
3.5 UNE POSSIBLE REMISE EN CAUSE PAR CE PROJET DE REGLEMENT DU DISPOSITIF FRANÇAIS DES ACCES PERMANENTS	29

Dans cette annexe sont décrits le cadre juridique d'ensemble des traitements de données de santé (partie 1), le processus d'accès aux données de santé (partie 2), et les évolutions prévues par le futur règlement européen sur l'espace des données de santé (partie 3).

1 Un cadre juridique des traitements de données de santé fixé par le RGPD et la loi « Informatique et libertés »

Si l'article 9 du RGPD pose, en son paragraphe 1, le principe selon lequel le traitement des données de santé à caractère personnel est interdit, son paragraphe 2, point j prévoit une exception à cette interdiction en matière de recherche scientifique ou historique¹². Conformément à l'article 89, paragraphe 1 du règlement, des garanties appropriées doivent ainsi encadrer cet accès aux données de santé à des fins de recherche. **Les modalités de cet accès sont ainsi régies par la section 3 du chapitre III du titre II de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, dite loi « Informatique et libertés ».**

L'article 66 de la loi « Informatique et libertés » impose aux traitements d'utilisation secondaire des données de santé de présenter une **finalité d'intérêt public**, laquelle obligation est également rappelée au premier alinéa de l'article L. 1460-1 du code de la santé publique (CSP). Il est intéressant de noter que la loi « Informatique et libertés » a pris soin de préciser que « *la garantie de normes élevées de qualité et de sécurité des soins de santé et des médicaments ou des dispositifs médicaux* » – qui correspond à une finalité spécifique de traitement des données de santé mentionnée à l'article 9, paragraphe 2, point i du RGPD – constitue une finalité d'intérêt public, consacrant ainsi la possibilité d'un accès aux données de santé notamment pour les projets d'évaluation en vie réelle des produits de santé et d'amélioration de la qualité du système de soins.

2 Un parcours d'accès règlementaire relativement complexe en France...

2.1 Un processus d'accès aux données de santé qui fait intervenir de nombreux acteurs

Lorsque le projet d'utilisation secondaire est présenté à des fins de recherche, d'étude ou d'évaluation dans le domaine de la santé, il doit :

- soit attester de sa **conformité à une méthodologie de référence** (MR), homologuée et publiée par la commission nationale de l'informatique et des libertés (CNIL), dans le cadre d'une déclaration de conformité adressée à cette dernière (article 73 de la loi « Informatique et libertés », second alinéa) ;
- soit obtenir une **autorisation accordée par la CNIL** (article 76 de la loi « Informatique et libertés »), après avis :
 - d'un comité de protection des personnes (CPP) pour les recherches impliquant la personne humaine ;

¹² Aux termes du considérant 159 du RGPD, « par "fins de recherche scientifique", il convient également d'entendre les études menées dans l'intérêt public dans le domaine de la santé publique ».

- du comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (CESREES) pour les recherches, études ou évaluations n'impliquant pas la personne humaine.

En application du dernier alinéa de l'article 76 de la loi « Informatique et libertés » et du 3° de l'article L. 1462-1 du CSP, le HDH assure le **secrétariat unique du dépôt de l'ensemble des projets de recherche n'impliquant pas la personne humaine et recourant à un traitement de données de santé à caractère personnel**. Il assure également, en vertu du 3° de l'article L. 1462-1 du CSP, le secrétariat du CESREES. Le HDH dispose d'un délai de 7 jours ouvrés pour transmettre les dossiers déposés auprès du secrétariat unique au CESREES (article 88 de la loi « Informatiques et libertés »). Le CESREES dispose d'un délai d'un mois pour rendre son avis.

Une fois obtenu l'avis favorable du CESREES, la demande est transmise par le HDH à la CNIL qui dispose alors d'un délai de deux mois, renouvelable une fois, pour délivrer l'autorisation, laquelle peut être tacite en l'absence de décision expresse à l'expiration de ce délai.

2.2 Les méthodologies de référence

Il existe aujourd'hui **8 méthodologies de référence** concernant les traitements de données de santé à des fins de recherche, d'étude ou d'évaluation dans le domaine de la santé, les deux dernières (MR-007 et MR-008) ayant été publiées le 12 octobre 2023. Elles sont établies en **concertation avec le Health Data Hub (HDH) ainsi qu'avec les organismes publics et privés représentatifs des acteurs concernés** :

Tableau 5 : Méthodologies de référence pour les traitements de données de santé

Si la recherche est qualifiée de recherche impliquant la personne humaine (RIPH)	
<u>MR-001</u>	<p>Applicable si la recherche nécessite le recueil du consentement des personnes concernées pour réaliser la recherche.</p> <p>Cela correspond :</p> <ul style="list-style-type: none"> • aux recherches interventionnelles sur la personne non justifiée par sa prise en charge habituelle (RIPH1) ; • aux recherches interventionnelles qui ne comportent que des risques et contraintes minimales (RIPH2) pour lesquelles une information individuelle est réalisée ; • aux recherches impliquant le recueil du consentement en raison de la réalisation d'un examen des caractéristiques génétiques des participants ; • aux essais cliniques de médicaments (régis par le règlement (UE) 536/2014 du 16 avril 2014 relatif aux essais cliniques de médicaments à usage humain), à l'exception des essais cliniques par grappes ; • aux investigations cliniques liées aux dispositifs médicaux (régies par le règlement (UE) 2017/745 du 5 avril 2017 relatif aux dispositifs médicaux.

<u>MR-002</u>	Applicable aux recherches non interventionnelles de performances concernant les dispositifs médicaux de diagnostic in vitro.
<u>MR-003</u>	Applicable aux recherches nécessitant une information des personnes concernées sans que ces dernières ne s'y opposent. Cela correspond : <ul style="list-style-type: none"> • aux recherches interventionnelles qui ne comportent que des risques et contraintes minimales (RIPH2) pour lesquelles une information collective est réalisée, après avis du comité de protection des personnes ; • aux recherches non interventionnelles qui ne comportent aucun risque ni contrainte sur la personne (RIPH3) ; • aux essais de médicaments par grappes.
Si la recherche est qualifiée de recherche n'impliquant pas la personne humaine (RNIPH)	
<u>MR-004</u>	Applicable aux recherches, études ou évaluations n'impliquant pas la personne humaine (autres que les traitements de données du SNDS exclusivement).
<u>MR-005</u>	Applicable aux recherches, études ou évaluations nécessitant l'accès aux données du PMSI et/ou des RPU par les établissements de santé et les fédérations hospitalières.
<u>MR-006</u>	Applicable aux recherches, études ou évaluations nécessitant l'accès aux données du PMSI par les industriels de santé.
<u>MR-007</u>	Applicable aux recherches, études ou évaluations nécessitant l'accès à la base principale du SNDS par les organismes agissant dans le cadre de leur mission d'intérêt public.
<u>MR-008</u>	Applicable aux recherches, études ou évaluations nécessitant l'accès à la base principale du SNDS par les organismes agissant dans le cadre de leurs intérêts légitimes (par exemple : industriels du secteur de la santé traitant les données du SNIIRAM).
<u>Référentiel ESND (ex-EGB)</u>	Encadre l'accès à l'échantillon national des données du SNDS (ENSD).

Source : CNIL

Il est à noter que, si un avis préalable du CESREES n'est pas requis pour la conformité aux MR-004, MR-005 et MR-006, il est nécessaire pour les traitements faisant l'objet d'une déclaration de conformité aux MR-007 et MR-008.

Si l'existence de méthodologies de référence est très appréciée par les acteurs, elles ne répondent pas toujours parfaitement aux besoins des acteurs, comme l'illustre l'encadré ci-après.

Les MR-007 et MR-008 : des méthodologies attendues par les acteurs du secteur, mais qui présentent encore quelques insuffisances

Les deux dernières méthodologies de référence MR-007 et MR-008 ont été conçues pour **faciliter l'accès à la base principale du SNDS** des organismes publics agissant dans le cadre de leurs missions d'intérêt public (MR-007) et des organismes privés – hors assureurs – pour lesquels la recherche est nécessaire à la poursuite d'un intérêt légitime (MR-008) : une série de finalités, identiques à celles mentionnées dans le référentiel encadrant la mise à disposition de l'échantillon du SNDS (ESND), sont ainsi présumées d'intérêt public, et les traitements s'inscrivant dans ces méthodologies pourront déroger à l'obligation d'information individuelle des personnes concernées pour les données provenant du SNDS¹³.

Toutefois, certains acteurs regrettent que, dans le cadre de ces deux nouvelles méthodologies :

- les seules données pouvant faire l'objet du traitement doivent être celles provenant exclusivement et directement des bases de données mises à disposition par la CNAM, toute réutilisation des données étant proscrite. Ce faisant, **ces méthodologies font obstacle à une réutilisation des données du SNDS**, qui auraient déjà été mobilisées dans le cadre d'un projet, sur un système fils au bénéfice d'un autre projet ;

- **sont également exclus les traitements nécessitant un appariement des données du SNDS avec les données à caractère personnel issues d'autres sources**, seules les données issues de la base principale du SNDS pouvant être traitées, alors que, dans une version préliminaire, ces chaînages étaient inclus ;

- les projets doivent continuer de recueillir un **avis favorable préalable du CESREES**.

2.3 Un avis du CESREES sur la méthodologie du projet

Saisi par le HDH, le CESREES se prononce, aux termes de l'article 90 du décret n° 2019-536 du 29 mai 2019¹⁴, « sur la méthodologie retenue, sur la nécessité du recours à des données à caractère personnel, sur la pertinence de celles-ci par rapport à la finalité du traitement et, s'il y a lieu, sur la pertinence scientifique et éthique du projet ». Le CESREES peut également se saisir ou être saisi, en application du second alinéa de l'article 72 de la loi « Informatique et libertés », par le président de la CNIL ou par le ministre chargé de la santé de la question du caractère d'intérêt public d'un traitement de données de santé à des fins de recherche.

Le comité dispose d'un délai d'un mois à compter de la réception du dossier complet, renouvelable une fois, pour rendre son avis.

À l'issue de l'avis du CESREES, le porteur de projet doit confirmer auprès du HDH qu'il souhaite que son projet – si celui-ci n'entre pas dans le cadre d'une MR – soit examiné par la CNIL.

¹³ Sous réserve de mesures appropriées pour protéger les droits, libertés et intérêts légitimes des personnes concernées, et d'un effort de transparence par la publication des informations disponibles.

¹⁴ Pris pour l'application de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

2.4 Un cadre spécifique pour l'accès aux données du SNDS qui prévoit notamment des accès permanents

Enfin, il convient de relever que l'accès aux données du SNDS, auquel s'appliquent les dispositions précitées de la loi « Informatique et libertés », est également encadré par des dispositions spécifiques du CSP. Les finalités d'un tel accès sont ainsi limitativement énumérées à son article L. 1461-3 qui prévoit qu'il ne peut servir qu'aux traitements :

- soit contribuant à une finalité mentionnée au III¹⁵ de l'article L. 1461-1 et répondant à un motif d'intérêt public ;
- soit nécessaires à l'accomplissement des missions des services de l'État, des établissements publics ou des organismes chargés d'une mission de service public compétents.

L'accès des services de l'État, des établissements publics ou des organismes chargés d'une mission de service public relève du **dispositif de l'accès permanent** : il concerne les services, établissements publics et organismes énumérés à l'article R. 1461-12 du CSP. La profondeur historique maximale de cet accès permanent varie selon la nature de l'organisme concerné (19 ans, 9 ans ou 5 ans, selon l'article R. 1461-13 du CSP). Aux termes de l'article R. 1461-15 du CSP, les traitements effectués dans le cadre d'un accès permanent ne peuvent pas excéder les limites fixées par l'article R. 1461-14 du même code, et notamment **pas impliquer d'appariement de données issues du SNDS avec d'autres données, ce type d'appariement requérant une autorisation préalable de la CNIL.**

Les organismes bénéficiant d'un accès permanent au SNDS

1° La direction de la recherche, des études, de l'évaluation et des statistiques, la direction générale de la santé et la direction générale de l'offre de soins, la direction de la sécurité sociale, le secrétariat général des ministères sociaux, la direction du budget, la direction générale du Trésor et le service de santé des armées ;

2° Les agences régionales de santé ;

3° Les caisses nationales des régimes de l'assurance maladie obligatoire, les organismes locaux et régionaux de l'assurance maladie obligatoire ;

4° La Caisse nationale de solidarité pour l'autonomie ;

5° La Haute Autorité de santé ;

6° L'Autorité de sûreté nucléaire ;

¹⁵ Qui prévoit que le SNDS a pour finalité la mise à disposition de données pour contribuer :

« 1° À l'information sur la santé ainsi que sur l'offre de soins, la prise en charge médico-sociale et leur qualité ;

2° À la définition, à la mise en œuvre et à l'évaluation des politiques de santé et de protection sociale ;

3° À la connaissance des dépenses de santé, des dépenses d'assurance maladie et des dépenses médico-sociales ;

4° À l'information des professionnels, des structures et des établissements de santé ou médico-sociaux sur leur activité ;

5° À la surveillance, à la veille et à la sécurité sanitaires ;

6° À la recherche, aux études, à l'évaluation et à l'innovation dans les domaines de la santé et de la prise en charge médico-sociale ».

- 7° L'Agence nationale de santé publique ;
- 8° L'Agence nationale de sécurité du médicament et des produits de santé ;
- 9° L'Agence de biomédecine ;
- 10° L'Institut de radioprotection et de sûreté nucléaire ;
- 11° L'Institut national du cancer ;
- 12° L'Établissement français du sang ;
- 13° L'Agence technique de l'information sur l'hospitalisation ;
- 14° L'Agence nationale d'appui à la performance des établissements de santé et médico-sociaux ;
- 15° La Plateforme des données de santé ;
- 16° L'Institut de recherche et documentation en économie de la santé ;
- 17° L'Institut national d'études démographiques ;
- 18° L'Observatoire français des drogues et toxicomanies ;
- 19° Le Haut Conseil pour l'avenir de l'assurance maladie ;
- 20° Le Fonds de financement de la couverture maladie universelle ;
- 21° Les observatoires régionaux de la santé ;
- 22° Les unions régionales de professionnels de santé ;
- 23° Les équipes de recherche de l'Institut national de la santé et de la recherche médicale ;
- 24° Les équipes de recherche des centres hospitaliers universitaires et des centres de lutte contre le cancer ;
- 25° Les équipes de recherche et de formation de l'École des hautes études en santé publique ;
- 26° L'Institut national de la statistique et des études économiques ;
- 27° L'Agence centrale des organismes de sécurité sociale ;
- 28° Les équipes de recherche du Centre national de la recherche scientifique dans le cadre de projets intéressant la santé publique ;
- 29° Les équipes de recherche de l'Institut national de recherche en informatique et en automatique dans le cadre de projets intéressant la santé publique ;
- 30° La Cour des comptes.

Par ailleurs, l'article L. 1461-4 du CSP pose le principe selon lequel le SNDS ne peut contenir ni les noms et prénoms des personnes, ni leur numéro d'inscription au répertoire national d'identification des personnes physiques (NIR), ni leur adresse, conformément à la règle, fixée par l'article L. 1460-1 du même code, selon laquelle les traitements de données de santé à caractère personnel ne doivent en aucun cas avoir pour fin l'identification directe ou indirecte des personnes concernées. **Dans cette logique, les appariements directs via le NIR entre les données du SNDS et celles issues d'autres bases de données de santé ne sont, pour l'heure, permis que grâce à la chaîne de pseudonymisation des NIR mise en place par la CNAM.**

3 ... appelé à évoluer dans la perspective du futur règlement européen relatif à l'Espace européen des données de santé

Le chapitre IV de la proposition de règlement du Parlement européen et du Conseil relatif à l'espace européen des données de santé, dit « règlement EHDS » (« *European Health Data Space* »), fixe un cadre pour l'utilisation secondaire des données de santé dans l'Union européenne. Ces dispositions sont encore en cours de négociation à l'heure de la finalisation du rapport.

3.1 La consécration d'une obligation pour tout détenteur de données de santé de les mettre à disposition à des fins d'utilisation secondaire

Est inscrite dans le projet de règlement une **obligation pour les détenteurs de ces données de les mettre à disposition à des fins d'utilisation secondaire**, cette obligation n'étant néanmoins pas applicable aux microentreprises, c'est-à-dire aux structures employant moins de dix personnes et ayant un chiffre d'affaires annuel inférieur à deux millions.

Sont cependant reconnus un droit à la propriété intellectuelle sur les données de santé et un droit à la protection du secret des affaires d'entreprises privées, notamment pour les entreprises de dispositifs médicaux ou les sociétés pharmaceutiques. Ces droits impliquent qu'en cas d'utilisation secondaire de ces données, **toutes les mesures nécessaires à la préservation des droits de propriété intellectuelle et des secrets des affaires devront être mises en œuvre**. De tels droits ne font cependant pas obstacle à l'accès des autorités publiques à ces données lorsqu'il s'agit de protéger la santé humaine, notamment dans un contexte de crise sanitaire pour la prévention et la gestion des risques pandémiques et la détection des dispositifs médicaux défectueux.

3.2 Une définition extensive des finalités de l'utilisation secondaire des données de santé

Le projet de règlement EHDS définit ainsi **15 catégories minimales de données destinées à une utilisation secondaire**, qui comprennent notamment les données relatives aux déterminants sociaux, environnementaux et comportementaux de la santé, en passant par les données génomiques, les données d'essais cliniques, les cohortes, les données de santé traitées à des fins de fourniture de soins ou à des fins de santé publique, de recherche, d'innovation, d'élaboration des politiques...

Le règlement définit les finalités auxquelles doivent répondre les demandeurs d'utilisation secondaire de données de santé qui sont au nombre de huit. Ces finalités sont relatives :

- aux activités pour des raisons d'intérêt public dans le domaine de la santé publique et de la santé au travail ;

- au fait d'aider les organismes du secteur public ou les institutions, organes et organismes de l'Union, dont les autorités réglementaires, dans le secteur de la santé ou des soins, à accomplir les tâches inscrites dans leur mandat ;
- au fait de produire des statistiques officielles à l'échelon national, plurinational et de l'Union en rapport avec les secteurs de la santé ou des soins ;
- aux activités d'éducation ou d'enseignement dans les secteurs de la santé ou des soins ;
- à la recherche scientifique ayant trait aux secteurs de la santé ou des soins ;
- aux activités de développement et d'innovation pour les produits ou services contribuant à la santé publique ou à la sécurité sociale, ou à la garantie d'un niveau élevé de qualité et de sécurité des soins de santé, des médicaments ou des dispositifs médicaux ;
- à la formation, au test et à l'évaluation des algorithmes ;
- à la fourniture de soins de santé personnalisés.

Le projet de règlement définit également les **finalités interdites** : utilisation des données pour des mesures préjudiciables à la personne, pour de la publicité commerciale, pour la majoration de primes d'assurance ou pour la mise au point de produits nocifs....

3.3 Un circuit d'accès aux données articulé autour du « *Health Data Access Body* » (HDAB)

Il reviendra à chaque État membre de désigner un ou plusieurs organismes responsables de l'accès aux données de santé à des fins d'utilisation secondaire, plus communément désigné sous le sigle « HDAB » (« *Health Data Access Body* »). En cas de pluralité de HDAB, il conviendra de désigner un HDAB coordonnateur.

Les HDAB auront pour missions essentiellement de **statuer sur les demandes d'accès, de rendre public un catalogue national des ensembles de données disponibles avec des informations sur la source et la nature de ces données et de préciser les conditions de mise à disposition**. Ils sont également chargés de constituer un répertoire des autorisations de traitement de données ainsi que d'assurer la publication des sanctions.

Les HDAB devront ainsi assurer la publication auprès du grand public des informations relatives à la mise à disposition des données de santé à des fins d'utilisation secondaire, en précisant la base juridique de l'octroi, en rappelant les droits des personnes physiques et les modalités d'exercice de ces droits, et en présentant les résultats des projets pour lesquels les données ont été utilisées. Un utilisateur des données de santé pourra, en outre, informer le HDAB d'une constatation ayant une incidence sur la santé de la personne dont les données ont été utilisées : le HDAB aura alors la possibilité d'en informer la personne physique et son professionnel de santé.

Il est également rappelé qu'en matière de données de santé, les organisations altruistes en matière de données ont l'obligation de se conformer aux règles du règlement *Data Governance Act* et de gérer leurs données dans un environnement sécurisé.

Tout détenteur de données de santé aura, par ailleurs, l'obligation de communiquer au HDAB le contenu de ses ensembles de données et de les mettre à disposition dans un délai, selon la dernière version du compromis connue de la mission, de trois mois à compter de la demande du HDAB (renouvelable une fois à titre exceptionnel).

Afin de faire respecter l'obligation par tout détenteur de données de santé de les partager, le HDAB se voit reconnaître un **pouvoir de sanction** par lequel il lui sera possible de révoquer l'autorisation de traitement délivrée, voire d'infliger une amende et, en cas de manquements répétés, d'interdire au détenteur fautif l'accès à l'EHDS pendant au maximum cinq ans.

Les demandes d'utilisation secondaire de données de santé devront être centralisées par un HDAB qui, le cas échéant, les transmettra aux autres HDAB et participants autorisés à avoir accès la plateforme européenne DonnéesDeSanté@UE (HealthData@EU) spécifiés dans la demande. Une réponse devra être apportée dans un délai, selon la dernière version du compromis connue de la mission, de trois mois à compter de la réception de la demande, ce délai pouvant être renouvelé une fois en fonction de la complexité de la demande.

À cela s'ajoute le fait qu'en application de l'article 46 du projet de règlement EHDS, une fois l'autorisation de traitement accordée, l'organisme responsable de l'accès aux données sera tenu de les mettre à disposition dans un **délai de deux mois après les avoir reçues du détenteur des données**, « *sauf si l'organisme responsable de l'accès aux données indique qu'il mettra à disposition les données dans un délai plus long* ».

Dans l'hypothèse où la demande ne porterait que sur des données de santé émanant d'un seul détenteur de données situé dans un seul État membre, il est envisagé, afin d'alléger la charge administrative que représente la gestion de ces demandes pour les HDAB coordonnateur au niveau central, de permettre au demandeur d'adresser sa demande directement au détenteur de données qui devra observer les mêmes exigences que celles applicables aux HDAB centraux.

Il convient de relever qu'en vertu du **principe de minimisation des données**, l'article 44 du projet de règlement EHDS prévoit la règle selon laquelle **les HDAB devront donner accès aux données de santé dans un format anonymisé dès lors que la finalité du traitement par l'utilisateur de données peut être atteinte à l'aide de telles données anonymisées**. Ce n'est donc que, **par exception**, dans l'hypothèse où la finalité du traitement ne pourrait être atteinte à l'aide de données anonymisées, que **les HDAB pourront donner accès aux données dans un format pseudonymisé**, les utilisateurs des données ayant alors l'interdiction de réidentifier les données pseudonymisées qui leur sont fournies.

Le HDAB ou le détenteur de données unique ont l'obligation de **fournir l'accès aux données de santé dans un environnement sécurisé**. Le HDAB et les utilisateurs de données seront responsables conjoints de traitement.

Est ouverte, à la discrétion des États membres, la possibilité d'une reconnaissance mutuelle d'une autorisation de traitement entre HDAB de différents États membres.

Tous les ensembles de données collectées et traitées avec le soutien d'un financement public devront faire l'objet d'un étiquetage de qualité et d'utilité des données. En l'absence de financement public, l'étiquetage sera facultatif.

3.4 Le principe d'une redevance comme contrepartie de la mise à disposition des données de santé

Il convient de rappeler que l'article 6 du règlement sur la gouvernance des données (*Data Governance Act*), en vigueur depuis septembre 2023, pose le principe d'une redevance qui peut être prélevée par tout organisme public qui autorise la réutilisation de catégories de données et qui est calculée sur la base des coûts liés à la conduite de la procédure de demande de réutilisation des catégories de données et doit être limitée aux coûts nécessaires relatifs : à la reproduction, à la fourniture et à la diffusion des données, à l'acquisition des droits, à l'anonymisation ou à d'autres formes de préparation des données à caractère personnel et des données commerciales confidentielles, à la maintenance de l'environnement de traitement sécurisé, à l'acquisition du droit d'autoriser la réutilisation conformément au présent chapitre par des tiers extérieurs au secteur public, et à l'assistance fournie aux ré-utilisateurs pour obtenir le consentement des personnes concernées et l'autorisation des détenteurs de données dont les droits et intérêts peuvent être affectés par cette réutilisation.

L'article 42 du projet de règlement EHDS introduit la possibilité pour les HDAB et les détenteurs de données uniques de percevoir une **redevance au titre la mise à disposition de leurs données à des fins d'utilisation secondaire**. Dans le projet de règlement initial, il est prévu, dans la même logique que le règlement sur la gouvernance des données, que le montant de cette redevance soit limité aux coûts de traitements des demandes de réutilisation des données (procédure, évaluation, réponse...), sans pouvoir couvrir les coûts complets liés à la collecte et à la mise à disposition des données. Une compensation pour une partie des coûts de collecte pour le détenteur est prévue uniquement si les données ne sont pas détenues par un HDAB ou par un organisme du secteur public. En outre, le projet de règlement prévoit que les données améliorées ou enrichies puissent être compensées au titre des ressources humaines et techniques par une redevance complémentaire.

Toutefois, la France plaide pour que la redevance appliquée par les HDAB puisse, pour l'ensemble des détenteurs de données, publics ou privés, **couvrir l'ensemble des coûts liés à la collecte et à la mise à disposition des données, y compris les frais de mise en qualité de ces données**. Les autres pays semblent sensibles à cette demande, poussée par la France pendant la présidence espagnole de l'Union européenne.

Le projet de règlement précise aussi que les redevances doivent être transparentes, proportionnées au coût de la mise à disposition et, le cas échéant, de la collecte. Elles pourraient être abaissées pour tenir compte de ce que le demandeur est une PME, un organisme public, un organisme de l'Union européenne, une université ou encore un établissement de santé. En cas de litiges, des mécanismes de règlement des différends sont prévus.

3.5 Une possible remise en cause par ce projet de règlement du dispositif français des accès permanents

Dans sa version initiale, le projet de règlement EHDS comprenait un article 48 prévoyait de dispenser les organismes du secteur public de l'obligation de justifier d'une autorisation de traitement pour accéder aux données. Toutefois, cette disposition n'a pu être maintenue à ce

stade des négociations entre États membres, un certain nombre d'entre eux, parfois pour des raisons historiques tenant à l'expérience de régimes autoritaires, redoutant la possibilité que des organismes relevant de la puissance publique puissent disposer d'une possibilité d'accès permanent à des données de santé à caractère personnel.

Si le dispositif des accès permanents tel qu'il existe aujourd'hui en France court donc le risque de ne pouvoir être maintenu sous l'empire du nouveau règlement européen, la délégation au numérique en santé indique avoir négocié au nom de la France une procédure simplifiée et accélérée de type « *fast track* » qui serait ouverte aux anciens bénéficiaires des accès permanents. Il pourrait s'agir pour ces organismes de bénéficier d'un accès sous réserve de démontrer la conformité des traitements concernés à une série d'exigences de type « méthodologie de référence », un tel accès pouvant alors être mobilisé quotidiennement pour répondre à des finalités différentes.

ANNEXE 3 : Comparaisons internationales en matière d'accès aux données de santé

Cette annexe expose des éléments de comparaison internationale sur l'accès aux données de santé à des fins de réutilisation en Europe (partie 1) ainsi qu'aux Etats-Unis et au Royaume-Uni (partie 2).

1 Les mécanismes d'accès aux données de santé à des fins d'utilisation secondaire en Europe

D'après les éléments transmis par le Health data hub, le pays le plus avancé en matière d'utilisation secondaire des données de santé en Europe, **la Finlande**, a accordé à Findata, un organisme très similaire au HDH, le rôle d'autoriser l'accès et de mise à disposition des données, sans intervention de l'autorité de protection des données. En Finlande, le processus de demande d'autorisation qui en résulte est plus lisible pour tous les acteurs et plus rapide que celui en place en France (environ 6 mois pour accéder aux données en Finlande vs. 18 mois en France) :

- Findata traite la demande d'accès et accorde ou refuse le permis d'accès aux données de santé dans un délai de 3 mois. En réalité, il y a une file d'attente (actuellement d'environ 3 mois) avant que le traitement effectif de la demande ne débute. Une fois en cours, il faut environ 2 mois pour que la demande soit complètement traitée ;
- Par la suite, Findata transmet la demande d'extraction de données au détenteur de données, qui dispose de 30 jours ouvrables pour fournir les données requises ;
- Les données sont collectées et combinées par Findata et mises à disposition dans un environnement de traitement sécurisé finlandais, dans un délai de 60 jours ouvrables.

La Finlande constitue aussi un modèle intéressant en matière de tarification de la mise à disposition des données, avec une grille tarifaire publique, adoptée par décret, décomposant la tarification :

- du permis d'accès aux données,
- des coûts d'extraction et de mise à disposition des données par le producteur de données,
- des coûts liés au processus de pseudonymisation et d'anonymisation des données (avec l'application d'un financement horaire forfaitaire) ;
- et des coûts d'accès à distance à l'environnement sécurisé.

Par ailleurs, dans le cadre de l'action conjointe européenne TEHDAS (« Towards European Health Data Space »), des visites organisées dans une dizaine de pays ont permis de réaliser en 2023 un benchmark¹⁶ sur les temps d'accès et de mise à disposition des données de santé à l'échelle européenne :

- En Belgique, l'accès aux données dépend de la complexité de la demande, mais il est fréquemment rapporté comme étant long, avec des délais s'étendant de 6 à 18 mois.
- En République tchèque, l'accès à certaines données semble généralement se dérouler de manière efficace. Pour l'Institut d'information et de statistiques de la santé, les délais d'accès sont définis par la loi, allant de 15 jours pour des demandes courantes à 30 à 60 jours pour des analyses plus approfondies. Pour les biobanques, la réception des échantillons demandés prend généralement de 2 à 3 mois.
- Au Danemark, le délai moyen d'accès aux données se situe entre 1 et 6 mois.
- En Estonie, le temps requis pour accéder aux données varie en fonction de la complexité des demandes, pouvant s'étendre jusqu'à 6 mois pour les requêtes les plus compliquées.
- En Hongrie, les procédures d'accès aux données varient selon les détenteurs de données. L'EESZT (Infrastructure nationale de santé électronique) est connue pour avoir des procédures d'accès chronophages. En revanche, le NEAK (Fonds national d'assurance maladie) dispose d'une procédure d'accès claire sur son site web. L'accès est principalement accordé à des données agrégées ou anonymisées, tandis que les données pseudonymisées au niveau individuel ne sont accessibles que sur place. Le délai moyen d'accès aux données est de 60 jours.
- Au Portugal, il faut généralement de 3 à 4 mois pour accéder aux données.

Tableau 6 : La prévalence des différents mécanismes d'accès aux données de santé à des fins d'utilisation secondaire en Europe*

Type d'accès	Nombre total d'États membres concernés	États membres concernés
L'accès est accordé après autorisation délivrée par un comité d'éthique de la recherche ou une agence de protection des données	22	BE, CZ, DK, DE, EE, IE, EL, ES, FR, HR, IT, CY, LV, LT, LU, HU, AT, PL, PT, RO, FI, SE, [UK]
Le détenteur des données fournit un accès direct sans intervention d'un comité	7	DK, HR, IT, NL, AT, SI, FI, [UK]

¹⁶ "Mapping health data management systems through country visits: development, needs and expectations of the EHDS",

Towards European Health Data Space, avril 2023 (<https://tehdas.eu/app/uploads/2023/04/tehdas-mapping-health-data-management-systems-through-country-visits.pdf>).

d'éthique ou d'une agence de protection des données		
Un organisme central de gouvernance de l'accès aux données existe dans le pays	13	BG, DK, DE, IE, EL, FR, CY, MT, NL, PT, SI, SK, FI, [UK]

* Certains pays peuvent appliquer différents mécanismes selon le type de données, de recherches ou de chercheurs, et donc apparaître sur plusieurs lignes du tableau

Source : Commission européenne, Direction générale de la santé et de la sécurité alimentaire, Évaluation des réglementations des États membres de l'Union européenne en matière de données de santé à la lumière du RGPD, 2021 (https://health.ec.europa.eu/system/files/2021-02/ms_rules_health_data_en_0.pdf).

2 Les pratiques aux États-Unis et au Royaume-Uni

De nombreux acteurs auditionnés par la mission, font valoir que, si la puissance des bases de données de santé en France, et en particulier du SNDS, est de plus en plus reconnue au niveau européen et international, **la complexité des démarches réglementaires et les délais d'accès continuent d'exercer un effet dissuasif qui peut conduire les équipes de recherche à renoncer à mobiliser les données de santé françaises.**

Sont souvent cités à titre de comparaisons internationales :

- les États-Unis, où un accès direct à des données de santé peut être obtenu en moyenne en une semaine dans le cadre des offres de services proposées par des firmes telles que HealthVerity Marketplace, Optum, Tempus ou encore MarketScan ;
- le Royaume-Uni, où l'accès aux principales bases de données nationales n'est pas subordonné à une autorisation délivrée par une agence de protection des données, mais est délivré à la discrétion de l'institution détentrice de la base, selon un processus d'examen interne qui lui est propre et faisant intervenir une analyse scientifique et éthique des demandes adaptée au degré de complexité de celles-ci. À titre d'exemple :
 - les modalités réglementaires d'accès à la base de données cliniques en soins primaires issues des cabinets médicaux, la « *Clinical Practice Research Datalink* » (CPRD), sont adaptées à la complexité du projet de recherche et permettraient une mise à disposition des données dans un délai de deux mois¹⁷ ;
 - de même, l'octroi d'un accès aux données de la « *UK Biobank* » relève de la responsabilité de son « *Board* », lequel délègue la supervision du processus d'examen des demandes, la mise en œuvre de la procédure d'accès ainsi que l'ensemble des décisions d'accès à son « *Access Sub-Committee* ». La « *UK Biobank Scientific Team* » procède à l'examen scientifique de la demande et l'expertise éthique du centre Ethox, basé à l'université d'Oxford, peut être mobilisée.

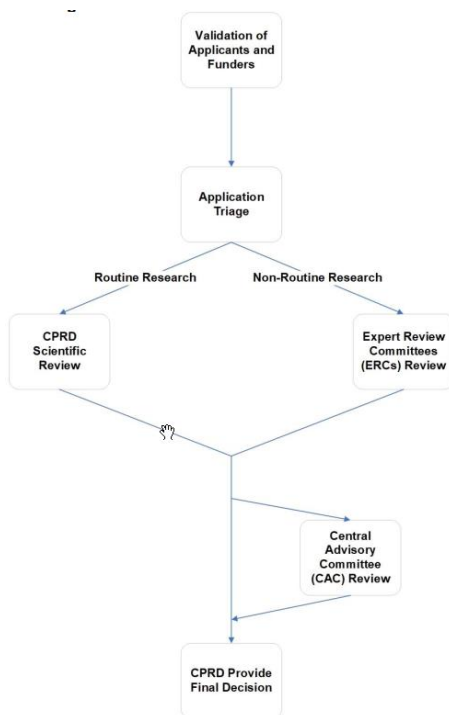
¹⁷ Cf. contribution du SNITEM.

Les modalités d'accès à la base CPRD au Royaume-Uni

Une première étape de validation consiste à s'assurer que l'accès aux données est sollicité par des chercheurs *bona fide* dans le domaine de la recherche en santé publique, financés par des organismes de confiance. Une attention particulière est donc portée au profil du ou des chercheurs et de l'organisme qui les emploie ou les finance.

Suit une étape de tri des demandes par des chercheurs de la CPRD selon qu'elles relèvent de la recherche dite de « routine » (« *routine research* ») ou de la recherche « hors routine » (« *non-routine research* »).

Les projets de recherche les moins complexes, de routine, sont ensuite évalués par des chercheurs de la CPRD. Les projets « hors routine » sont, eux, soumis à l'examen d'un comité d'experts (« *Expert review committee* »). À l'issue de cette évaluation, éventuellement complétée par un examen par le comité central consultatif (« *Central advisory committee* ») si les chercheurs de la CPRD ou le comité d'experts l'ont jugé nécessaire, la CPRD rend sa décision sur la délivrance de l'accès à sa base de données.



Source : CPRD (<https://cprd.com/sites/default/files/2022-02/CPRD%20Research%20Data%20Governance%20Operating%20Framework.pdf>).

ANNEXE 4 : Le Health Data Hub : missions et réalisations

Table des matières

ANNEXE 4 : LE HEALTH DATA HUB : MISSIONS ET REALISATIONS	35
1 UNE MISSION DE MISE A DISPOSITION DES DONNEES DU SNDS SUSPENDUE A LA PROBLEMATIQUE DE L'HEBERGEMENT DE LA COPIE DE CE DERNIER	35
2 UNE MISSION D'ENRICHISSEMENT DU SNDS QUI SUSCITE DE FORTES ATTENTES.....	37
3 UNE MISSION DE GUICHET UNIQUE ET D'ACCOMPAGNEMENT APPRECIEE PAR LES ACTEURS DU SECTEUR	38
4 UNE MISSION D'INFORMATION DES PATIENTS ET DE FACILITATION DE L'EXERCICE DE LEURS DROITS ENCORE A DEVELOPPER	39
5 UNE MISSION D'ANIMATION DE L'ECOSYSTEME A APPROFONDIR	41

Créée par l'article 41 de la loi n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé, dite loi « OTSS », la plateforme des données de santé (PDS), plus connue sous le nom de « Health Data Hub » (HDH), a succédé à l'Institut national des données de santé (INDS), groupement d'intérêt public (GIP), à compter de l'approbation, par un arrêté du 29 novembre 2019, d'un avenant à la convention constitutive de l'INDS et portant création de la PDS.

1 Une mission de mise à disposition des données du SNDS suspendue à la problématique de l'hébergement de la copie de ce dernier

En application du 1° de l'article L. 1462-1 du code de la santé publique (CSP), le HDH est d'abord chargé « *de réunir, organiser et mettre à disposition les données du système national des données de santé (...) et de promouvoir l'innovation dans l'utilisation des données de santé* ». Dans sa nouvelle feuille de route pluriannuelle pour la période 2023-2025, le HDH consacre un de ses quatre axes stratégiques à « Mettre à disposition les données de sa base principale, l'enrichir et faciliter sa réutilisation ».

À l'heure actuelle, en l'absence de transfert d'une copie de la base principale du SNDS sur sa plateforme, le HDH doit se contenter d'intervenir dans le circuit de mise à disposition des données de la base principale du SNDS *via* la caisse nationale d'assurance maladie (CNAM). Les principales opérations, les plus chronophages, préalables à l'extraction (expression des besoins avec validation du périmètre des données demandées, identification et préparation de la

population ciblée, préparation, validation et signature de la convention avec le porteur de projet...) restent de la responsabilité de la CNAM, le cas échéant sur son portail.

Néanmoins, le HDH a fourni un appui à la CNAM dans le cadre de la préparation de données de la base principale du SNDS, s'agissant des ciblages de populations d'étude, des appariements à d'autres bases de données et des extractions de données :

- une partie des extractions de données de la base principale du SNDS est ainsi actuellement prise en charge par le HDH, par la mise à disposition d'1,6 ETP. 192 extractions ont été réalisées en 2022 et 181 depuis janvier 2023 ;
- le HDH est impliqué dans les travaux de ciblage de populations d'étude et les appariements aux données de la base principale du SNDS, principalement pour les projets réalisant leur étude sur la plateforme du HDH mais aussi pour des projets se faisant sur le portail de la CNAM (de manière plus ponctuelle), en y consacrant 2 ETP, ce qui a correspondu, en termes de chiffres, depuis le début de l'année 2023, à 5 ciblages, 3 appariements indirects (+ 4 en cours) et plus de 10 appariement directs.

S'il peut intervenir en appui pour la réalisation des opérations précitées, le HDH n'effectue pas à proprement parler la mise à disposition des données de la base principale du SNDS qui continue d'être assurée par la CNAM. Toutefois, afin de fluidifier l'étape de finalisation du périmètre des données à extraire, le HDH a **conclu une convention avec la CNAM de co-responsabilité de traitement**, permettant au HDH de gérer, par délégation, les extractions et le conventionnement pour les projets hébergés sur sa plateforme.

Le HDH le reconnaît lui-même, « **le circuit actuel de mise à disposition des données de la base principale du SNDS via la CNAM demeure loin de la vision cible** ». En janvier 2022, le HDH a retiré « temporairement » la demande qu'il avait déposée auprès de la CNIL pour opérer le transfert d'une copie de la base principale du SNDS sur la plateforme Azure.

Néanmoins, le HDH a réussi à développer une plateforme à l'état de l'art répondant aux besoins de calcul et de stockage d'un certain nombre de projets et qui offre des espaces d'analyse pour la réalisation des projets. Sur les 97 projets accompagnés par le HDH après avoir été sélectionnés dans le cadre de ses appels à projets, les deux tiers (68 projets) ont fait le choix d'être hébergés sur sa plateforme technologique. L'autre tiers des projets accompagnés n'ayant pas fait ce choix se décompose en projets s'appuyant sur la plateforme de la CNAM (11 projets), sur des infrastructures non adaptées au traitement de données pseudonymisées (8 projets), sur d'autres systèmes fils existants (8 projets) et sur des systèmes internes des responsables de traitement (10 projets).

En termes de délais d'ingestion des données sur sa plateforme, ceux-ci dépendent de la complexité et de la volumétrie des données en jeu. Néanmoins, s'agissant des cas les plus standard, l'ingestion des données sur la plateforme requiert en moyenne 17 jours, avec 12 jours comme minimum et 31 jours comme maximum, pour des ingestions d'un volume moyen de 4,2 téraoctets.

2 Une mission d'enrichissement du SNDS qui suscite de fortes attentes

Le HDH est chargé sur le plan opérationnel, en vertu du 1^o du III de l'article R. 1461-3 du CSP, d'enrichir la base principale du SNDS par une série de données limitativement énumérées, le comité stratégique des données de santé étant chargé, pour sa part, d'identifier les composantes futures de cette base principale :

- l'échantillon représentatif des données de remboursement par bénéficiaire transmises par des organismes d'assurance maladie complémentaire ;
- les données recueillies à l'occasion d'activités de prévention, de diagnostic, de soins ou de suivi social et médico-social. À ce titre, le HDH participe à l'effort de « médicalisation » de la base principale du SNDS par l'apport de données cliniques, en accompagnant notamment les lauréats de l'appel à projets pour la constitution d'entrepôts de données de santé hospitaliers et un projet de constitution d'un entrepôt de données de médecine de ville, dans le cadre du consortium « P4DP »¹⁸, financé à hauteur de 11 millions d'euros par l'État ;
- les données relatives à la perte d'autonomie ;
- les données à caractère personnel des enquêtes dans le domaine de la santé ;
- les données recueillies à l'occasion des visites et dépistages obligatoires dans l'éducation nationale ;
- les données recueillies par les services de la protection maternelle et infantile ;
- les données issues des dossiers médicaux en santé au travail.

Cette mission d'enrichissement de la base principale du SNDS, qui doit permettre de compléter cette base essentiellement médico-administrative par des données médicalisées, sociales ou socio-médicales, est très attendue du monde de la recherche. Toutefois, si le versement sur une plateforme centralisée de bases médicalisées complémentaires au SNIIRAM, au PMSI et au CépiDC ne pose pas de difficulté dans son principe, la centralisation du stockage d'une copie des bases de données issues des EDS hospitaliers crispe une partie des acteurs du secteur, qui y voient une logique de centralisation de l'ensemble des bases de données de santé, contraire au principe de subsidiarité dans la constitution de telles bases.

Il convient de souligner que, dans l'absolu, ne peuvent venir alimenter la base principale du SNDS, conformément à l'article R. 1461-2 du CSP, que les traitements qui couvrent l'ensemble de la population, étant entendu que l'identification des composantes futures de la base principale relève du comité stratégique des données de santé. Dans ces conditions, les traitements qui ne remplissent pas cette condition, ont d'abord vocation à être versés dans le catalogue du SNDS. Ainsi, un arrêté du 12 mai 2022 est venu compléter la base principale du SNDS par les bases de

¹⁸ Dont fait partie le HDH, aux côtés du Collège national des généralistes enseignants (CNGE), le CHU de Rouen, l'université de Nice, l'université de Normandie et l'OAMIX.

données vaccinales et de dépistage « Vaccin-covid » et « SI-DEP », et a listé dix bases de données constitutives du catalogue du SNDS.

Les dix bases de données constituant le catalogue du SNDS

1° La base de données relative au dépistage du cancer du sein des départements du Gard et de la Lozère, dénommée « base E-SIS » ;

2° La base de données relative à l'enquête statistique sur l'état de santé et les conditions de vie en lien avec l'épidémie due au coronavirus, dénommée « EpiCov » ;

3° La base de données relative à la banque nationale des maladies rares centralisant les dossiers patients informatisés créés par les centres experts, dénommée « BNDMR » ;

4° La base de données relative aux données de surveillance de 33 maladies à déclaration obligatoire permettant de prévenir les risques d'épidémie, dénommée « Maladies déclaration obligatoire » ;

5° La base de données relative à l'accès précoce de 1402 patients atteints de cancer bronchique à petites cellules traités par TECENTRIQ en association au carboplatine et à l'étoposide en première ligne de traitement de stade étendu et présentant un score ECOG de 0 ou 1, dénommée « ATU CBPC atezolizumabR » ;

6° La base de données relative aux patients atteints d'hépatite B ou C, dénommée « HEPATHER » ;

7° La base de données relative à l'exploitation des données de passages aux urgences, dénommée « OSCOUR » ;

8° La base de données relative à l'entrepôt de données structurées issues des dossiers médicaux de patientes atteintes d'un cancer du sein métastatique prises en charge dans des Centres régionaux de lutte contre le cancer, dénommée « ESME CSM » ;

9° La base de données relative à une étude sur l'histoire naturelle de la maladie d'Alzheimer, dénommée « MEMENTO » ;

10° La base de données relative aux infarctus du myocarde pris en charge par les services d'aide médicale urgente (SAMU) et les structures mobiles d'urgences et de réanimation (SMUR) d'Ile-de-France, dénommée « registre E-MUST ».

Conformément à l'article 4 de l'arrêté précité, les responsables des bases de données alimentant le catalogue du SNDS les conservent dans leurs propres SI sous leur responsabilité. La copie de ces bases n'est transférée pour stockage par le HDH qu'après la signature d'une convention entre ce dernier et le responsable de la base concernée.

3 Une mission de guichet unique et d'accompagnement appréciée par les acteurs du secteur

Le dernier alinéa de l'article 76 de la loi informatique et libertés charge le HDH d'assurer le guichet unique des dossiers de traitement de données de santé à des fins de recherche n'impliquant pas la personne humaine et de les orienter vers les instances compétentes. Cette mission de guichet

unique et d'accompagnement des porteurs de projets est saluée par l'ensemble des acteurs de la recherche entendus par la mission.

Le HDH s'investit ainsi fortement dans la préparation du dossier réglementaire avant sa soumission au CESREES, qui prend en moyenne deux mois. Afin de sécuriser le dossier, le HDH procède notamment à la relecture du protocole et de la note d'information, ainsi que de l'analyse d'impact relative à la protection des données (AIPD). Il assiste également les porteurs de projets dans les réponses aux questions de la CNIL, dans l'expression des besoins en données du SNDS, dans la définition de la stratégie d'information des patients et d'exercice par ces derniers de leurs droits. Le HDH s'emploie également à orienter les projets qui le justifient vers une déclaration de conformité à une méthodologie de référence.

Une fois les prérequis réglementaires validés, les porteurs de projets sont ensuite accompagnés par le HDH pour commencer à travailler sur la plateforme de ce dernier, lequel assure l'ingestion des données extraites par la CNAM et déploie l'espace d'analyse.

Le HDH a en outre déployé un catalogue de formations à destination des porteurs de projets, notamment des formations au SNDS et aux données qu'il contient ou encore des formations sur les procédures réglementaires d'accès aux données de santé.

Le HDH tient également un répertoire public de l'ensemble des projets de traitement n'impliquant pas la personne humaine qui lui sont déposés, que ces projets nécessitent une autorisation de la CNIL ou bénéficient d'une procédure d'accès simplifié aux données de santé.

Le HDH assure enfin le secrétariat du CESREES.

4 Une mission d'information des patients et de facilitation de l'exercice de leurs droits encore à développer

Conformément à la mission qui lui incombe en application du 2° de l'article L. 1462-1 du CSP, le HDH a fait de l'accompagnement des patients dans l'exercice de leurs droits une priorité.

Outre des actions à destination des citoyens de sensibilisation à l'utilisation secondaire de leurs données de santé et d'acculturation au SNDS, conduites par sa direction de la participation citoyenne, le HDH met en œuvre un formulaire national d'exercice des droits permettant aux patients d'exercer leurs droits, notamment d'opposition, relatifs à l'utilisation de leurs données dans le cadre de la base principale du SNDS et des bases du catalogue.

Le HDH envisage d'accompagner ce formulaire par « *un espace d'information et de sensibilisation sur la seconde vie des données de santé et sur les projets utilisant des données afin que les droits soient exercés de façon éclairée* ».

Par ailleurs, en termes de transparence vis-à-vis des patients sur l'utilisation secondaire de leurs données de santé, le HDH a annoncé l'enrichissement de son répertoire public des projets qui lui sont déposés avec les résultats de ces projets.

Le HDH s'investit fortement dans la sensibilisation des citoyens et des responsables de traitement aux règles entourant l'usage des données de santé et à l'exercice par les personnes concernées de leurs droits. À cet effet, il convient de relever les actions suivantes :

Les actions de sensibilisation du HDH auprès des citoyens sur l'exercice de leurs droits

1° La construction d'un portail pédagogique d'information pour fournir aux citoyens toutes les réponses nécessaires pour prendre des décisions éclairées lors de l'exercice des droits ou simplement lors de la recherche d'informations, et éviter que les choix relatifs à l'exercice des droits ne se fassent par défaut ou par manque d'information. Ceci correspond à une demande des citoyens eux-mêmes (baromètre du HDH et de France Assos Santé¹⁹, rapport Tehdas²⁰). Ce portail est porté par deux comités : comité de pilotage composé de la CNIL, la DREES et de France Assos Santé ; comité de citoyens (comprenant notamment des représentants d'associations de patients et d'étudiants).

2° La publication régulière de documents et d'information, et animation de formations sur l'exercice des droits, destinés aux citoyens. L'objectif est surtout de fournir des informations dans un langage accessible pour expliquer que les données de santé pour la recherche à des fins d'utilité publique respectent un cadre légal, et informer sur leurs droits. Par exemple :

- une note d'engagement²¹ du HDH vis-à-vis des données de santé des citoyens et une note "vos droits²²" pour les connaître ;

- des modules de formations citoyennes sur le SNDS et les entrepôts de données qui abordent le sujet des droits des patients, dont un prévu en 2024 sur la différence entre consentement / information individuelle et collective. Ces notes sont complémentaires, co-construites avec les citoyens et représentants d'associations, écrites dans un langage accessible et pédagogique, et téléchargeables sur plusieurs pages du site du HDH pour les trouver facilement ;

- la **publication systématique des projets au sein du répertoire public du HDH**, ainsi que des bases au sein du catalogue des données, avec les caractéristiques des traitements de données associés et les modalités qui leur sont offertes pour exercer leurs droits auprès des responsables de traitement et responsables de données.

3° Les publications régulières de documents et l'animation d'ateliers et de formations sur l'exercice des droits, destinés aux responsables de traitements (voire, aux responsables de données). L'objectif est de leur fournir un appui qui se matérialise de la façon suivante :

¹⁹ "Lancement du premier baromètre de connaissances sur les données de santé et leur usage secondaire", Health Data Hub, 9 novembre 2022 (<https://www.health-data-hub.fr/actualites/lancement-du-premier-barometre-de-connaissances-sur-les-donnees-de-sante-et-leur-usage>).

²⁰ 3 "TEHDAS outlines key issues and considerations on data altruism, Towards European Health Data Space, 5 décembre 2023.

lien : <https://tehdas.eu/results/tehdas-outlines-key-issues-and-considerations-on-data-altruism/>

²¹ "Nos engagements vis-à-vis des citoyens", Health Data Hub, 2020. Lien :

https://www.health-data-hub.fr/sites/default/files/2021-04/Engagements_Citoyens.pdf

²² "Données de santé, mes droits", Health Data Hub, 2020. Lien :

https://www.health-data-hub.fr/sites/default/files/2021-04/Mes_droits.pdf

- Des ateliers réguliers : (i) de design d'affiches²³ (pour co-construire des affiches (dans les salles d'attente) pour accompagner les responsables des entrepôts de données de santé à faire de l'information adaptée et les sensibiliser sur les droits) ; (ii) d'information et exercice des droits (pour leur présenter leurs obligations et la manière dont ils peuvent les respecter, en lien avec les spécifications techniques de la plateformes) ;
- Des documents de conseils juridiques et organisationnels : ils permettent de définir les modalités de gestion des droits en lien avec les responsables de traitement, notamment les droits applicables et leur périmètre, les rôles et responsabilités de chaque acteur (HDH, responsable de traitement, responsable de données) dans le processus.
- La répartition des rôles des différentes parties prenantes en matière de gestion des droits. En la matière, une convention de co-responsabilité est en cours de finalisation entre la CNAM et le HDH de manière à assurer la bonne prise en compte des demandes d'exercice des droits portant sur les données du SNDS.

Source : *Éléments transmis par le HDH.*

5 Une mission d'animation de l'écosystème à approfondir

En vertu du 7° de l'article L. 1462-1 du CSP, le HDH est chargé « *de contribuer à diffuser les normes de standardisation pour l'échange et l'exploitation des données de santé, en tenant compte des standards européens et internationaux* ».

En matière de standardisation, bien qu'il n'ait pas encore élaboré de doctrine, le HDH a d'abord concentré ses efforts sur les données de la base principale du SNDS dont les terminologies propres à la France et les schémas de données non standard rendent difficile leur exploitation dans le cadre de projets de recherche au niveau européen.

Dans le cadre d'appels à projets organisés par le réseau EHDEN (« *European Health Data and Evidence Network* », branche européenne du réseau OHDSI, responsable d'OMOP-CDM) et participant au financement de l'omopisation de bases de données européennes, le HDH s'est ainsi attaché à l'« omopisation » de la base principale du SNDS, permettant une réduction notable du nombre de ses tables, passées de 180 (plus de 4 500 variables) à aujourd'hui 14 tables au format OMOP-CDM. Le HDH souligne que cette transformation a permis de réduire par cinq la volumétrie des données de la base principale du SNDS, de garder uniquement les informations principales sur le parcours de soin du patient, améliorant la compréhension de cette base par des personnes non expertes, et de centrer les données autour du patient, favorisant ainsi l'étude des parcours de soins.

Le HDH a également procédé à une étude comparative des standards de données de santé utilisés en France et à l'international dans l'optique d'ouvrir à tout l'écosystème une documentation à ce

²³ Des affiches dans les salles d'attente des établissements de soins pour sensibiliser aux données de santé : quel premier impact ?", Health Data Hub, 05 septembre 2023. Lien : <https://www.health-data-hub.fr/actualites/des-affiches-dans-les-salles-dattente-des-etablissements-de-soins-pour-sensibiliser-aux>

sujet. Les standards actuellement couverts par cette étude sont : i2b2, OMOP-CDM, FHIR, SNOMED-CT, Beacon, CDA, OSIRIS, CIM-11, LOINC, ICHI.

Il est à noter, en outre, que le comité stratégique des données de santé, après avoir conduit des travaux sur le socle commun de données des entrepôts de données de santé hospitaliers, devrait très prochainement mettre en place un groupe de travail chargé de formuler des propositions pour accompagner les acteurs dans le choix d'un standard à appliquer, dans un premier temps, aux données de ce socle.

Enfin, au titre du 8° de l'article L. 1462-1 du CSP, le HDH est également chargé d'accompagner, notamment financièrement, les porteurs de projets sélectionnés dans le cadre d'appels à projets lancés à son initiative et les producteurs de données associés aux projets retenus. Dans ce cadre, il a mené, en trois ans, un total de 14 appels à projets pour favoriser la recherche et l'innovation grâce aux données de santé.