



HEALTH
DATA HUB

Rapport annuel 2023

SOMMAIRE

PRÉFACE

Gérard Raymond 4

ÉDITORIAL

Stéphanie Combes 5

PARTIE 1

LE HEALTH DATA HUB, UNE INFRASTRUCTURE CLÉ POUR LA RECHERCHE EN SANTÉ

Les quatre grandes missions du Health Data Hub 7

La gouvernance et l'équipe du Health Data Hub 8

Les temps forts et chiffres clés 2023
du Health Data Hub 10

PARTIE 2

LES AVANCÉES 2023 DE LA FEUILLE DE ROUTE PLURIANNUELLE 2023-2025 DU HEALTH DATA HUB

Enjeu 1

Réduire les délais d'accès aux données de santé et multiplier les projets impactants 14

Simplifier et clarifier les procédures de création
d'accès et de réutilisation des données 15

Mettre à disposition des données d'intérêt
et de qualité 18

Soutenir les projets 22

Mutualiser les connaissances et les expertises 27

Renforcer l'attractivité du Health Data Hub 31

Enjeu 2

Mettre à disposition les données de la base principale du SNDS, l'enrichir et faciliter sa réutilisation 35

Développer l'expertise du Health Data Hub
dans la mise à disposition des données 36

Faciliter l'exploitation de la base principale du
SNDS 37

Contribuer à enrichir la base de données
de la base principale 41

Maintenir la plateforme à l'état de l'art
fonctionnellement et en matière de sécurité 42

Enjeu 3

Renforcer les connexions du Health Data Hub aux acteurs de l'écosystème 45

Renforcer le travail partenarial
avec les acteurs locaux 46

Renforcer la présence du Health Data Hub
au niveau européen et international 51

Réunir les conditions d'un partage efficace
en termes de standards, gouvernance d'accès
et de valorisation dans un contexte national
et européen 56

Réunir les conditions d'un partage efficace
en termes de normes de sécurité, de souveraineté
dans un contexte national et européen 58

Enjeu 4

Être à l'écoute de la société civile et co-construire une culture de la donnée de santé 60

Informier largement sur les enjeux
de la réutilisation des données de santé 61

Écouter et impliquer la société civile 63

Développer les partenariats avec
les associations d'utilisateurs 65

Accompagner les personnes concernées
dans l'exercice des droits 67

PARTIE 3

BILAN DE L'ACCÈS AUX DONNÉES 68

Un accès réglementé 69

L'accès en quelques chiffres (chiffres 2023) 71

Une offre de services destinée aux porteurs de projets 72

L'usage des données de santé et transparence 73

Gérard Raymond,
Vice-Président
du Health Data Hub



La France a été précurseur de l'exploitation secondaire des données de santé à travers de nombreuses initiatives depuis plus de 20 ans : la création du Sniiram en 1999, élargi en Système National des Données de Santé (SNDS) en 2016, et auquel l'Assurance Maladie a ouvert davantage l'accès et la mise à disposition. Puis, la loi a élargi le périmètre des données de santé partageables, notamment les données de certaines cohortes, registres et entrepôts ; tout en encourageant les projets en intelligence artificielle (IA) qui peuvent notamment recourir à l'offre de service du Health Data Hub (HDH), créé à cet effet. L'Assurance Maladie et le HDH ont ainsi largement oeuvré, en lien avec les autres acteurs de l'écosystème et ses partenaires, à poser les fondations pour permettre une exploitation plus importante du patrimoine de données. Plus précisément, il y a déjà 10 ans, un grand nombre d'acteurs dont les associations appelaient à "libérer les données". Mais pour exploiter de manière impactante les données, il faut plus que simplement les libérer ! Il faut les produire, les documenter et les standardiser pour les rendre utilisables. Ensuite, il faut les croiser avec différentes sources, car seules, elles sont souvent insuffisantes. Même si l'on souhaite progresser rapidement, compte tenu des enjeux pour la médecine et les patients, mais aussi de la compétitivité internationale, il est complexe de mettre en oeuvre tous ces chantiers à la fois. De fait, la création du Sniiram il y a plus de vingt ans, des grandes cohortes et plus récemment des entrepôts hospitaliers, sont autant de briques posées au fil du temps, et auxquelles le HDH a continué de contribuer en 2023.

Ainsi, le bilan de l'année 2023 est que le HDH a contribué au renforcement et à la solidification pour construire un édifice porteur de sens pour les patients, aux bénéfices déjà visibles. Les bénéfices des projets d'intérêt public ne se limitent pas aux résultats finaux, mais incluent également le temps de recherche et de développement de ces projets, y compris des phases telles que l'appariement des données et les tests d'utilisation des applications développées. Ce rapport annuel 2023 est l'occasion de rappeler l'importante créativité et innovation dont sont capables les équipes qui portent les projets d'intérêt public pour utiliser les données et les chaîner ; mais aussi de rappeler que les projets franchissent progressivement des jalons, et l'apport du HDH pour les y aider, très souvent en appui des équipes de l'Assurance Maladie.

Stéphanie Combes,
Directrice
du Health Data Hub



La France dispose de l'une des bases de données médico-administratives les plus importantes au monde. En exploitant ses données, la France a le potentiel pour faire progresser collectivement le système de santé. C'est dans ce cadre, et pour prolonger des initiatives déjà mises en oeuvre, que le HDH a fait son entrée dans le paysage, fin 2019. Objectif : accélérer, simplifier, unifier et sécuriser l'accès aux données de santé, en toute transparence, notamment grâce à l'IA. En quatre ans, les usages de ces données de santé se sont multipliés, avec des premiers résultats pour les projets accompagnés ; le périmètre de ces données s'est élargi et s'est ouvert à des gisements de données en France. Ce projet d'ouverture des données nécessitait à la fois de relever des défis techniques et, pour beaucoup, de s'accoutumer à l'usage et au partage des données de santé. C'est pourquoi, l'ambition du HDH était multiple : fédérer et animer les acteurs de l'écosystème ; offrir une plateforme de services très étendue permettant d'accompagner les porteurs de projets de l'idée jusqu'aux résultats ; et susciter et co-construire une culture partagée de la donnée de santé. En quatre ans, le HDH s'est aussi inscrit dans le paysage européen, notamment en tant qu'autorité compétente dans l'action conjointe TEHDaS "Vers un espace européen des données de santé".

La nouvelle feuille de route 2023-2025, co-construite avec l'écosystème, a permis de définir les nouvelles priorités : appels à projets, appui aux entrepôts, contribution au Comité stratégique des données, espace européen des données de santé (EHDS) et initiatives citoyennes. Ainsi, 2023 a été l'année du renforcement des nombreuses initiatives démarrées. Les travaux issus de la gouvernance stratégique pour formuler la vision à moyen-long terme du patrimoine national des données de santé et les règles du jeu du partage se sont poursuivis et multipliés. Sur un plan plus opérationnel, des travaux ont continué d'être mis à la disposition des publics avec une offre de formations et des guides, variés et complémentaires ; ou encore une offre pour accompagner les entrepôts de données hospitaliers, en articulation avec le ministère du Travail, de la Santé et des Solidarités. Enfin, des travaux contribuant à la structuration d'un paysage de données à l'échelle européenne se sont intensifiés.

Les efforts de 2023 permettront d'engager une année 2024 de façon prometteuse, avec la matérialisation des ponts entre les différentes structures, que ce soit au niveau local, national ou encore européen. Une étape qui sera décisive pour l'éclosion du dispositif à son plein potentiel.



1

PARTIE

LE HEALTH DATA HUB, UNE INFRASTRUCTURE CLÉ POUR LA RECHERCHE EN SANTÉ

Les quatre grandes missions du Health Data Hub

Créé par la loi du 24 juillet 2019 relative à l'organisation et la transformation du système de santé, le Health Data Hub (HDH) est une structure publique permettant aux acteurs d'accéder facilement à des données de santé non nominatives, hébergées sur une plateforme sécurisée. Les porteurs de projets peuvent y croiser les données et les analyser pour améliorer la qualité des soins et l'accompagnement des patients.

Pour ce faire, le HDH assure quatre missions :

- **Offrir un guichet unique** pour accompagner les porteurs de projets dans leurs démarches administratives. Ces projets sont expertisés par le Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (CESREES). Ce dernier émet un avis avant que les projets soient soumis à la Commission Nationale de l'Informatique et des Libertés (CNIL) pour autorisation, ou lorsque les projets nécessitent uniquement son expertise dans le cadre de procédures simplifiées.
- **Ouvrir l'accès à un catalogue de bases de données partenaires.** Construit de manière itérative, le contenu de ce catalogue est porté par un arrêté pris après avis de la CNIL. Les bases du catalogue sont répliquées dans la plateforme technologique pour pouvoir être croisées avec d'autres bases de données, en accord avec les acteurs à l'origine de ces données.
- **Fournir une plateforme sécurisée** à l'état de l'art, offrant des capacités avancées de stockage, de calcul, de rapprochement et d'analyse des données.
- **Animer l'écosystème pour accélérer l'innovation** en favorisant le partage d'expériences et de connaissances.

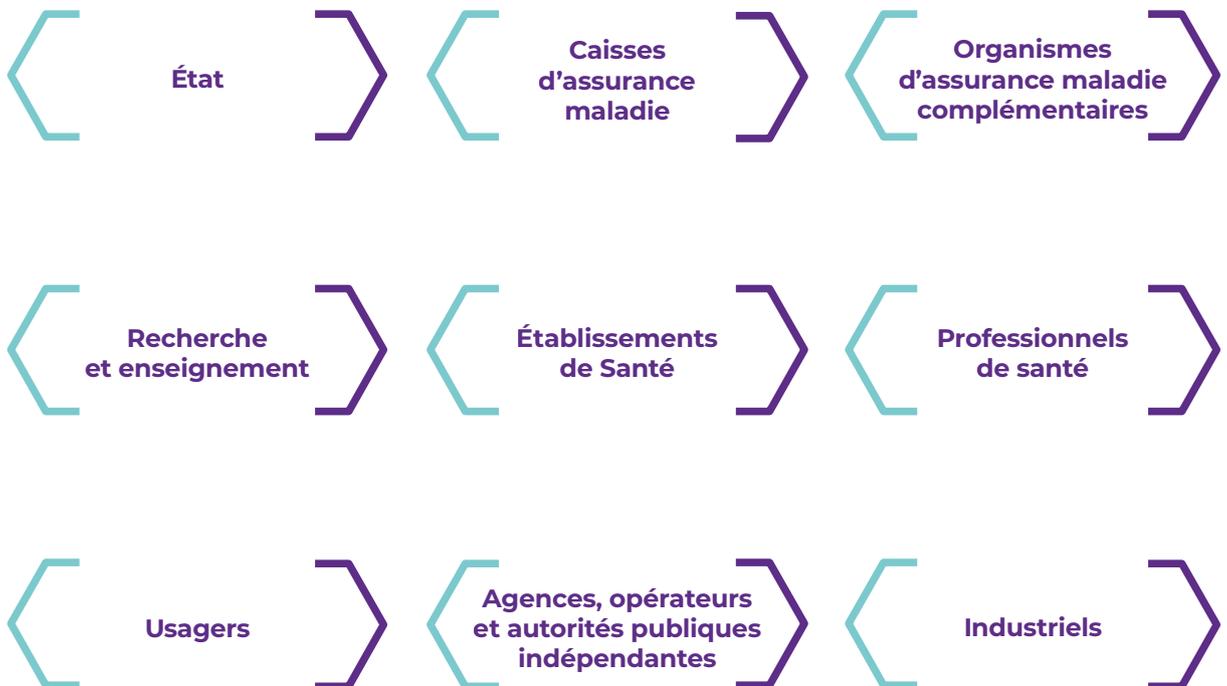
L'offre de service du HDH est complémentaire de celle que peuvent offrir les acteurs publics ou privés qui administrent les bases de données administratives, registres, cohortes, ou enquêtes. Elle répond aux besoins de deux types de projets :

- Les projets visant à croiser différentes sources de données, lorsque ces croisements, chaînages ou appariements sont complexes à réaliser.
- Les projets qui nécessitent d'importants volumes de données pour que les méthodes statistiques utilisées puissent être considérées comme fiables.

Dans les deux cas, la plateforme technologique du HDH agit comme un opérateur chargé de réunir de manière sécurisée les différentes sources de données et permettre qu'elles soient exploitées avec des technologies à l'état de l'art.

La gouvernance et l'équipe du Health Data Hub

Le HDH est un groupement d'intérêt public, créé par la loi du 24 juillet 2019 relative à l'organisation. Son financement est principalement public. Il a ainsi été doté de 12,610 millions d'euros pour l'exercice 2023, provenant de l'objectif national des dépenses d'assurance maladie (ONDAM). Il réunit 56 membres, répartis en neuf collèges : l'État, les caisses d'assurance maladie, les organismes d'assurance maladie complémentaires, la recherche et l'enseignement, les établissements de santé, les professionnels de santé, les agences, opérateurs et autorités publiques indépendantes, les usagers du système de santé et les industriels. L'assemblée générale fixe les grandes orientations du programme de travail annuel du HDH.





Un représentant de chaque collège siège au conseil d'administration, sauf pour l'État qui en a deux : la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES), pour le ministère du Travail, de la Santé et des Solidarités ainsi que la Direction générale de la recherche et de l'innovation (DGRI), pour le ministère de l'Enseignement supérieur et de la Recherche. La Caisse nationale de l'Assurance Maladie (CNAM), l'Union Nationale des Complémentaires Santé (UNOCAM), l'Institut national de la santé et de la recherche médicale (Inserm), le Centre Hospitalier Universitaire (CHU) de Toulouse, l'Union Nationale des Professionnels de Santé (UNPS), France Assos Santé (FAS), Santé publique France (SpF) et l'Association française des Clinical Research Organisations (AFCROs) en sont également membres. L'assemblée générale et le conseil d'administration sont présidés par Gérard Raymond, également président de France Assos Santé.

Depuis janvier 2021, le HDH s'est doté d'un conseil scientifique consultatif (CSC), composé de 28 experts, dont sept experts internationaux, aux compétences complémentaires : médecins, pharmaciens, spécialistes de intelligence artificielle (IA), juristes et spécialistes de l'éthique, généticiens, épidémiologistes, économistes de la santé,

etc. Depuis septembre 2022, le conseil est présidé par la généticienne et épidémiologiste Ségolène Aymé. Le Pr Bertrand Lukacs, chirurgien urologue, spécialiste des données en urologie et cancérologie, en est le vice-président. Ce conseil émet des avis pour éclairer les décisions de la direction, dans un contexte international hautement compétitif.

L'équipe du HDH est animée par l'envie commune d'accompagner les acteurs de la recherche et de l'innovation à mener à bien leurs projets. Au 31 décembre 2023, le HDH comptait 95 collaborateurs, dont 59 femmes (soit 62,1 %) travaillant dans des équipes métier qui accompagnent les projets de bout en bout (le guichet, la direction des partenariats, la direction des données, la direction technique et le produit, la direction des projets et services utilisateurs), dans des directions stratégiques (la direction scientifique, la direction médicale et la direction citoyenne) ou des fonctions support (la direction de la sécurité des systèmes d'information, le secrétariat général, la direction de la communication, la direction de la déléguée à la protection des données (DPO) et la direction juridique).

Les temps forts et chiffres clés 2023 du Health Data Hub

100 PROJETS INNOVANTS
dont...

41 projets font appel à
l'intelligence artificielle

8 projets pour le programme
Unibase - Unicancer et Ligue
contre le cancer

10 projets relatifs à l'épidémie
de la Covid-19

19 projets portant sur la
prévention en santé

4 projets de recherche en
santé-environnement

15 projets visant à constituer
une Bibliothèque Ouverte
d'Algorithmes en Santé
(BOAS)

UNE CONTRIBUTION À L'ENRICHISSEMENT DU PATRIMOINE DE DONNÉES

- 1** projet d'entrepôt national de données de médecine de ville (P4DP)
- 10** bases de données constituent le catalogue du SNDS et **24** autres sont pressenties pour le rejoindre
- 1** entrepôt multicentrique EMC2 intégrant des données de la base principale du SNDS et des données issues de **4** établissements de santé

UN RAYONNEMENT INTERNATIONAL

- 4** projets internationaux sélectionnés avec une participation du HDH : Quantum, Direct Grant, Invents, Darwin EU
- 4** projets issus de l'AAP France-Québec
- 300** participants à la Summer School AI4Health

UNE OUVERTURE À LA SOCIÉTÉ

- 1** conférence de consensus avec **70** citoyens
- 2** groupes de dialogue pérennes avec les associations de patients et d'étudiants
- 5** programmes de formations citoyennes
- 3** Data Challenges lancés
- 176** interventions du HDH à des événements nationaux et internationaux
- 20** articles universitaires publiés

PARTIE 2

**LES AVANCÉES
2023 DE LA
FEUILLE DE ROUTE
PLURIANNUELLE
2023-2025 DU
HEALTH DATA HUB**

En juin 2022, le HDH a publié sa feuille de route stratégique. Elle s'articule autour de quatre orientations stratégiques :

Dans ce rapport annuel 2023, le HDH présente un état d'avancement sur les quatre orientations stratégiques définies dans sa feuille de route pluriannuelle.

1 Réduire les délais d'accès aux données de santé et multiplier les projets impactants

2 Mettre à disposition les données de la base principale, l'enrichir et faciliter sa réutilisation

3 Renforcer les connexions du Health Data Hub aux acteurs de l'écosystème

4 Être à l'écoute de la société civile et co-construire une culture de la donnée de santé

ENJEU

1

Réduire les délais d'accès aux données de santé et multiplier les projets impactants



janv	\$ 31 555,00	\$ 82 710,00	\$ 4 139,00
févr	\$ 56 032,00	\$ 43 665,00	\$ 37 128,00
mars	\$ 68 726,00	\$ 34 549,00	\$ 52 101,00
avr	\$ 27 982,00	\$ 15 001,00	\$ 7 307,00
mai	\$ 21 764,00	\$ 9 822,00	\$ 60 496,00
juin	\$ 53 225,00	\$ 30 359,00	\$ 29 905,00
juil	\$ 16 477,00	\$ 27 178,00	\$ 92 545,00
août	\$ 47 572,00	\$ 18 918,00	\$ 42 796,00
sept	\$ 41 374,00	\$ 39 266,00	\$ 11 922,00
oct	\$ 71 259,00	\$ 82 336,00	\$ 43 820,00



Simplifier et clarifier les procédures de création d'accès et de réutilisation des données

Baromètre de pilotage des délais d'accès

Le HDH a lancé en septembre 2023 un **baromètre poursuivant trois objectifs** : avoir une meilleure connaissance de l'état d'avancement des projets, identifier les points d'attention susceptibles de freiner leur mise en œuvre, et préciser la valorisation des résultats qui en a été faite le cas échéant. Il s'agit de disposer d'indicateurs sur le devenir des projets nécessitant une autorisation de la CNIL ou mis en œuvre dans le cadre d'une procédure d'accès simplifiée. Les résultats issus de ce baromètre permettront de proposer des pistes d'action en lien avec les acteurs concernés, par exemple eu égard à la mise en œuvre de leur traitement. Au total, pour cette première édition, **près de 2 000 porteurs de projets à l'initiative de 5 500 études** ont été invités à renseigner ce baromètre. Il a vocation à être reconduit tous les ans permettant de suivre l'évolution de différents indicateurs dans le temps.

Près de
540
projets ont alimenté le
baromètre

Mise en visibilité des données disponibles : le catalogue de métadonnées

Le HDH développe un catalogue de métadonnées. Il permet aux différents acteurs de prendre connaissance du patrimoine de santé français, et ainsi de juger de leur intérêt et des bases listées qui leur seraient utiles pour leurs études. Ceci est rendu possible à travers la mise à disposition d'informations granulaires, relatives aux bases de données des partenaires¹. L'avancée des travaux en 2023 a permis de mettre en ligne une **première version "bêta" du portail**.

Des activités de développement ont été conduites afin d'offrir aux utilisateurs une interface ergonomique, simple d'utilisation et répondant aux attentes des utilisateurs de pouvoir facilement **visualiser les bases de données de santé disponibles** en France. À terme, ce catalogue a pour objectif d'être disponible en marque blanche, afin d'ouvrir la possibilité aux autres structures de santé françaises de **créer et maintenir leur propre catalogue** de métadonnées. Aussi, les équipes travaillent au développement d'outils d'ingestion des métadonnées du catalogue, permettant d'alléger l'effort d'alimentation du catalogue.

¹ Ces informations sont disponibles au niveau du jeu de données (ensemble cohérent de bases de données, telle qu'une base structurée patients associée à une base d'images IRM - le jeu de données comprend alors deux bases de données), des bases qui le composent et des composantes de ces bases (par exemple, tables ou variables pour les bases de données structurées).

Travaux d'accompagnement des initiatives d'entrepôts de données de santé

Le HDH accompagne les initiatives d'entrepôts de données de santé (EDS), notamment pour favoriser le partage de bonnes pratiques et assurer l'homogénéisation des procédures au niveau national. Ainsi, en 2023, il a assuré **l'animation de trois groupes de travail** lancés par le comité stratégique des données de santé : sur **la gouvernance**, sur **le financement** et sur **les variables communes de données** à tous les entrepôts.

Le groupe de travail **“Gouvernance d'accès aux données de santé”**, piloté par le Pr Bernard Nordlinger, président du CESREES, a réuni et auditionné divers représentants de bases de données et d'institutions ayant de l'expérience dans la gouvernance d'accès aux données. En 2023 **un kit documentaire**² permettant la mise en place d'un Comité Éthique et Scientifique (CSE) opérationnel et conforme à la réglementation a été mis au point. Il comprend :

- **Une trame de règlement intérieur** : destinée aux responsables de traitement des EDS, elle définit le cadre et les règles de fonctionnement du CSE.
- **Une fiche de recevabilité** : destinée au secrétariat du CSE et de l'équipe opérationnelle de l'EDS, elle permet de s'assurer de la complétude du dossier et de la faisabilité globale de l'étude, d'un point de vue méthodologique et technique.
- **Une fiche d'évaluation** : destinée aux rapporteurs, elle réunit l'ensemble des critères permettant au CSE d'étudier de façon approfondie les objectifs, la méthodologie et la pertinence scientifique, voire l'intérêt public.
- **Un modèle d'avis du CSE** : dans la continuité de la fiche d'évaluation, ce livrable permet de formaliser l'avis rendu sur un projet.

Le groupe de travail **“Financement des bases de données de santé publiques”**, lancé lui aussi par le comité stratégique des données de santé et piloté par Jean-François Lefebvre, directeur général du CHU de Toulouse, visait à garantir le financement durable des bases de données de santé en France. Il a rendu en 2023 **une première note soulignant l'importance stratégique des EDS** hospitaliers et la nécessité de soutenir la création d'un réseau favorisant l'ouverture et le partage des données hospitalières. Pour ce faire, le groupe de travail a recueilli les retours d'expériences de l'écosystème et sélectionné les outils de calcul de redevances de l'AP-HP et du Ouest Datahub pour initier **une phase pilote** débutant au premier semestre 2024³, et visant notamment à obtenir des métriques précises sur la réalité de la contractualisation.

Le groupe de travail **“Ensemble des variables communes à tous les entrepôts”**, co-piloté par la Pr Geneviève Chêne, professeur des universités et praticien hospitalier de santé publique à l'université de Bordeaux, et le Dr Cyrille Delpierre, directeur de recherche à l'Inserm, mandaté en janvier 2023 par le comité stratégique des données de santé, propose **un socle commun de 51 items** ayant vocation à être adopté par l'ensemble des EDS hospitaliers du territoire national. Ce socle permettra de **caractériser l'ensemble des séjours et des patients**, quels que soient les motifs de leur recours aux soins. Le socle est composé principalement des données les plus couramment collectées par les établissements dans le cadre des soins, notamment des données cliniques, médicamenteuses, des données socio-démographiques, ainsi que des résultats biologiques. Ces données sont généralement collectées de manière structurée et fiable au sein des systèmes d'information hospitaliers.

2 Si ces livrables constituent des guides, les structures devront adapter le processus de gestion des demandes à leurs propres contraintes pour aboutir à un fonctionnement cohérent et pérenne.

3 L'objectif de cette phase pilote est d'assurer la cohérence et la comparabilité des résultats en vue de finaliser le modèle commun de valorisation économique des données de santé hospitalières.

Par ailleurs, pour offrir aux acteurs une vision d'ensemble des démarches et tâches à accomplir, une première version d'un kit documentaire a été constituée en 2023. Le **“starter kit EDS”** fournit une collection d'outils opérationnels et pédagogiques pouvant faciliter la mise en œuvre du projet tout au long de son cycle de vie. Il sera enrichi au fur et à mesure de l'avancement des travaux et réflexions nationales autour des entrepôts de données de santé.

De plus, **six sessions collectives** à destination de l'ensemble de l'écosystème des EDS ont été organisées en 2023 sur des thématiques variées : la présentation du projet d'exercice des droits (animation par le HDH), la conformité réglementaire (animation par la CNIL), la gestion de grands projets numériques (animation par la Direction interministérielle du numérique), la présentation des livrables du groupe de travail “Gouvernance d'accès aux données de santé” (animation par le HDH) et la présentation du modèle OMOP par les équipes du HDH.

Espace européen des données de santé : position française et partage d'expérience

Le HDH participe aussi aux réflexions portant sur le **projet de règlement** pour l'espace européen des données de santé (European Health Data Space, dit EHDS)⁴. Il a notamment contribué, en 2023, à l'élaboration de la position française dans le cadre de l'instruction du règlement, en lien avec la Délégation au Numérique en Santé (DNS) et sous la coordination du Secrétariat Général des Affaires Européennes (SGAE).

Aussi, reconnu au niveau européen pour son expertise et son rôle de modèle au sein du futur EHDS, et du fait de son rôle de coordinateur du projet pilote HealthData@EU, le HDH a été sollicité en avril 2023 par les parties prenantes européennes pour **partager son retour d'expérience**. Cela a été réalisé dans le cadre d'un séminaire au siège de la représentation permanente de la France auprès de l'UE, en collaboration avec la DNS, pour présenter le modèle français aux représentants des autres États membres.

4 La Commission européenne a présenté en mai 2022 son projet de règlement pour établir l'EHDS. Après deux ans de négociations, le projet de règlement a été adopté par les États membres le 22 mars 2024, puis par le Parlement européen le 24 avril 2024. La publication du texte au journal officiel est prévue pour l'automne 2024.

Mettre à disposition des données d'intérêt et de qualité

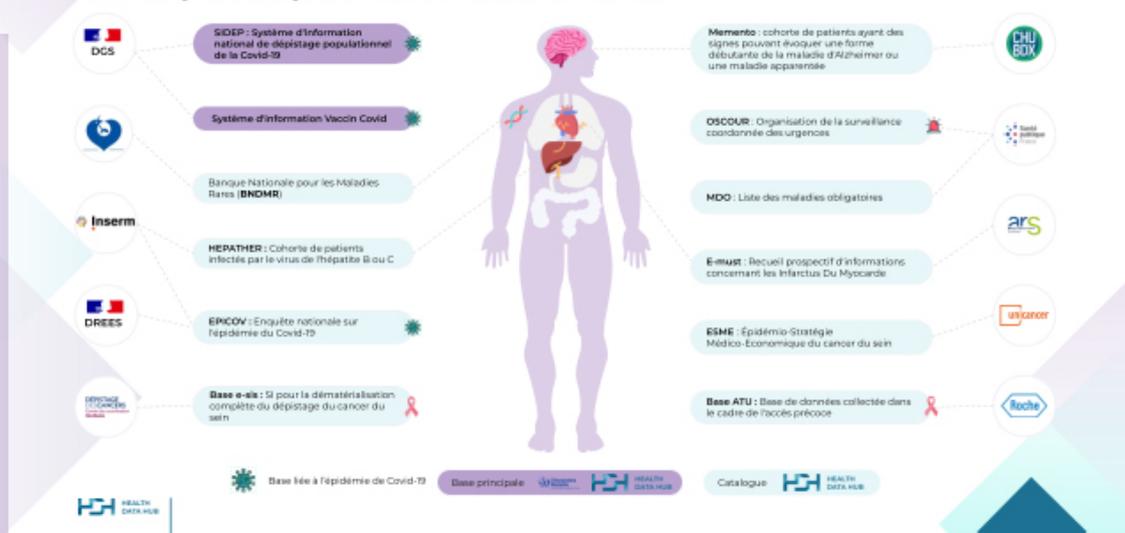
Catalogue : accompagner les responsables de données

Concernant le catalogue de données du SNDS, le HDH a vocation à rassembler des **copies d'une collection de bases**, pour la plupart déjà existantes. Ce travail se réalise de manière progressive et itérative en partenariat avec les responsables de la collecte des données concernées et le comité stratégique des données de santé. Ce dernier approuve le contenu du catalogue, finalement porté par un arrêté pris après avis de la CNIL. Les critères de sélection se fondent sur **la pertinence**

de la base et **son potentiel de valorisation scientifique** (intérêt pour la surveillance épidémiologique, l'évaluation des produits de santé, le pilotage du système de santé, etc) ; **son caractère incontournable** du fait de ses spécificités ; **le périmètre et la qualité** des données ; ou encore **son taux d'actualisation**.

En 2023, le HDH a accompagné les dix responsables de données de la première version du catalogue du SNDS (listés sur l'arrêté du 12 mai 2022) dans sa mise en œuvre opérationnelle.

Première liste des bases de la base principale et du catalogue du SNDS portée par l'arrêté du 12 mai 2022



Cette **mise en œuvre du catalogue du SNDS** a avancé de manière significative, notamment via :

- **La signature de cinq contrats entre les responsables de données et le HDH** : Banque Nationale de Données Maladies Rares (BNDMR) - AP-HP ; ESME - Unicancer ; Organisation de la surveillance coordonnée des urgences (OSCOUR) - SpF ; Autorisation temporaire d'utilisation, Cancer bronchique à petites cellules (ATU CBPC) - Roche ; e-SIS - Centre Régional de Coordination du Dépistage des Cancers (CRCDC Occitanie).
- **Les travaux de documentation des bases de données réalisés pour sept bases** (BNDMR - APHP ; ESME - Unicancer ; e-SIS - CRCDC Occitanie ; OSCOUR - SpF ; ATU CPBC - Roche ; Hepather - Inserm ; e-must - Agence régionale de santé (ARS) Île-de-France).
- **La préparation des données (standardisation et pseudonymisation) réalisée pour six bases de données** (BNDMR - APHP ; ESME - Unicancer ; e-SIS - CRCDC Occitanie, ATU CPBC - Roche ; Hepather - Inserm ; e-must - ARS Île-de-France).

- **La mise en place d'un canal de transfert** permettant de transférer les données entre le responsable de données et le HDH de manière sécurisée réalisée par cinq responsables de données (BNDMR - APHP ; ESME - Unicancer ; e-SIS - CRCDC Occitanie ; e-must - ARS Île-de-France ; OSCOUR - SpF).

Par ailleurs, après avoir lancé en 2022 les premiers ateliers pour aider à la mise en conformité des bases de données souhaitant intégrer le catalogue, le HDH a poursuivi son accompagnement en 2023. Ainsi, il a pu conseiller **quatorze responsables de base de données** en 2023 souhaitant soit créer leur base de données, soit la mettre en conformité.

Catalogue : mise en valeur des bases de données à travers des projets pilotes

Des projets pilotes ont été lancés sur les bases de données déjà identifiées afin de réduire significativement les délais d'accès aux bases de données du catalogue pour les futurs porteurs de projets (chercheurs, start-up, industriels, etc.). Ainsi, les données seront prêtes à être utilisées et déjà appariées aux données de la base principale du SNDS. En 2023, **cinq bases de données sur les dix** figurant dans l'arrêté catalogue du 12 mai 2022 ont pu être mises en valeur à travers la réalisation de **sept projets pilotes** :

- **La BNDMR**, gérée par l'AP-HP, est un EDS national rassemblant les données de plus d'un million de patients atteints de 4 800 maladies rares et suivis dans les centres experts français. Il vise à décrire les parcours de soins des patients atteints de maladies rares (**projet DROMOS**) ; valider un algorithme de ciblage de personnes atteintes de maladies auto-immunes et auto-inflammatoires systémiques (**projet EMIR ALGO**) ; améliorer le parcours de soins des enfants ayant une variation du développement génital (**projet VarGen**).

10 responsables de données de la première version du catalogue du SNDS accompagnés dans sa mise en œuvre opérationnelle

- **La base de données de dépistage du cancer du sein (e-SIS)** du CRCDC en Occitanie rassemble les données cliniques et mammographies de 150 000 femmes ayant participé au dépistage organisé du cancer du sein du Gard et de la Lozère entre 2004 et 2019. Le projet vise à évaluer l'apport de l'IA dans le dépistage organisé du cancer du sein (**projet DEEP.PISTE**).
- **La base de données du registre relatif aux infarctus du myocarde (registre e-MUST)**, gérée par l'ARS de l'Île-de-France, est un registre issu des prises en charge des patients par les SAMU et SMUR d'Île-de-France pour un infarctus du myocarde depuis 2014. Il vise à étudier les parcours de soin après un infarctus du myocarde (**projet parcours IDM en Île-de-France**).
- **La base relative à l'Autorisation Temporaire d'Utilisation de cohorte (ATUc)** pour le traitement des cancers bronchiques à petites cellules (ATUc CBPC) du laboratoire ROCHE SAS rassemble les données de plus de 1 400 patients ayant bénéficié de l'ATUc en France d'un traitement des cancers bronchiques à petites cellules, autorisation délivrée par l'Agence nationale de sécurité du médicament et des produits de santé (ANSM) en avril 2019. Cette base permet de décrire la survie globale des patients atteints d'un cancer pulmonaire à petites cellules et ayant été inclus dans un programme d'ATU (**projet LINKATEZO**).
- **La base de données relative à la cohorte de patients infectés par le virus de l'hépatite B ou C (HEPATHER)** est portée par l'Agence nationale de recherches sur le sida et les hépatites virales (ANRS) et l'Inserm. Elle rassemble depuis 2014 les données d'une cohorte de plus de 20 000 patients atteints d'hépatite B ou C en France. Elle permet de développer et valider des algorithmes de ciblage de patients avec une infection VHC ou VHB chronique à partir de données médico-administratives de la base principale du SNDS (**projet HEPAVIR ALGO**).

Travaux d'enrichissement du catalogue de données

Afin de contribuer à l'enrichissement des bases de données du catalogue du SNDS, le groupe de travail "Catalogue" du comité stratégique des données de santé a été créé. Il est chargé d'identifier les bases de données existantes qui ont vocation à être inscrites sur l'arrêté portant la liste des bases composant la base principale et le catalogue du SNDS, de recommander leur inscription et, le cas échéant, d'émettre un avis sur l'ordre de priorité de leur inscription.

En 2023, ce groupe de travail du comité stratégique des données de santé, dont le secrétariat est assuré par le HDH, a travaillé sur une **deuxième version de l'arrêté catalogue**. Un total de **vingt-quatre nouvelles bases de données** a été identifié pour intégrer le catalogue de données, venant s'ajouter au dix préalablement visées par le premier arrêté catalogue du 12 mai 2022. La publication du second arrêté ministériel portant la liste des bases composant la base principale et le catalogue pourrait intervenir au cours du second semestre 2024, après avis de la CNIL.

5 bases de données sur les 10 figurant dans l'arrêté catalogue utilisées pour 7 projets pilotes

Catalogue de métadonnées : projets France - Europe

Le catalogue de métadonnées du HDH, dont une [version bêta](#) a été développée, audité et mise en ligne en 2023, sera finalisé et mis à disposition de l'écosystème en open source et en marque blanche en 2024. Cette version finale intégrera notamment **le modèle de métadonnées européen HealthDCAT-AP**, dont une première version a été partagée fin décembre 2023. Dans un premier temps, l'alimentation du catalogue se concentrera sur les bases des partenaires du HDH dans le cadre du catalogue du SNDS, ainsi que de son consortium lauréat du Direct Grant for setting up services by Health Data Access Bodies pour construire l'EHDS.

Ce projet de développement d'un catalogue de métadonnées s'inscrit également dans le cadre du [projet pilote pour la mise en place de l'EHDS](#). En effet, un chantier se consacre à la définition d'une ontologie standard de métadonnées, basée sur le modèle générique de référence "[DCAT-AP](#)" porté par la Commission européenne. Ces travaux doivent permettre, entre autres, d'accélérer l'interopérabilité entre les catalogues de métadonnées européens, toujours dans le but de faciliter la recherche en santé.

Construction de labels de qualité des bases de données

L'appel à projets "**Developing a Data Quality and Utility Label for the European Health Data Space**", porté par la Commission européenne a été lancé afin de garantir la qualité et l'utilité des données partagées dans le cadre de l'EHDS. Il vise entre autres à **développer et implémenter un label de qualité et d'utilité des données de santé** (tel que prévu par la proposition de règlement EHDS), permettant de faciliter l'exploitation des bases de données dans le cadre de la recherche. Portée par l'institut espagnol IACS, la candidature du **consortium européen QUANTUM**, rassemblant vingt-huit partenaires européens dont le HDH et huit partenaires français (l'Inserm, les HCL, la HAS, l'AP-HP, l'AP-HM, les CHU de Bordeaux,

Lille et Toulouse), a été soumise le 13 avril 2023. La candidature du consortium a été retenue et l'accord de financement avec la Commission européenne a, quant à lui, été signé en décembre 2023. Le projet a officiellement débuté le 1^{er} janvier 2024. Le HDH participe aux travaux de réflexion autour de la construction de ce label de qualité et d'utilité des bases de données de santé, notamment sur les critères et les métriques qui le composeront. A terme, il s'agira également de piloter et évaluer l'utilisation du label proposé et de permettre un cadre de réflexion évolutif et dynamique.

Services dédiés aux acteurs souhaitant ouvrir leurs données en open data

Le HDH s'inscrit dans une démarche d'open data et œuvre pour la mise en place de services dédiés aux acteurs désireux de partager leurs données de santé anonymisées en open data. En 2023, le HDH a initié le développement de solutions technologiques et de services d'accompagnement afin de faciliter la mise à disposition des données de santé de manière sécurisée, dans le respect des standards de l'open data. La plateforme gouvernementale **data.gouv.fr** sera d'ailleurs prochainement utilisée pour le référencement des bases de données de santé anonymisées. D'autres travaux se poursuivront en 2024 notamment avec l'ouverture de quatre bases de données.

Soutenir les projets

Du processus d'accompagnement au suivi des projets

Bien que chaque projet soit singulier de par sa finalité, tous passent par de nombreuses étapes communes au cours de leur préparation. Le HDH a ainsi construit en 2022 **une offre d'accompagnement standardisée**, modulable et déployable pour l'ensemble des projets. En 2023, une démarche d'amélioration continue a été mise en place sur ce processus d'accompagnement. Le HDH a procédé à une analyse des délais et difficultés relatifs aux différentes étapes rencontrées par les porteurs de projets. Plusieurs actions ont abouti fin 2023, permettant **des simplifications pour les futurs projets** notamment : la transformation des sessions de formation en MOOC ; la refonte du contrat d'utilisation de la plateforme ; la simplification de l'ajout d'utilisateur ou encore l'amélioration des processus de travail avec la CNAM pour la préparation des données de la base principale du SNDS. Par ailleurs, **un audit sur le processus d'accompagnement** a été conduit en 2023, afin d'assurer sa bonne application par les opérateurs projets du HDH.

De plus, en 2023, le HDH a mis en place **un tableau de bord interne dynamique**, reprenant les différentes caractéristiques des projets (objectifs, partenaires, démarches réglementaires, avancement, etc.) et permettant de réaliser des focus selon différents axes (les projets de la plateforme, l'année de sélection, le domaine médical, etc.). Le HDH en assurant ainsi le passage d'un pilotage "projet par projet" à un **pilotage plus transversal**

du portefeuille, se dote des moyens nécessaires pour répondre à l'enjeu du passage à l'échelle. Enfin, dans un contexte d'accroissement du portefeuille et d'attentes de plus en plus fortes sur la qualité de service et les délais, et parce qu'il est devenu nécessaire d'outiller les projets pour coordonner l'ensemble des acteurs (médecins, juristes, informaticiens, experts de la donnée) et institutions (responsables de traitement, responsables des données, HDH, CNIL, CNAM), plusieurs actions ont été initiées :

- **Un outil de gestion des tickets** permettant au HDH d'assurer le suivi des demandes, en interne et en externe des utilisateurs, existe depuis 2022. En 2023, un chantier autour de la mise en place de tickets standardisés a été initié, et se poursuivra en 2024.
- **Une solution de gestion de la relation client (GRC)** a été paramétrée et testée en 2023. Son déploiement global doit intervenir début 2024 pour un suivi transversal des échanges entre les équipes du HDH et les acteurs de l'écosystème.
- **Un outil de gestion des flux de travail** a été testé en 2023, sur le périmètre des projets et opérations. Il sera déployé plus largement en 2024 et permettra de fluidifier la réalisation des tâches associées au processus projet.

Accompagnement des acteurs dans la relecture des expressions de besoins

Depuis 2021, un **starter kit** est disponible pour tous les acteurs désirant mener un projet nécessitant une autorisation de la CNIL. Composé de documents pédagogiques, il **facilite la constitution du dossier de soumission**. Une expérimentation a été lancée fin 2022 pour accompagner les porteurs de projets, novices de l'usage du SNDS, à compléter les pièces constitutives de leur dossier afin d'améliorer la qualité de leur projet soumis à l'évaluation du CESREES et faciliter les travaux d'extraction de ces données. Concrètement, des experts dans le traitement des données de la base principale ont été mobilisés pour aider les porteurs de projets, n'ayant pas d'expertise sur les données, à ajuster leur protocole scientifique et l'expression de leurs besoins afin de tenir compte des réserves formulées par le comité.

Fin 2023, **neuf porteurs de projets ont ainsi pu bénéficier de cet appui** : pour six d'entre eux, l'avis final du CESREES a été favorable estimant que les documents ajustés répondaient aux observations formulées, deux n'avaient pas encore soumis à nouveau leurs projets, un seul a eu un avis défavorable de ce comité. Cette initiative prendra un nouvel élan en 2024 avec une proposition d'accompagnement des porteurs de projets étendue.

Outils de pédagogie et d'information à destination des acteurs

Des actions d'enrichissement du kit pédagogique à destination des acteurs ont été mises en place :

- Eu égard au nouvel Échantillon du Système National des Données de Santé (ESND), **un kit pédagogique précisant les conditions à respecter et les modalités de soumission** a été mis en ligne au premier trimestre 2023.
- Ont été publiés **deux nouveaux starter kits dédiés aux premières Méthodologies de Référence (MR)** visant à simplifier l'accès aux données de la base principale du SNDS (dénommées [MR-007](#) et [MR-008](#)). Co-construits avec la

CNIL et la CNAM, ils visent à accompagner les porteurs de projets pas à pas dans la mise en œuvre de leur projet.

- La création d'**un kit pédagogique dédié aux projets mis en œuvre dans le cadre d'une procédure simplifiée, dite MR-004**, a été lancée. Elle comprend notamment un modèle de protocole scientifique intégrant des éléments d'aide au remplissage, un guide pédagogique ainsi qu'une checklist permettant aux porteurs de projets de vérifier leur conformité au référentiel. Ce starter kit sera finalisé au premier semestre 2024.

De plus, une large palette de formations a été mise en place par le HDH sur les démarches réglementaires ou sur la base principale du SNDS et ses modalités d'accès. Afin de les compléter, des MOOC (Massive Open Online Course) ont été créés en 2023 :

- Le premier **MOOC "organisation d'un Data Challenge en santé"**, consulté par plus de 220 personnes au cours de l'année 2023, vise à décrire les bonnes pratiques pour que les porteurs de projets potentiels disposent de toutes les clés pour organiser un événement de ce type. Les Data Challenges sont des compétitions en science des données ayant pour but de développer le meilleur algorithme possible pour répondre à une question médicale précise à partir de l'analyse d'un large jeu de données.
- Le **second MOOC**, consulté par plus de 600 personnes en 2023, présente quant à lui **l'architecture de la base principale du SNDS**. Il détaille chacune de ses composantes (DCIR, PMSI, CépIdc), et liste les bonnes pratiques de data management et les outils open source. Décliné en quatre modules, leur contenu se fonde sur les ressources documentaires de la CNAM, l'Agence technique de l'information sur l'hospitalisation (ATIH) et le Centre d'épidémiologie sur les causes médicales de Décès (Cépidc).

Le déploiement de ces formations numériques se poursuivra en 2024.

Guides d'accompagnement sur la sécurité et protection des données

Pour faciliter le travail de mise en conformité des chercheurs, le HDH a fait évoluer ses outils d'accompagnement à destination des utilisateurs dont les projets se déroulent sur la plateforme technologique du HDH. En 2023, suite à plusieurs ateliers avec la CNIL, **un kit documentaire incluant les spécificités de la plateforme technologique du HDH** a été rédigé afin de construire une Analyse d'Impact relative à la Protection des Données (AIPD) conforme aux attentes de la Commission lorsqu'elle instruit les demandes d'autorisation. Il comprend aussi une méthodologie afin de réaliser l'homologation, en particulier l'analyse de risques et la mise en place de mesures de sécurité propres au projet, offrant ainsi au porteur de projet une meilleure maîtrise des risques.

Ces outils font l'objet d'améliorations continues au regard des retours d'expérience que les services de la CNIL présentent régulièrement au HDH. Au cours de la complétion du kit par les responsables de traitement, les équipes du HDH se tiennent à leur disposition pour effectuer plusieurs retours, relectures et ateliers en vue de fluidifier par la suite l'instruction de leur demande d'autorisation auprès de la CNIL.

Poursuite des appels à projets pour soutenir les projets innovants

Chargé de soutenir l'innovation, le HDH assure une veille permanente sur les dernières avancées en science des données et ouvre régulièrement des appels à projets (AAP) destinés à l'écosystème de la santé. Ces multiples appels à projets ou appels à manifestation d'intérêt (AMI), en lien avec les priorités du ministère du Travail, de la Santé et des Solidarités, visent à **appréhender la diversité des données et des usages** (IA, chaînage, ciblage, données de vie réelle, etc) ainsi que tous **les enjeux de santé publique** (cancérologie, essais cliniques, prévention, etc.).

Onze appels à projets en cours ou venant de se conclure :

- **AMI BOAS (vagues 4 et 5)** : cet AMI vise à constituer une Bibliothèque Ouverte d'Algorithmes en Santé (BOAS) et vise à soutenir le développement et la validation d'algorithmes de ciblage dans le SNDS. En 2023, deux vagues ont été lancées.
- **AAP France-Québec – IA en santé (vague 2)** : cet AAP transatlantique franco-québécois, en partenariat avec le Fonds de recherche Québec-santé, vise à renforcer la valorisation des données de santé au service de la recherche. Deux nouveaux projets lauréats ont été sélectionnés lors de la deuxième vague.
- **AAP Accompagnement et soutien à la constitution d'entrepôts de données de santé hospitaliers (vagues 1 et 2)** : en partenariat avec la Direction générale de l'offre de soins (DGOS) et Bpifrance, ce dispositif de la stratégie d'accélération "santé numérique" vise à constituer un réseau d'EDS hospitalier. Six projets lauréats sont issus de la vague 1, dix projets lauréats sont issus de la vague 2.
- **Data Challenges en santé (vague 3)** : les Data Challenges en santé sont des compétitions en science des données qui visent à répondre à des problématiques médicales précises à l'aide de l'analyse de jeux de données anonymisées de haute qualité, mis à la disposition de l'ensemble des participants. Les compétiteurs doivent répondre à la question posée en développant des algorithmes en un temps imparti. Cet AAP a été lancé en partenariat avec la DNS et Bpifrance.
- **AAP CSF-ITS** : le comité stratégique de filière des industries et technologies de santé (CSF-ITS) a lancé l'AAP "data, IA et parcours de santé". L'objectif est d'analyser des parcours de santé pour en cibler les ruptures et proposer des innovations. Dans ce cadre, le HDH s'est associé au collectif organisateur afin d'apporter un soutien adapté aux besoins des lauréats. A l'issue de cet AAP, cinq lauréats ont été sélectionnés.

- **AAP sur le croisement de données de santé et environnementales (vagues 1 et 2) :** le Green Data For Health (GD4H) et le HDH ont conjointement organisé deux AAP successifs afin d'améliorer la mobilisation des bases de données environnementales et de santé au service des projets de recherche et d'innovation. Quatre nouveaux projets lauréats ont été sélectionnés dans le cadre de la vague 1 ; quatre projets lauréats sont attendus dans le cadre de la vague 2, qui sera clôturée au premier semestre 2024.
- **AMI Unicancer (vague 2) :** en partenariat avec Unicancer et la Ligue contre le cancer, le HDH a lancé un AMI visant à accélérer la recherche à partir de données de vie réelle en cancérologie en construisant des bases de données de référence en la matière. Cinq projets lauréats sont accompagnés dans ce cadre.
- **AAP DATAE :** lancé par la DGOS et dont le HDH est l'opérateur, DATAE est un AAP ouvert aux cas d'usage dans tous les domaines de la recherche appliquée en santé. Les projets doivent nécessairement utiliser des données issues d'un entrepôt de données hospitalier.

11 appels à projets en cours ou venant de se conclure

Zoom sur l'appel à projets conjoint en santé-environnement avec le Green Data for Health

En 2022, le HDH s'est associé au Green Data for Health (GD4H), dispositif du Commissariat général au développement durable (CGDD) visant à favoriser le recueil de données environnementales.

Ensemble, ils ont lancé un premier AAP conjoint "La donnée pour la recherche et l'innovation en santé-environnement", centré sur l'impact du bruit, des bassins industriels et des

perturbateurs endocriniens sur la santé. Les quatre lauréats ont été annoncés en mai 2023 et bénéficient d'un accompagnement des équipes du GD4H et du HDH.

Les lauréats de la première vague de l'AAP "santé-environnement" :

Le projet MRC-PE (Maladie Rénale Chronique - Profils Environnementaux) porté par l'université de Lille, en

partenariat avec l'université d'Artois, l'IMT-Nord Europe, Junia et l'Inserm, vise à étudier l'association entre les profils environnementaux et l'incidence d'insuffisance rénale terminale en France, tout en s'intéressant à la médiation potentielle de cette relation par le diabète, l'hypertension artérielle ainsi qu'à d'autres déterminants environnementaux.

Le projet BIS (Bassins Industriels et Santé)

porté par Santé publique France (SpF), en partenariat avec l'Institut national de l'environnement industriel et des risques (INERIS), vise à déterminer l'association entre la proximité de grands bassins industriels et l'état de santé de la population riveraine en construisant des indicateurs d'exposition et de pressions environnementales liés aux émissions industrielles, puis en les croisant avec des indicateurs sanitaires.

Le projet SOMNIBRUIT

porté par BRUITPARIF, en partenariat avec l'Observatoire régional de santé Île-de-France (ORS), la Ville de Paris (VDP) et le Centre du Sommeil et de la Vigilance de l'Hôtel Dieu (AP-HP), vise à connaître et quantifier les effets du bruit environnemental sur le sommeil en mettant en relation la consommation de médicaments prescrits pour les troubles du

sommeil à l'échelle de la commune avec les statistiques d'exposition au bruit environnemental au sein des communes concernées, et à explorer les différences observées dans ces relations en fonction des sources de bruit.

Le projet Noises (cardio NeurO vascular diseases effects of noiSe exposure combined with socioeconomic inequalities)

porté par le CHU de Bordeaux, l'Inserm, l'université de Bordeaux et l'Institut Bergonié, en partenariat avec le Registre de Strasbourg, le Laboratoire Image Ville environnement (LIVE), et l'Unité mixte de recherche du CNRS et de l'université de Strasbourg, vise à cartographier et décrire la distribution spatiale des indicateurs socioéconomiques et environnementaux (pollution de l'air et bruit) de maladies cardio-neuro-vasculaires, à l'échelle

des Îlots Regroupés pour l'Information Statistique (IRIS) pour les territoires de Bordeaux et de Strasbourg afin d'identifier les parties du territoire nécessitant une action prioritaire.

Cette coopération est renouvelée par **le lancement d'un second AAP**, s'inscrivant également dans le cadre du 4e Plan National Santé-Environnement (PNSE4) et de France Nation Verte. Ce dernier vise à **faciliter les croisements de données au service de la prévention**. En enrichissant la base principale du SNDS avec des données de divers domaines de la vie, et en les croisant avec des données de contexte, l'objectif de cet AAP est d'étoffer les connaissances et de mieux caractériser les facteurs de risques. En 2024, quatre projets lauréats seront annoncés lors de cette seconde vague.

Mutualiser les connaissances et les expertises

Participation aux travaux de l'AMDAC : offre de service pour une ouverture, une description et des référencements des données

L'Administration Ministérielle des Données, Algorithmes et Codes sources du ministère du Travail, de la Santé et des Solidarités (AMDAC) a rédigé sa **nouvelle feuille de route 2024-2026**, fruit d'un travail collectif mené en 2023 avec l'ensemble des référents AMDAC, dont le HDH fait partie. Son ambition est de dessiner un futur prometteur et sécurisé pour l'utilisation des données, algorithmes et codes sources au bénéfice d'une action collective. Que ce soit par la démocratisation de l'accès aux données, la simplification des procédures administratives ou la mise à disposition de jeux de données pour le grand public, **l'objectif est de se tenir au plus près des attentes citoyennes**. Cette feuille de route vise également à **encourager la recherche et l'innovation**, en offrant aux chercheurs et aux entreprises les ressources nécessaires pour réutiliser des données de santé dans le respect du RGPD. En proposant des initiatives pour optimiser le travail des acteurs, elle facilite la création de nouveaux services et permet de réorienter les équipes vers des tâches à plus forte valeur ajoutée.

Le HDH, aux côtés des autres contributeurs de la démarche, s'est notamment engagé dans les actions suivantes :

- Lancer le "groupe de travail open source" afin de partager les bonnes pratiques sur la diffusion de codes sources en open source.
- Partager les bonnes pratiques sur les plateformes de données.
- Produire des documents de référence pour la constitution d'EDS.
- Mettre à disposition des règles de gestion sur la pseudonymisation.
- Mettre à disposition des procédures partagées d'anonymisation des données.
- Créer un espace commun de données de "Solidarités".
- Cataloguer les données et algorithmes de la sphère pour mieux les maîtriser.
- Proposer des contenus de formation.
- Rendre plus accessibles les jeux de données phares en open data de la sphère "Solidarités".

Poursuite de la démarche open source du HDH

Le HDH s'inscrit dans une démarche open source et propose, en accès libre, des informations, de la documentation et des formations afin d'utiliser les données plus facilement. En particulier, le HDH propose :

- **Une documentation ouverte collaborative**, actualisée régulièrement en partenariat avec la DREES, SpF, la CNAM, l'ANSM, les ARS et l'ATIH. En 2023, cette documentation comporte 61 fiches thématiques.
- **Un dictionnaire des variables**, développé initialement par la DREES et maintenu par le HDH, permettant de naviguer au sein des tables et variables de la base principale du SNDS. Les mises à jour de ce dictionnaire interactif s'effectuent progressivement, pour intégrer les nouvelles terminologies et variables de la base principale du SNDS. Le site enregistre 22 200 visites en moyenne par an.

- **Un forum d'entraide** à destination des nouveaux utilisateurs de données de santé mais aussi des experts souhaitant échanger avec d'autres membres. Ce forum est ouvert à tous, 281 posts et 51 nouveaux comptes ont été créés au total en 2023.
- **Une communauté Meetup** rassemblant 2 419 membres en 2023 avec plus de 220 inscrits depuis début 2023.
- **Un guide relatif au standard OMOP** permettant d'accompagner les acteurs dans la transformation de leurs données selon le standard international OMOP-CDM.
- **La construction d'un programme visant à mettre en place une Bibliothèque Ouverte d'Algorithmes en Santé (BOAS)** avec l'appui du ReDSiam, la DREES, l'Institut Pierre-Louis Epidémiologie et Santé Publique (IPLESP), Bordeaux PharmacoEpi, PELyon, le CépiDC, Reperes, l'ATIH, SpF et la CNAM a continué en 2023. Deux nouvelles vagues de l'AMI BOAS ont été lancées en 2023. Ainsi, cinq vagues successives d'AAP ont été lancées dans le cadre de ce programme, et ont permis de sélectionner quinze projets. Au premier trimestre 2024, une première version de la BOAS sera finalisée.
- **Plusieurs MOOC** disponibles gratuitement en ligne visent à décrire l'architecture de la base principale du SNDS et de ses composantes. Quatre modules ont été mis en place dont les liens sont à retrouver [ici](#).

Poursuite du programme Data Challenges

Trois nouvelles compétitions Data Challenges ont eu lieu en 2023 :

- Le **Data Challenge VisioMel** a été organisé par la Société Française de Pathologie (SFP), la Société Française de Dermatologie (SFD), le Conseil National Professionnel des Pathologistes (CNPath) et le Groupe de Cancérologie Cutanée (GCC). L'objectif était de prédire la probabilité de récurrence de mélanomes dans les cinq ans suivant le diagnostic initial à partir de l'analyse de lames de biopsies numérisées de près de

2 000 patients. [La compétition](#) s'est déroulée de mars à mai 2023 et a rassemblé près de 600 participants du monde entier qui ont soumis plus de 600 algorithmes. Une cérémonie de remise des prix a été organisée en mai 2023, un prix de 25 000 € a été partagé entre les trois lauréats ayant soumis les algorithmes les plus performants. À la suite de cette compétition et dans une démarche d'ouverture de la science, VisioMel organise l'accessibilité la plus large des données et des [algorithmes des lauréats](#).

- Le **Data Challenge D-IA-GNO-DENT**, porté par les Hôpitaux Universitaires de Strasbourg (HUS), visait à développer une solution d'IA d'aide au diagnostic de maladies rares à manifestations bucco-dentaires à partir de l'analyse de photographies de la bouche et des dents de 200 patients. [La compétition](#) a eu lieu d'avril à juillet 2023, et a rassemblé une centaine d'équipes qui ont soumis plus de 1 450 modèles. Une remise des prix a été organisée en décembre 2023 afin de récompenser les gagnants et de présenter les premiers résultats et la suite des travaux. Les algorithmes les plus performants ont été mis en [open source](#).



compétitions
Data Challenges
ont eu lieu
en 2023

- L'**Allergen Chip Challenge**, mené par la Société française d'allergologie (SFA) portait sur l'amélioration du diagnostic des allergies en reliant l'historique clinique au profil immunologique de 3 000 patients. [La compétition](#) a duré trois mois et a rassemblé près de 300 participants qui ont soumis plus de 3 000 algorithmes. Une remise des prix a été organisée en novembre 2023, afin de récompenser les lauréats. A ce jour, les scripts des lauréats sont disponibles en [open source](#) et la base de données support du Data Challenge est vouée à être partagée en open data afin de permettre à la recherche de poursuivre sa progression.

Le HDH en partenariat avec la DNS, le Secrétariat général pour l'investissement (SGPI) et Bpifrance dans le cadre du plan France 2030 lance régulièrement des AAP visant à sélectionner de nouveaux projets Data Challenges à accompagner. **Trois nouveaux projets ont été sélectionnés en 2023 et sont en cours d'organisation :**

- Le **Data Challenge Cytolog-IA**, porté par le Groupe Francophone d'Hématologie Cellulaire (GFHC) et la société Algoscope, a pour objectif de créer un outil d'IA d'aide à la classification automatique de leucocytes dans le but d'automatiser la lecture de frottis sanguin et d'aider les praticiens dans la démarche diagnostique.
- Le **Data Challenge Cardi-Hack**, de l'Institut de Cardiométabolisme et de Nutrition de l'AP-HP (IHU ICAN), souhaite développer une solution d'IA capable de prédire le pronostic des patients de moins de 40 ans atteints de cardiomyopathies hypertrophiques.
- Le **Data Challenge ANNITIA**, porté par une autre équipe de l'IHU ICAN, a pour but de développer des algorithmes capables de prédire l'évolution de la stéatose hépatique non alcoolique et la survenue d'événements hépatiques graves à partir de données issues de tests non invasifs.



En parallèle, le HDH a poursuivi en 2023 l'accompagnement des challenges sélectionnés précédemment, et dont les compétitions auront lieu en 2024 et 2025 :

- Le **Data Challenge DigiLUT**, porté par l'Hôpital Foch, a pour objectif de développer un algorithme capable de détecter un rejet de greffe à la suite d'une transplantation pulmonaire à partir de l'analyse de lames virtuelles de biopsie de greffon.
- Le **Data Challenge Dat-Hub** de la Société Française de Médecine Nucléaire (SFMN) pour développer un algorithme contribuant au diagnostic des syndromes parkinsoniens à partir de données d'imagerie cérébrale fonctionnelle Datscan du système dopaminergique.

Animation de l'écosystème français sur l'utilisation du modèle de données OMOP-CDM

Le HDH a lancé, depuis 2020, des travaux pour standardiser la base principale du SNDS. En effet, **la standardisation des données de santé** pour permettre leur interopérabilité représente un enjeu majeur pour faciliter et accélérer la recherche en santé. Par exemple, le standard OMOP-CDM permet, entre autres, d'adapter la structure des données pour **faciliter la réalisation de projets de recherche épidémiologique**.

Il bénéficie d'une reconnaissance internationale, portée par l'association Observational Health Data Sciences and Informatics (OHDSI) dont les efforts ont permis de standardiser les dossiers médicaux de plus de 800 millions de patients.

Les efforts du HDH en ce domaine lui ont permis de contribuer à la montée en compétence globale de l'écosystème français à ce sujet. **De nombreuses actions ont été réalisées en 2023** à ce sujet, telles que :

- **L'ouverture des scripts de standardisation de l'intégralité de la base principale du SNDS**, la création de l'entrepôt multicentrique EMC2 basé sur le modèle OMOP.
- Le lancement d'un groupe de travail visant à **standardiser le socle commun des données de santé** des principaux entrepôts de données de santé français.
- Le lancement d'un groupe de travail sur **la standardisation du socle commun de données de santé des EDS** français.
- La **présentation des travaux** sur ce standard lors du congrès EMOIS.

Identification de besoins communs (outils, modèles), et actions en articulation avec la Stratégie Nationale pour l'Intelligence Artificielle

En avril 2022, le HDH a été mandaté par le coordinateur de la Stratégie Nationale pour l'Intelligence Artificielle (SNIA) pour animer un groupe de travail visant à proposer **une feuille de route afin d'accélérer le développement d'outils du Traitement Automatique du Langage (TAL)**⁵ dans le domaine de la santé, et ce à l'échelle nationale. Le HDH a ainsi réuni et piloté un groupe de travail de dix chercheurs spécialisés dans le TAL sur les données de santé afin de proposer en janvier 2023 au coordinateur de la SNIA une note de synthèse présentant des actions stratégiques déployables dès 2023.

Suite au lancement grand public de ChatGPT, la feuille de route stratégique de la SNIA a évolué ce qui l'a conduit à solliciter à nouveau le HDH en 2023, cette fois afin de proposer **des cas d'usage stratégiques de l'IA générative en santé**. Le HDH et le groupe de travail "TAL en santé" se sont mobilisés très rapidement afin de rendre une note de synthèse identifiant notamment **trois grandes actions à mener** : l'amélioration d'un grand modèle de langue sur des données médicales, la création d'un corpus de comptes rendus médicaux fictifs en accès libre ainsi que la génération de données synthétiques via l'IA générative.

La SNIA a annoncé, en juin 2023, **le lancement d'une procédure d'AAP** afin d'accélérer **la création et la mise en accessibilité de communs numériques en IA générative**. Le HDH a réuni un consortium de haut niveau afin de proposer un projet candidat avec l'ambition d'accélérer fortement l'appropriation et le développement de solutions TAL en santé en livrant un ensemble d'outils open source de pointe. Il réunit le groupe de travail d'experts TAL en santé et dix-huit établissements de santé répartis sur tout le territoire – dont notamment l'AP-HP, le Centre Léon Bérard, le Groupement de coopération sanitaire (GCS) HOURAA – et un acteur deeptech qui collabore actuellement avec la Direction Interministérielle du Numérique sur un projet mobilisant des Large Language Models (LLM). Le **projet "PARTAGES"** (développement Avancé de communs numériques pour l'Intelligence Artificielle Générative En Santé) a été présélectionné par Bpifrance en décembre 2023. La procédure de sélection se poursuit au premier trimestre 2024.

⁵ Le traitement automatique du langage (TAL) est un des champs de recherche les plus actifs de l'IA. Avec le TAL, il est possible d'analyser des données textuelles de manière automatique afin d'assister le travail des médecins, que ce soit dans le cadre de l'aide au diagnostic ou pour réaliser des tâches récurrentes chronophages.

Renforcer l'attractivité du Health Data Hub

Mise en visibilité de l'offre de service du HDH à des publics ciblés

Depuis 2019, **plus de 100 projets** ont rejoint le portefeuille de projets du HDH, qui a connu une accélération importante sur la période 2022-2023 avec **plus de vingt-cinq nouveaux projets par an**. L'ambition est à présent de passer le cap des 200 projets accompagnés. En 2023, le HDH a initié des présentations de son offre de service, ce qui lui a permis en fin d'année de structurer cette démarche de manière plus industrielle et proactive, afin d'aller à la rencontre des acteurs. Une démarche qui permet de stimuler le montage de nouveaux projets, ou d'aider à accélérer ceux en cours.

Un quart des projets que le HDH accompagne implique une start-up, et une offre de service dédiée lui est consacrée, avec le soutien de ses partenaires. Pour contribuer à ce que davantage de start-up accèdent à la base principale du SNDS, dans le respect de la réglementation en vigueur, et les aider à l'utiliser, des actions ont été engagées. Ainsi, s'est tenue la première édition de **"Start-up connect : rencontre avec l'écosystème"**, organisée par Parisanté Campus et le HDH en juin 2023. Cet évènement a permis de réunir plus de 200 acteurs parmi lesquels des investisseurs, des structures publiques et des experts en données de santé, des établissements de soins, ainsi que plus de 115 start-up.

Plus de
200
participants à
l'évènement "Start-up
connect : rencontre
avec l'écosystème"

Faire connaître l'écosystème des données de santé et les actions du HDH

L'expérience française en matière de réutilisation des données de santé mise en avant lors d'un séminaire à Bruxelles

Dans le contexte des négociations sur le règlement pour un espace européen des données de santé, le HDH et la DNS ont organisé le séminaire **"Towards EHDS: The French experience in secondary use of health data"**, le 25 avril 2023, à la Représentation permanente de la France auprès de l'UE, à Bruxelles. La quarantaine de représentants des pays européens ont ainsi pu découvrir la stratégie française en matière de réutilisation des données de santé, ainsi que l'offre de services du HDH grâce aux interventions respectives de Raphaël Beaufret, co-responsable du numérique en santé à la DNS, et d'Emmanuel Bacry, directeur scientifique du HDH. Plusieurs porteurs de projets sont par ailleurs intervenus pour **illustrer**

la stratégie française en matière de réutilisation des données : Arnaud Rosier, CEO d'Implicit, porteur du projet HYDRO (qui vise à mieux prédire/prévenir les insuffisances cardiaques), Jean-Yves Blay, président d'Unicancer, et directeur général du Centre Léon Bérard, porteur du projet DEEP.SARC (qui vise à améliorer le parcours de soins des patients atteints de sarcomes), Christel Daniel, du département Innovation et données de l'AP-HP, et le professeur Cyrille Confavreux, des HCL, pour évoquer la création d'EDS hospitaliers.

Retour sur la dernière édition de la Summer School AI4Health

Chaque année, AI4Health offre à des étudiants, chercheurs ou professionnels **la possibilité de se former aux techniques de l'IA en santé**, en mobilisant des intervenants de renommée internationale qui assurent pendant une semaine des cours pléniers et des sessions pratiques. Alors que l'IA en santé, notamment générative, est un sujet d'actualité majeur, une édition d'été a été organisée sur une semaine en juillet 2023. Le HDH a ainsi co-organisé avec trois instituts interdisciplinaires en IA (Prairie, 3IA Côte d'Azur et MIAI Grenoble) et avec le soutien de quatorze partenaires industriels cet événement. La séance inaugurale s'est déroulée avec les interventions de : Alessandro Blasimme (maître de conférences en bioéthique à l'Ecole Polytechnique Fédérale de Suisse - ETH Zürich), Andrzej Rys (conseiller scientifique principal, Direction générale de la santé et de la sécurité alimentaire (DG SANTE), Commission européenne) et Mihaela Van der Schaar (professeure

d'apprentissage automatique, d'IA et de médecine à l'université de Cambridge et membre de l'Alan Turing Institute à Londres).

Étudier les opportunités des données de santé grâce au conseil scientifique consultatif (CSC)

En janvier 2021, un CSC a été créé avec pour vocation d'accompagner la direction du HDH dans ses missions, notamment par le biais d'avis et de propositions visant à l'éclairer dans ses décisions. Y siègent vingt-huit personnalités issues de nombreuses disciplines telles que la médecine, la pharmacie, la génétique, l'épidémiologie, la statistique, l'économie de la santé, le juridique, les sciences des données et intelligence artificielle ou éthique. Dès la fin de l'année 2021, le conseil s'est élargi pour accueillir sept membres internationaux.

Durant l'année 2023, en poursuivant les activités initiées l'année précédente, les membres du CSC ont mis à disposition leur expertise scientifique pour soutenir les équipes du HDH dans le déroulement de leurs activités. À titre d'exemple, nous pouvons citer l'implication de Pierre Zweigenbaum, directeur de recherche au CNRS et à l'Université Paris Saclay, spécialisé en IA, qui a grandement contribué à la construction d'**une réponse ambitieuse à un AAP de Bpifrance portant sur les communs numériques en IA générative**. Cela a conduit à la création d'un consortium candidat d'une trentaine d'acteurs majeurs de la santé en France, parmi lesquels des CHU, organismes de recherche, etc.

En ce qui concerne l'appui aux travaux européens, Ségolène Aymé et Alain Livartowski ont participé en tant qu'experts au projet de réponse à un appel à **financement de la Commission européenne, visant à aider les pays de l'Union européenne à préparer le futur règlement de l'EHDS**. Grâce à leur contribution, le HDH a pu former **un consortium fort de dix-neuf partenaires nationaux stratégiques** qui collaborent aux côtés du HDH pour rendre de nombreuses données de santé disponibles à des fins d'utilisation secondaire.

Plus de
300
participants issus de
27 pays à la Summer
School AI4Health



Ségolène Aymé, présidente du CSC, a participé au jury de sélection de la vague 2 de l'AAP visant à développer les entrepôts de données de santé hospitaliers, porté par la DGOS et Bpifrance dans le cadre de France 2030, ainsi qu'au jury de sélection des lauréats de la seconde vague de l'AAP commun avec le Fonds de Recherche Québec-santé (FRQS), aux côtés de Pierre Zweigenbaum. Bertrand Lukacs, vice-président du CSC, a quant à lui participé au jury de sélection des lauréats de l'AAP données de santé-environnement porté avec le Green Data for Health (GD4H), tandis que Bernard Hamelin a pour sa part, participé au jury de la seconde vague de l'AAP Unibase, porté conjointement avec Unicancer et la Ligue contre le cancer. De plus, Bertrand Lukacs participe aux travaux de l' AIS dont un des objectifs est de renforcer le suivi de l'efficacité des produits de santé en vie réelle afin de faciliter l'évaluation et une prise en charge de ces derniers.

En 2023, le CSC a également lancé et/ou poursuivi ses propres travaux sur des sujets de prospective, identifiés en collaboration avec la direction du HDH. Parmi ces sujets figurent les enjeux éthiques liés à l'utilisation des données de santé, la structuration des registres de données sur les maladies rares, ainsi que les possibilités offertes par les données génomiques et en santé mentale.

La thématique de l'éthique en santé numérique a fait l'objet d'un groupe de travail dirigé par Ségolène Aymé dans le but de répondre aux attentes et aux préoccupations suscitées par la disponibilité des données de santé à des fins de

recherche. L'objectif est de **permettre un débat éclairé sur la recherche en données de santé.**

Les travaux ont donné lieu à la publication d'un article scientifique (disponible sur [HAL](#)) et ont permis d'apporter des éléments dans le cadre d'une audition menée par un groupe de travail du comité éthique de l'Inserm à ce sujet. Une des principales conclusions de cette réflexion est que **les avantages (mesure de l'efficacité des soins, surveillance épidémiologique, amélioration des connaissances en santé, etc.) l'emportent sur les risques liés à l'utilisation secondaire des données de santé**, notamment en raison du niveau de sécurité appliqué en France, où les données sont pseudonymisées ou anonymisées. Face à ce constat, les rédacteurs suggèrent que **l'évaluation du rapport bénéfices/risques serve de base aux décisions d'autorisation pour l'utilisation des données de santé à des fins de recherche**, pour en maximiser les bénéfices.

Les **registres de maladies rares** sont des bases de données spécialisées, et de ce fait une source de données indispensable à la recherche en santé, qui soulèvent plusieurs enjeux, notamment liés à leur nombre et à la multiplicité des maladies rares ainsi qu'à l'absence de codes spécifiques dans la CIM10. Ce sujet a fait l'objet d'un groupe de travail mené par Martine Gilard, membre du CSC, spécialiste en cardiologie.

Afin d'identifier les pistes à des actions de soutien aux registres, le groupe de travail du CSC a mené des auditions avec les représentants de plusieurs registres d'importance. Les registres

apparaissent hétérogènes et limités dans leurs missions. Aucun registre ne bénéficie de financement à la hauteur de ses besoins. Il existe cependant en France une base nationale, la BNDMR, qui est exhaustive car basée sur les consultations dans les centres de référence nationaux où tous les patients consultent une fois par an. **La BNDMR offre la possibilité aux établissements hospitaliers de collecter au-delà du set de données minimal** obligatoire connecté au dossier patient informatisé. De plus, **la BNDMR a entrepris un chaînage de toutes les maladies rares avec la base principale du SNDS**. Le HDH, qui a déjà inscrit la BNDMR à son catalogue, pourrait apporter son soutien aux projets de la BNDMR afin de valoriser les données de ces maladies très invalidantes et coûteuses, et faire progresser les connaissances et les soins.

Afin de **mieux comprendre les nouveaux enjeux liés aux données génomiques ainsi qu'aux données de psychiatrie et santé mentale**, le conseil s'est appuyé sur des experts afin de mener des états des lieux de ces enjeux. À ce jour, **l'état des lieux sur les données de psychiatrie et en santé mentale est en cours de finalisation**, tandis qu'un **support de formation sur les données génomiques** est prévu à destination de l'écosystème. Les travaux réalisés en 2023 mettent en évidence les bases de données existantes et les perspectives qu'elles représentent pour la recherche. Ils permettent d'identifier des acteurs avec lesquels le HDH pourrait être susceptible de nouer des partenariats d'intérêt dans le futur.

L'intérêt pour **les données d'anatomie et cytopathologie pathologiques, les données sociales et d'oncologie** se multipliant, le CSC a suivi de près l'évolution des travaux en la matière, qui ont connu plusieurs évolutions durant l'année 2023. En effet, une mission sur la pathologie digitale a été lancée en juin par le ministère du Travail, de la Santé et des Solidarités, tandis qu'un décret portant la création d'une mission interministérielle chargée du pilotage des données sociales a été publié en décembre. Enfin, le **Paris Saclay cancer cluster**

(PSCC) poursuit son développement. Le CSC et les équipes du HDH suivent avec intérêt ces initiatives, le HDH participant notamment à la Task Force "Anatomopathologie numérique" de France biotech, qui a pour objectif de stimuler la collaboration au sein de l'écosystème sur ce sujet et d'identifier les solutions innovantes en la matière, dans l'optique de rédiger un livre blanc sur l'ACP en 2024.

Le CSC s'est réuni à deux reprises au cours de l'année 2023. La première réunion a eu lieu le 7 juin 2023, au cours de laquelle les équipes du HDH ont pu échanger sur les premiers résultats des projets accompagnés par le HDH. La seconde réunion, tenue à Paris le 22 novembre 2023, a marqué la clôture de l'année du CSC et a permis d'avoir un regard rétrospectif sur les réalisations du HDH quatre ans après sa création. Ainsi, les équipes du HDH ont eu l'occasion de présenter au conseil le bilan de l'année 2023 et le programme de travail pour 2024, tandis que les travaux des différents groupes de travail ont ensuite été présentés.

Travaux autour des moyens à apporter à l'écosystème en termes de financement

En 2023, le HDH a initié une réflexion sur la possibilité de bénéficier de financements privés pour apporter plus de moyens à l'écosystème de la recherche et à ses porteurs de projets. Des structures privées et publiques, proches de la santé ont été identifiées car ayant bénéficié de financements privés, ou d'un mélange de financements privés et publics, pour mener à bien des projets d'intérêt général. Plusieurs dizaines d'acteurs ont été analysés sur la base d'une grille décrivant leurs activités, leur structure juridique et leur mode de financement.

ENJEU

2

Mettre à disposition les données de la base principale du SNDS, l'enrichir et faciliter sa réutilisation



Développer l'expertise du Health Data Hub dans la mise à disposition des données

En 2023, **plus de 200 extractions de données**⁶, une quinzaine d'appariements directs, quatre appariements indirects et cinq ciblage ont été réalisés par le HDH en appui à la CNAM. Aussi, pour assurer ces travaux ou se former aux données, neuf nouveaux collaborateurs ont suivi en 2023 les formations sur la base principale du SNDS proposées par la CNAM afin de pouvoir accéder au portail du SNDS de la CNAM.

Le HDH a aussi mis en place des leviers permettant de gagner en efficacité sur ces différentes activités : développement d'un outil (Python2sas) servant à paramétrer automatiquement et rapidement l'ensemble des programmes d'extraction ; implémentation du programme de génération automatique des bilans d'appariement direct destinés aux porteurs de projets ; conduite d'ateliers réguliers pour mieux impliquer les porteurs de projets dans la stratégie d'appariement indirect.

Aussi, dans l'objectif de fluidifier le processus de mise à disposition des données, et de réduire les délais d'accès aux données, le HDH et la CNAM ont signé un **contrat clarifiant les rôles et responsabilités attribués à chacun** par le décret SNDS. A l'instar de la CNAM, le HDH est ainsi désormais en mesure de signer avec les porteurs, dont les projets se déroulent sur la plateforme technologique du HDH, les conventions encadrant la mise à disposition des données.

6 L'identification de la population d'étude via un ciblage ou un appariement, puis l'extraction des données pour le projet constituent des étapes clés de la préparation des données de la base principale du SNDS.

Faciliter l'exploitation de la base principale du SNDS

Avancées du groupe de réflexion autour de l'accélération des extractions

En 2023, la CNAM, aux côtés du HDH, de la DREES et de la DNS ont engagé une réflexion dans le cadre d'un groupe de travail, visant à identifier de nouvelles modalités de mise à disposition des données susceptibles de **réduire les délais d'accès**. L'une d'elles pourrait consister à déléguer une partie des travaux d'appariement et de ciblage, portés aujourd'hui seulement par les services de la CNAM, à des porteurs de projets experts dans le traitement des données.

Plus de
600
projets mobilisant la base principale du SNDS ont été soumis au HDH depuis sa création

Cela reviendrait à leur donner temporairement accès à la base principale du SNDS afin qu'ils réalisent les travaux de préparation de leurs données, en conformité avec le périmètre autorisé. Cette piste est toujours à l'étude.

Actions du comité de contributeurs de la base principale du SNDS

En 2023, un comité des contributeurs (dit, "comité de suivi du SNDS") à la base principale du SNDS a été mis en place et piloté par la DREES⁷. Il vise d'une part à **assurer un suivi opérationnel des différents chantiers portés par le comité stratégique, à identifier les éventuelles difficultés rencontrées et à décider ensemble d'un plan d'action pour y répondre**. D'autre part, ce comité a pour objectif de **donner de la visibilité sur l'alimentation de la base principale du SNDS** par de nouveaux flux de données.

Ainsi, certains flux sont déjà identifiés et leur intégration a continué en 2023 : c'est notamment le cas de l'intégration des données relatives au handicap, des causes médicales de décès ou des données liées à la vaccination contre la covid-19. D'autres flux sont en cours de création au niveau de la collecte des données pour l'usage primaire, et leur intégration pour l'usage secondaire ne pourra se faire que dans un second temps.

⁷ Créé par l'arrêté du 29 juin 2022, le comité stratégique des données de santé a comme mission principale de proposer des orientations pour le développement du SNDS au ministère du Travail, de la Santé et des Solidarités.

Poursuite du développement de la documentation et des outils ouverts pour faciliter la manipulation de la base principale du SNDS

Le HDH maintient et développe des outils en accès libre à destination de la communauté permettant de mieux appréhender les données de la base principale du SNDS, dont certains ont évolué suite à des travaux mis en oeuvre en 2023 :

- Un [dictionnaire interactif](#) et son schéma de données associé permettant de naviguer au sein des tables et variables. Il comprend désormais les métadonnées de 2021 de la base principale du SNDS, ainsi qu'un ajustement du schéma du SNDS, qui décrit le contenu et les relations entre les fichiers de données.
- Une [documentation collaborative](#) en ligne autour des données de la base principale du SNDS. Elle a été complétée avec les règles de la transformation OMOP (faisant suite à leur validation par le consortium EH DEN en 2023), un guide d'initiation à la base principale du SNDS ainsi que de nouvelles fiches techniques, dont celles sur les requêtes types au sein du [PMSI HAD](#), [SSR](#) et [RIM-P](#).
- Un [générateur de données synthétiques](#) pour une première prise en main des données fictives de la base principale du SNDS avant l'accès effectif à cette base. Il intègre désormais des technologies (Polars, Spark) permettant d'optimiser les temps de traitement pour la génération de grosses volumétries.
- Le [forum](#), pour échanger et poser des questions sur les données de la base principale du SNDS et partager les dernières actualités, ainsi que le site [écosystème SNDS](#) permettant de répertorier et de rechercher des travaux mobilisant les données de la base principale du SNDS sont alimentés au quotidien.

Renforcement du programme Bibliothèque Ouverte d'Algorithmes en Santé (BOAS)

De nombreux partenaires participent à la démarche de science ouverte portée par le HDH, parmi lesquels, la DREES, la CNAM, l'ATIH, SpF, le ReDSiam ou les ARS. Le HDH a lancé en juin 2021 la première vague de l'appel à manifestation d'intérêt (AMI) BOAS. Cet AMI vise à soutenir le développement et la validation d'algorithmes de ciblage facilitant la manipulation de la base principale du SNDS afin d'alimenter une bibliothèque de ressources en open source. **En 2023, ont été lancées la quatrième et la cinquième vagues de l'AMI.** Ce dernier est désormais **ouvert en continu**, et de premiers algorithmes ont déjà pu être développés, validés et présentés par les porteurs au travers des Meetups organisés par le HDH.

Aussi, en 2023 a été initié le développement de la BOAS dont l'objectif de **répertorier les algorithmes de traitement des données de santé**, facilitant ainsi leur réutilisation dans divers projets de recherche et d'innovation. La BOAS est disponible en ligne sur le site du HDH. Elle est alimentée au fil du temps et vise à lister, documenter et rendre accessibles les codes et algorithmes relatifs aux données de santé.

Améliorer la pose de diagnostic dans le cas du cancer des poumons

Le projet Poumalgo, lauréat de la 4e vague de l'AMI BOAS, porté par la Clinique Beausoleil, en association avec les registres de l'Hérault et de l'Isère.

Avec près de 47 000 nouveaux cas diagnostiqués chaque année en France métropolitaine, le cancer du poumon reste l'un des cancers les plus fréquents. Il est d'ailleurs la première cause de mortalité par cancer (20 % des décès par cancer), avec près de 33 000 décès en 2019, selon SpF. En raison de l'absence de symptômes au stade précoce, la majorité des cancers sont souvent découverts à un stade avancé (environ 55 % des cancers du poumon), sans possibilité de recourir à un traitement, expliquant le pronostic et la mortalité élevée.

15

projets lauréats
issus de l'AMI BOAS

La CNAM a développé des algorithmes dont un, dans la base principale du SNDS, identifiant les cas de cancer du poumon (actifs ou sous surveillance). Le **projet POUMALGO** vise à valider et sophistiquer cet algorithme pour améliorer ses performances. Cela permettra lors d'études sur les cancers pulmonaires utilisant la base principale du SNDS de **n'inclure que les vrais cas de cancer du poumon et ainsi mieux répondre aux objectifs**. Les chercheurs pourront, par exemple, mener des études comparant les différentes approches thérapeutiques en terme de survie et de consommation de soins, concernant l'analyse des parcours, identifier les examens réalisés avant le diagnostic et déterminer quelle prise en charge amène à un diagnostic précoce et ainsi à un meilleur pronostic, rechercher les comorbidités associées au cancer etc. Le projet est en cours et l'algorithme développé sera partagé en open source.

Poursuite des travaux OMOP-CDM

Depuis 2020, le HDH standardise les données de la base principale du SNDS, et les transforme vers le modèle de données commun OMOP-CDM. Cette standardisation, qui permet de **passer de plus de 180 tables à moins de 20 clairement documentées en anglais** et adaptées aux études observationnelles longitudinales, ouvre donc son utilisation à des projets européens. Des internes en médecine ont également travaillé sur les alignements de plusieurs milliers de codes et chapitres des terminologies présentes dans la base principale du SNDS (CCAM, NABM, LPP, CSARR, etc.) vers des équivalents internationaux (SNOMED-CT, LOINC, RxNorm, etc.). Les scripts (ETL) permettant de transformer les données au format OMOP-CDM ont été ouverts sur [Gitlab](#), et une documentation détaillant les règles de transformation a été publiée sur la [documentation](#) collaborative du SNDS. Initialement réalisés sur les données de la base principale du SNDS Fast Track (patients avec un diagnostic hospitalier de Covid-19), ces travaux ont été étendus en 2023 à tous les bénéficiaires sur la période 2015-2021 : un travail validé par EHDEN, institution chargée du standard OMOP-CDM en Europe.

Développement d'un outil de data visualisation des parcours de soins

Depuis 2022, le HDH accompagne, dans le cadre d'un partenariat de recherche, **le développement et l'intégration à sa plateforme d'un outil de data visualisation des parcours de soins dans la base principale du SNDS**. Développé par l'équipe Aviz (Inria et CNRS) et le centre de mathématiques appliquées (Polytechnique), **l'outil ParcoursVis** permet de transformer des données de la base principale du SNDS en événements médicaux significatifs et de les visualiser de manière interactive. En 2023, l'équipe Aviz a ajouté de nouvelles fonctionnalités, et **l'outil a été intégré à la plateforme du HDH** pour permettre aux porteurs de projets de bénéficier de données réelles, alors que les chercheurs travaillaient jusque-là sur des données synthétiques. De nouveaux travaux auront lieu en 2024.

Poursuite du développement du concentrateur, pour industrialiser les appariements

Piloté par le HDH, et suivi par la CNAM et la Caisse Nationale d'Assurance Vieillesse (CNAV), un système d'information, dit "concentrateur", est en cours de construction. Il vise à :

- **Proposer une interface à destination des citoyens** pour leur fournir des informations concernant leurs droits et leur permettre de les exercer.
- **Proposer une interface unique aux responsables de données** pour qu'ils puissent initier l'appariement de leurs jeux de données via le transfert de tables d'accrochage contenant des données directement identifiantes (NIR ou informations d'identité).
- **Réduire l'effort requis** tant du côté de la CNAM pour pseudonymiser les NIR, que du côté de la CNAV pour reconstituer les NIR, à partir des traits identifiants en proposant un processus automatisé de bout en bout impliquant leurs SI respectifs.
- **Permettre la centralisation des tables d'accrochage pour les projets multicentriques** qui se multiplient, c'est-à-dire des projets qui nécessitent le croisement entre plusieurs bases de données hospitalières et les données de la CNAM par exemple.

Le HDH a poursuivi ses travaux sur ce concentrateur en 2023 en réalisant le **Document d'Architecture Technique (DAT)**, ainsi qu'une **analyse de risques, une analyse d'impact de la protection des données (AIPD) et l'application SI d'industrialisation des appariements avec la base principale du SNDS**. Sa mise en production est prévue pour 2024.

Actions pour créer une offre de formations à l'utilisation des données de la base principale du SNDS

En 2023, le HDH a animé un groupe de travail chargé d'identifier les concepts clés à maîtriser pour disposer des compétences nécessaires, et de proposer des pistes d'action pour augmenter le vivier d'experts dans le traitement des données de la base principale du SNDS. Ces premiers travaux ont conduit le HDH à **produire un premier template de grille de compétences**. Le HDH a ensuite identifié des acteurs intéressés pour intégrer, dans leur dispositif pédagogique (masters, écoles d'ingénieurs), un **module dédié à la présentation de la base principale du SNDS**.

Contribuer à enrichir la base de données de la base principale du SNDS

Alimentation de la doctrine nationale relative aux EDS : un socle de données commun

En 2023, un groupe de travail dédié a proposé un socle commun de données qui servirait de cible à atteindre pour l'ensemble des EDS hospitaliers : **51 items** répondant aux enjeux stratégiques de développement du réseau d'EDS hospitaliers et adapté aux limites de collectes actuelles pour permettre une mise en œuvre efficiente dans chaque établissement. Ce socle est composé principalement des **données les plus couramment collectées par les établissements** dans le cadre des soins, notamment les données cliniques, médicamenteuses, les données socio-démographiques, ainsi que les résultats biologiques. Ce socle pourrait favoriser la mutualisation des données entre les établissements, leur collecte à grande échelle et leur rapprochement avec d'autres bases de données nationales. En tant que référence généraliste, ce socle soutiendrait également la gestion des établissements de santé, l'évaluation des parcours et des pratiques de soins **en offrant une vision commune** des patients et de leur prise en charge à l'hôpital.

Les travaux autour du projet d'entrepôt de données de médecine de ville

Le projet d'entrepôt national de données de médecine de ville **"P4DP"** (Platform for Data in Primary care), et lauréat 2022 de l'AMI Santé numérique, est porté par l'université Côte d'Azur, l'université de Rouen Normandie, le CHU de Rouen, le Collège national des généralistes enseignants (CNGE), le HDH et Loamics (acteur privé apportant la brique technologique). En effet, il permettra d'**utiliser les données de santé de la patientèle des médecins généralistes** pour améliorer la qualité des soins en médecine générale et la prise en charge des patients, de soutenir la recherche et les essais cliniques et de contribuer à la surveillance épidémiologique. Ces données issues des soins primaires seront chaînées à celles de la CNAM.

Afin de contribuer à constituer cet entrepôt, plusieurs travaux ont été réalisés en 2023, se concluant par la signature d'un contrat de collaboration avec deux éditeurs de logiciels, et **l'identification de plus de 600 médecins participants**. Les briques techniques de l'entrepôt sont en cours de construction et la convention de mise au catalogue du HDH a été signée.

Maintenir la plateforme à l'état de l'art fonctionnellement et en matière de sécurité

Les travaux autour de l'expérience utilisateur de la plateforme technologique du HDH

Dans le but d'établir une communication transparente entre les développeurs et les utilisateurs finaux, favorisant ainsi un cycle continu d'amélioration basé sur des retours d'expérience concrets, le HDH vise à **inclure les porteurs de projets dans l'évolution de nouvelles fonctionnalités**. Cette collaboration étroite permet de garantir que les mises à jour et les améliorations apportées à la plateforme sont en phase avec **les besoins réels des utilisateurs**, renforçant ainsi les possibilités d'analyses pour l'écosystème de données de santé. Aussi le HDH propose des sessions régulières de collecte entre opérateurs projets et porteurs de projets. Il a également mis en place en 2023 **un comité des utilisateurs**, afin de partager l'avancement des travaux et collecter les nouveaux besoins.

L'enrichissement de la plateforme technologique

Le HDH propose à tous les porteurs de projets une offre logicielle qui évolue en permanence, dans un environnement sécurisé et à l'état de l'art. En 2023, le HDH a développé de nouvelles fonctionnalités sur la plateforme du HDH avec pour objectif de **faciliter les projets d'analyse des porteurs de projets**. En effet, il est dorénavant possible d'importer facilement des paquets (conteneur de fonctions) Python et R, ainsi que des modèles pré-entraînés au sein de l'environnement sécurisé de la plateforme. L'accès à **l'outil d'analyse**

Superset a également été facilité et un accès aux logs a été pensé pour que les utilisateurs puissent résoudre plus facilement des problèmes de code. De plus, la plateforme technologique s'est enrichie pour accélérer les ingestions des données, avec la gestion de nouveau format et des outils opérationnels en interne pour faciliter les ingestions.

L'offre logicielle du HDH comprend à ce jour :

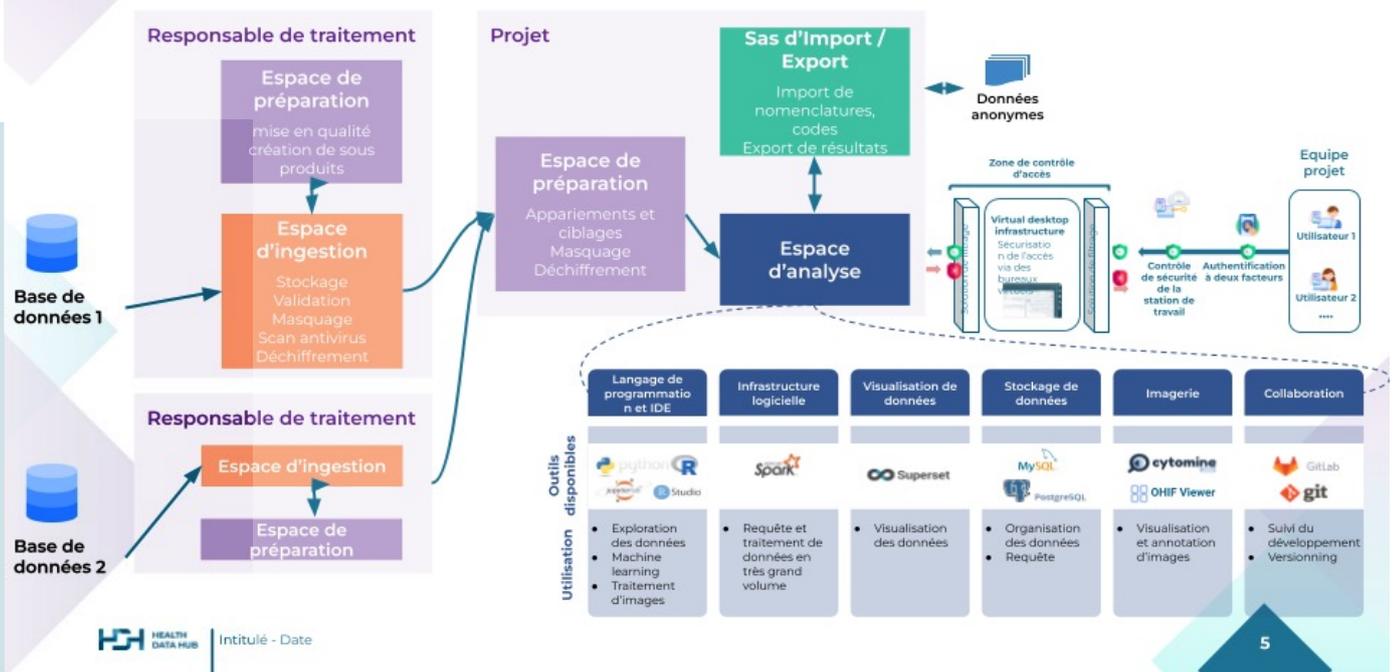
- **Les environnements de développement intégré (IDE) JupyterLab et RStudio** qui proposent de travailler respectivement avec les langages de programmation Python et R.
- Les outils de collaboration **Git et Gitlab**.
- **L'outil de calcul distribué Spark pour Jupyterlab**, permettant d'effectuer des analyses complexes à grande échelle.
- **Le logiciel Superset** offre la possibilité aux profils non data scientists (notamment aux médecins) d'explorer et de visualiser les données de façon simple et sans avoir besoin de connaissances en développement.
- Les systèmes de gestion de bases de données **PostgreSQL et MySQL**.
- Les outils de visualisation et d'annotation d'images via **Cytomine** (JPEG, PNG, TIFF, etc.) et **OHIF Viewer** (DICOM).
- **L'outil d'analyse de données Atlas**, permettant de créer des cohortes et d'effectuer des études scientifiques poussées sur des données converties au format OMOP.

Travaux de maintien des conditions de sécurité et de mise à jour de l'homologation

Le HDH met à disposition des porteurs de projets qu'il accompagne et qui le souhaitent, un **"espace d'analyse" sécurisé sur une plateforme technologique**. Cet espace permet notamment de recevoir, stocker, traiter et visualiser les données de santé nécessaires aux projets sans pouvoir les exporter. Le contexte de sécurité de la plateforme technologique du HDH a été défini, en respectant les principes de défense en profondeur, dans le cadre d'une démarche conduite en collaboration avec l'Agence nationale de la sécurité des systèmes d'information (ANSSI) et le fonctionnaire sécurité des systèmes d'information (FSSI) du ministère du Travail, de la Santé et des Solidarités.

La sécurité de la plateforme technologique est, en outre, testée et revue périodiquement par des prestataires spécialisés en sécurité des systèmes d'information indépendants qualifiés par l'ANSSI, dont trois ont déjà été réalisés avec succès depuis sa mise en œuvre. Depuis son homologation initiale en 2020 et la revoyure de 2021, le niveau de sécurité de la plateforme a continué d'être amélioré, et s'est soldé par **le succès de l'audit PASSI de novembre 2023**. En effet, un audit d'architecture ainsi que des tests d'intrusion de trois catégories ont été réalisés sur la plateforme technique. Ils ont permis de démontrer un excellent niveau de maturité de l'architecture de la plateforme, ainsi qu'un soin particulier pour garantir la sécurité des données, en termes de chiffrement et de maîtrise des mouvements. L'ANSSI et la CNIL sont invitées à l'homologation de chaque nouvelle version de la plateforme technique.

Focus : comment se compose la plateforme technologique ?



Par ailleurs, afin de garantir le plus haut niveau de sécurité des données de santé, **le HDH met en place son propre centre opérationnel de sécurité (SOC)**, qui vise à détecter au plus tôt les comportements anormaux afin de répondre aux incidents de sécurité de la manière la plus rapide et efficace possible, aussi bien au niveau de la plateforme technologique qu'au niveau de son patrimoine informationnel. En 2023, les travaux de développement se sont concentrés sur la construction du socle technique de la solution et sur son intégration dans l'écosystème du HDH. Son déploiement est prévu pour 2024.

Plateforme technologique : travaux de réduction des ressources mobilisées

Afin de diminuer les coûts d'exploitation de la plateforme technologique du HDH sans diminuer la qualité du service, le HDH a approfondi et systématisé en 2023 ses contrôles des ressources cloud utilisées. Pour ce faire, adoptant une approche FinOps, les équipes ont instauré **une surveillance quotidienne des consommations et de l'utilisation des services** fondée sur un reporting fin des variations de coûts. En parallèle, afin d'améliorer les prévisions de dépenses engagées, **la caractérisation des projets de recherche et de leurs besoins en ressources** a été revue. L'affinage se poursuivra en 2024 pour augmenter la précision des prédictions. Enfin, le HDH a élaboré en 2023 **une politique de maîtrise des services et de réduction des coûts associés** (recensement et optimisation des services utilisés, mesure de la volumétrie de stockage de données, pré-réservations de machines virtuelles). Ces actions, qui dans une logique d'amélioration continue se poursuivront en 2024, entendent concilier service de haut niveau et coûts maîtrisés.

Travaux pour répondre aux besoins croissants du portefeuille de projets

En 2023, dans le cadre de l'engagement continu dans l'optimisation des opérations, le HDH a renforcé ses efforts pour garantir **une documentation exhaustive, une mise en œuvre efficace et une vérification rigoureuse** de celle-ci, particulièrement face à un portefeuille de projets croissant. En outre, il a **intégré et consolidé l'usage d'outils de gestion** permettant l'amélioration de la gestion des incidents, l'accélération de la prise en charge des tickets, le renforcement des processus opérationnels tout en permettant une communication fluidifiée au sein des équipes. En parallèle, les comitologies dédiées ont été mises en place. Le HDH continue à explorer de nouvelles solutions technologiques et à les intégrer dans ses processus afin d'optimiser davantage les opérations et de répondre aux besoins croissants de son portefeuille de projets.

ENJEU

3

Renforcer les connexions du Health Data Hub aux acteurs de l'écosystème



Renforcer le travail partenarial avec les acteurs locaux

Accompagnement du projet EMC2, de l'Agence Européenne du médicament (EMA)

En 2021, l'EMA a retenu la candidature du HDH et de ses partenaires dans le cadre d'un **appel d'offre sur des données en vie réelle**. L'objectif de cet appel d'offre est de pouvoir **accéder à des bases de données individuelles hospitalières**

européennes, à la maille patient et au format OMOP-CDM. Pour y répondre, le HDH s'est associé à quatre établissements de santé : les Hospices civils de Lyon (HCL), le Centre Léon Bérard, le CHRU de Nancy et l'hôpital Saint Joseph, ainsi qu'à la plateforme de pharmaco-épidémiologie de Bordeaux, l'ANS et Capgemini. Il s'agit du projet "EMC2".



En 2023, la CNIL a accordé une autorisation de trois ans pour cet entrepôt, **premier EDS multicentrique⁸ contenant des données cliniques, chaînées à la base principale du SNDS, dans un modèle commun international**, permettant de les rendre interoperables. L'entrepôt sera hébergé sur la plateforme technologique du HDH, à laquelle pourront accéder les partenaires hospitaliers, l'EMA, ainsi que les porteurs de projets autorisés sous réserve de la gouvernance sur laquelle les partenaires s'accorderont.

Accompagnement du programme Unibase

Le HDH a signé, en juillet 2021, un partenariat avec **Unicancer**, dans le cadre du programme Unibase. Le 25 novembre 2021, les deux partenaires ont lancé un AMI ouvert aux dix centres de lutte contre le cancer (CLCC) disposant d'entrepôts de données ConSoRe, moteur de recherche en cancérologie. Trois projets pilotes, en phase avec le Plan cancer 2021-2030, ont été sélectionnés en 2022, avec, à la clé, un accompagnement et un financement du HDH.

Puis, une seconde vague de l'AMI, associant un troisième partenaire, **La Ligue contre le cancer**, a été ouverte notamment aux CLCC n'utilisant pas ConSoRe, ainsi qu'à d'autres structures hospitalières. **Cinq projets pilotes, en phase avec le Plan cancer 2021-2030, ont été sélectionnés en 2023**, bénéficiant d'un financement conjoint du HDH et de la Ligue contre le cancer, et d'un accompagnement dans le cadre de la réalisation de leur projet :

CASEBA est porté par l'Institut Curie. Le projet porte sur les cancers sévères de bas grade de l'ovaire, pathologie émergente et diagnostiquée en tant que telle depuis peu. L'objectif principal du projet est de nourrir une base de données multimodale et la mise en lien avec les données de la CNAM portant sur des patientes atteintes de ce cancer. Cela permettra de mesurer l'effet de la filière de prise en charge des patientes sur leur survie, ainsi que d'améliorer le diagnostic radiologique de ce

cancer afin de proposer une prise en charge spécifique et optimale aux patientes.

HORUS est porté par Gustave Roussy et l'équipe de l'Inserm "Épidémiologie des Radiations" du CESP-Unité 1018. Il vise à créer une cohorte inédite de patients traités avant l'âge de 25 ans pour un cancer pédiatrique et ayant reçu un traitement anticancéreux innovant. Cette cohorte sera constituée à partir de l'agrégation de données venant de différentes sources sur l'histoire néoplasique du jeune patient et de l'appariement avec la base principale du SNDS. Grâce à ces données, l'équipe évaluera la morbidité et la mortalité en vie réelle de ces patients afin d'évaluer leur toxicité potentielle pour mieux les prévenir, dépister et traiter.

HYPO-RT Prostate est porté par le Centre Oscar Lambret. Le projet porte sur les nouvelles techniques de radiothérapie, dites hypo-fractionnées, appliquées au cancer de la prostate. Celles-ci consistent à augmenter la dose délivrée au patient à chaque séance, en diminuant le nombre de séances par rapport aux radiothérapies classiques. Le projet vise à déterminer la diffusion de ces nouvelles techniques, leur accessibilité par les patients et à évaluer en vie réelle leur efficacité et toxicité.

UNI-AJA est porté par le Centre Léon Bérard et l'Institut de cancérologie de Lorraine. Il a pour objectif de constituer une cohorte rétrospective d'adolescents et de jeunes adultes (15-29 ans) atteints de cancer ou d'hémopathie maligne et pris en charge dans 10 établissements de santé sur la période 2009-2022. Cette cohorte a pour objectif d'étudier le parcours de soin et le devenir de ces patients. Elle sera mise en lien avec les données de la base principale du SNDS afin d'enrichir les informations concernant cette population.

⁸ Les projets multicentriques répondent à de nombreux enjeux dont la mutualisation de données issues d'environnements différents, l'investigation de questions communes, etc.

UNIRIV est porté par le Centre Léon Bérard et l'Institut Gustave Roussy. Il est consacré à une nouvelle thérapie du cancer de la prostate métastatique par radiothérapie interne vectorisée. Ce projet vise à identifier des biomarqueurs issus de l'imagerie scintigraphique qui permettront de prédire l'efficacité du traitement pour une meilleure prise en charge du patient grâce à la centralisation des données cliniques, biologiques et d'imagerie d'un grand nombre de patients issus de plusieurs centres.

Accompagnement des EDS "locaux"

Comme mentionné dans ce document, pour accélérer l'émergence des EDS à l'échelle nationale, le ministère du Travail, de la Santé et des Solidarités, en lien avec le HDH, a lancé en 2022 un AAP en deux vagues, doté d'une enveloppe totale de 75 millions d'euros. En 2023, **16 lauréats ont été retenus** pour la pertinence et le potentiel de leurs projets. Le HDH participe à leur accompagnement : sessions collectives de formation pour soutenir les porteurs de projets dans la construction de leur EDS (mobilisant des experts du HDH, de la CNIL et de la DINUM), organisation de groupes de travail dans le cadre du comité stratégique des données de santé et auxquels les lauréats de l'AAP EDS ont été sollicités pour participer. En 2024, le HDH poursuivra ces actions.

Accompagnement de l'appel à projets DATAE

Afin d'entretenir la forte dynamique mise en œuvre par les établissements de santé dans le développement de leurs EDS, la DGOS a lancé en 2023 l'AAP DATAE (Données de santé et Applications) dont l'opération est confiée au HDH (constitution du cahier des charges, conception du processus de sélection, composition du jury, etc.). Cet AAP, destiné aux projets de recherche mobilisant des données collectées dans les EDS, **visent à être un dispositif pérenne, reconduit annuellement, afin de répondre au fort intérêt et aux usages croissants** dans les établissements. DATAE est ouvert aux cas d'usage dans tous les domaines de la recherche appliquée en santé mais les projets

doivent nécessairement utiliser des données issues d'un EDS hospitalier. Les lauréats seront annoncés en 2024.

Contribution aux initiatives des acteurs de l'innovation en matière de données de santé

PariSanté Campus : coeur du numérique en santé et partenaire clé pour diffuser la connaissance

Créé en 2021, PariSanté Campus (PSC) est un pôle d'excellence unique au monde, pour la recherche, l'innovation, la formation et l'entrepreneuriat en santé numérique. Il a été cofondé par l'Inserm, Inria, l'université Paris Sciences & Lettres (PSL), l'ANS et le HDH. En tant que membre co-fondateur, le HDH participe aux instances de PSC et a notamment suivi avec intérêt la construction du programme scientifique du campus, visant à faciliter les interactions entre les différents acteurs de la e-santé. En 2023, le HDH a contribué à **un programme d'accélération de prévention numérique à destination des start-up**, porté par PSC et la Bpifrance. Par ailleurs, en 2023, le HDH a organisé **treize événements à PSC** contribuant à la diffusion d'une culture de la donnée de santé, et s'est impliqué dans l'animation de PSC, en l'occurrence lors du Campus live!, de l'AI4Health, de France biotech et de Vivatech.

Le Health Data Hub, membre fondateur de l'association Filière Intelligence Artificielle et Cancers

La Filière IA et Cancers (FIAC) est une association qui fédère les acteurs publics et privés au sein d'une initiative nationale. Le projet souhaite bénéficier à **l'ensemble de l'écosystème en oncologie**, et permettre le développement de solutions thérapeutiques et diagnostiques innovantes. L'association réunit à ce jour, l'Institut national du cancer, le HDH, l'Alliance pour la Recherche et l'Innovation des Industries de Santé (ARIIS), France Biotech et neuf industriels de santé : Amgen, AstraZeneca, Bristol-Myers Squibb, Janssen-Cilag, MSD, Novartis, Pfizer, Pierre fabre, et Roche Diagnostics. **Cinq projets reposant sur la réutilisation de données en cancérologie (PRC)** ont



été approuvés par les membres fondateurs de l'association. Ils concernent notamment les cancers du poumon, broncho-pulmonaire et bronchique. Depuis le début de son implication, et également en 2023, la FIAC et le HDH travaillent à la définition de convergences communes en oncologie.

Lancement d'un projet conjoint avec le Green Data For Health, dispositif du Commissariat général au développement durable (CGDD)

En 2022, le Health Data Hub s'est associé avec le GD4H, dispositif du CGDD visant à favoriser le recueil de données environnementales. Ensemble ils ont lancé un AAP conjoint "La donnée pour la recherche et l'innovation en santé-environnement". **Quatre projets de recherche ont été sélectionnés** et bénéficieront d'un accompagnement humain et technique dans leur mise en œuvre pour une durée de 18 mois. Fort de ce succès, **une seconde vague de cet AAP conjoint a été lancée en décembre 2023**. Ouvert aux acteurs publics ou privés, cet AAP s'adresse à tout projet d'excellence en recherche reliant données de santé et facteurs environnementaux.

Par ailleurs, en 2023, la coopération entre le GD4H et le HDH s'est étoffée avec la conception d'une **formation citoyenne**, bénéficiant de l'expertise de l'INERIS, et visant à sensibiliser les citoyens à ces données et leur usage pour la recherche. Aussi, le GD4H et le HDH ont co-animé des groupes de travail du Conseil National de l'Information Géolocalisée (CNIG), dans le cadre desquels **trois livrables ont été produits conjointement** : un rapport sur les possibilités de géolocalisation des données de santé et une synthèse des besoins et usages clés des données environnementales et sanitaires en santé-environnement. Dans ce cadre, le GD4H a également produit un état des lieux des ressources structurantes par sous-groupe (référentiels, standards, ressources structurantes).

Paris-Saclay Cancer Cluster, accélérateur de l'innovation en oncologie

Créée en février 2022 à l'initiative de l'université Paris-Saclay, de l'Institut polytechnique de Paris, de l'Inserm, de l'Institut Gustave Roussy et de Sanofi, PSCC rassemble **un continuum d'excellence de près de quatre-vingt membres de neuf nationalités** : organismes de recherche, universités, hôpitaux, start-up, scale-up, industriels, investisseurs en capital, associations de patients, gouvernement, collectivités et acteurs du territoire. Ce cluster en oncologie vise à **favoriser les échanges au sein de la communauté, à renforcer l'offre disponible pour les projets innovants et à fournir aux chercheurs et développeurs européens les plus prometteurs l'accès aux meilleures expertises, à des outils et des services du domaine** : mentors, réseaux, financements, données, échantillons, technologies, infrastructures et laboratoires. Le HDH échange ainsi avec PSCC afin de définir des axes de coopération.

Identification d'autres acteurs : focus prévention

En 2023, dans le cadre de ses réflexions, le HDH s'est penché sur le sujet de **l'utilisation des données de santé dans le cadre de la prévention en santé**. Ainsi a été mené un grand nombre de rencontres et d'échanges avec des acteurs institutionnels (DNS, AIS, Bpifrance, DGCS, Assemblée nationale, etc.) qui mènent des réflexions stratégiques, en cohérence avec les futurs objectifs de la stratégie nationale de santé et les priorités définies par le ministère du Travail, de la Santé et des Solidarités, ainsi qu'avec des acteurs proposant des projets concrets de prévention en santé (notamment sur le "bien vieillir", la télémédecine, le financement des actions de prévention). Ces rencontres ont permis au HDH d'identifier des sources de données de santé supplémentaires et aux porteurs de projets, de bénéficier des expertises du HDH, en termes, notamment, de soutien à la collecte, de consolidation et de mise en valeur de leur patrimoine de données.

Renforcer la présence du Health Data Hub au niveau européen et international

Participation au projet « Toward a European Health Data Space » (TEHDaS)

L'action conjointe TEHDaS, lancée en février 2021 coordonnée par SITRA, le fonds pour l'innovation finlandais et impliquant vingt-six pays européens, avait pour objectif de **formuler des recommandations à la Commission européenne pour le futur EHDS**. Elle a pris fin le 31 juillet 2023, marquée par une assemblée générale à Helsinki, où les partenaires ont passé en revue les résultats du projet. Pendant cette période, le HDH a coordonné les contributions des acteurs français au sein de TEHDaS, mobilisant ainsi l'ANS, l'Inserm, les HCL, les universités Aix-Marseille et Toulouse II. La contribution française s'est concentrée sur deux aspects, **l'engagement citoyen et la gouvernance**.

Plus de
6000
contributions lors
d'une consultation en
ligne pour la création
de l'EHDS

Pour ce qui est de l'engagement citoyen, une consultation en ligne organisée par le HDH, avec Sciensano et la NHS Confederation, a rassemblé **plus de 6 000 contributions**, une étape qui a pris fin en avril 2022. Celle-ci a conduit à **la publication de douze recommandations** en mars 2023, visant à mieux informer le public sur l'utilisation des données de santé et à favoriser sa participation dans les processus décisionnels liés à l'EHDS. Concernant la thématique de la gouvernance, un atelier organisé le 29 mars 2023 par le HDH a réuni une quarantaine de participants européens, notamment d'Espagne, de France, des Pays-Bas, de Suède et de Finlande. Le rapport de cet atelier a souligné le besoin de clarifier les rôles et responsabilités des différents acteurs dans le futur espace européen des données de santé et a formulé plusieurs recommandations pour éclairer les aspects ambigus du projet de règlement. Les travaux initiés dans le cadre de TEHDaS se poursuivront en 2024 dans une nouvelle action conjointe, qui devrait également être coordonnée par SITRA.

Travaux dans le cadre du projet HealthData@EU

Lancé en octobre 2022 et coordonné par le HDH, le projet HealthData@EU Pilot réunit dix-sept acteurs européens majeurs, parmi lesquels les plateformes nationales de données de santé de plusieurs Etats membres (Finlande, Norvège, Danemark, Allemagne, Belgique, Hongrie, Croatie et Espagne), deux agences européennes (l'EMA et Centre européen de prévention et de contrôle des maladies - ECDC), ainsi que des infrastructures de recherche européennes

(BBMRI, Elixir, EBRAINS). Les travaux de ce projet portent sur **la création d'une première version de l'infrastructure transfrontière HealthData@EU**, prévue dans le projet de règlement pour l'EHDS. Ce réseau connectera tous les partenaires fournisseurs de données – nœuds de données –, ainsi que les services centraux établis au niveau de la Commission européenne. Les activités liées à la conception et au déploiement de l'infrastructure technique pour le réseau de nœuds et la liaison avec les services centraux de la Commission européenne sont menées conjointement par le HDH et le Health Data Lab allemand. Cette infrastructure technique doit permettre la réalisation des deux services prioritaires du parcours utilisateur (consultation de métadonnées dans un portail européen, remplissage d'un formulaire de demande d'accès).

La phase de cadrage, qui s'est terminée au premier trimestre 2023, a abouti à **un plan de développement et de déploiement du réseau en plusieurs phases**. La forte implication des équipes a permis de livrer une première version opérationnelle (MVP) de l'infrastructure en décembre 2023. Les travaux se poursuivront en 2024 avec l'adaptation de l'infrastructure à partir de plusieurs livrables clés du projet, tel que le formulaire unique de demande d'accès et le nouveau standard de métadonnées, Health DCAT-AP et la connexion à l'infrastructure de partenaires supplémentaires.

Participation au “Direct Grant”

Le HDH a mobilisé en 2023 **un consortium de dix-neuf partenaires** responsables de données, pour répondre à l'appel à financement Direct Grant Setting up services by Health Data Access Bodies (HDAB) du programme EU4Health visant à accélérer la création des organismes responsables de l'accès aux données de santé (HDAB) prévus dans le projet de règlement EHDS dans les États membres. Les partenaires du projet sont : l'AP-HP, le Centre Léon Bérard, le CHU de Lille, le CHU de Bordeaux, le CHU de Brest, le CH de Chartres, le CHU de Dijon, le CRCDC en Occitanie, FranCIM, l'Inserm, Unicancer la CNAM, le CHU de Montpellier, le GCS Groupement Grand Est, l'Institut national du cancer, Roche, GCS hôpitaux universitaires

du Grand Ouest, l'AFM-Téléthon et l'Association Vaincre la Mucoviscidose. La Commission européenne a approuvé le projet français, pour un budget total de 4,6 millions d'euros, favorisant la réutilisation des données de santé à l'échelle nationale et européenne. Ces fonds renforcent l'engagement de la France et du HDH dans la mise en place de l'EHDS.

Participation au projet TEF-Health

Le HDH participe, avec plus de cinquante organisations partenaires, au projet européen Testing and Experimentation Facility, Health (TEF-Health). Le consortium comprend **des partenaires issus de neuf pays** (Allemagne, Belgique, France, Finlande, Italie, Portugal, République tchèque, Slovaquie et Suède) ainsi que des partenaires opérant au niveau européen, comme l'infrastructure de recherche EBRAINS.

Lancé en janvier 2023 pour une durée de cinq ans, ce projet est dirigé par l'hôpital universitaire de Berlin la Charité et est cofinancé à hauteur de 50% par l'UE. Son objectif est de **soutenir le développement de l'IA en santé en fournissant aux innovateurs des infrastructures et services de tests** pour accélérer la mise sur le marché de solutions novatrices en matière d'algorithmes et de robotique. TEF-Health vise à toucher les acteurs de l'innovation en santé dans toute l'UE, avec une quarantaine de start-up et PME ayant exprimé leur intérêt dès son lancement.

Le HDH est engagé aux côtés de nombreux autres acteurs français au sein du “noeud français” de l'initiative : le Laboratoire national de métrologie et d'essais (LNE) en tant que leader du noeud français, le Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA) Grenoble, le CHU de Grenoble, l'Institut Hospitalo-Universitaire (IHU) de Strasbourg, le CHU de Rennes, les HCL et EIT Health France. Au sein du projet, le HDH contribue aux travaux relatifs au **développement d'un catalogue de services pour les futurs utilisateurs de TEF et à l'identification de cas d'usage concrets soutenant des start-up** avec l'offre de services du HDH et de TEF plus largement.

Identification des appels à projets d'intérêt

Le HDH suit attentivement les AAP européens en matière de santé numérique, mettant un accent particulier sur les programmes de financement Horizon Europe, "DIGITAL" Europe et l'UE pour la Santé (EU4Health). En 2023, on compte parmi les projets identifiés, et auxquels le HDH encourage des partenaires français à contribuer : le Direct Grant Setting up Services by Health Data Access Bodies, l'action conjointe de prévention des cancers et des maladies non-transmissibles (JA PREVENT NCD), l'action conjointe EUCanScreen, visant à améliorer la détection précoce du cancer grâce au dépistage. Dans le cadre de ces actions conjointes, et grâce à son expertise autour de l'interopérabilité des données, le HDH contribuera à deux enjeux majeurs de santé publique : **la prévention et le dépistage**. En 2024, le HDH ambitionne de participer à nouveau à plusieurs appels à financements européens.

Identification des acteurs internationaux d'intérêt

En 2022, le HDH est allé pour la première fois à la rencontre d'acteurs⁹ israéliens et japonais pour en savoir plus sur la réutilisation des données de santé dans leurs contextes nationaux respectifs. Malgré des cadres réglementaires et des contextes nationaux sensiblement différents, ces acteurs partagent à beaucoup d'égards, les préoccupations françaises et européennes en matière de collecte et de réutilisation des données de santé. Le HDH a œuvré en 2023 pour maintenir ces échanges réguliers et développer des pistes de collaborations comme la participation de représentants du réseau japonais MID-NET aux auditions organisées dans le cadre des travaux sur les entrepôts de données de santé¹⁰ ; la mise en place d'un partenariat avec le Technion-Rambam permettant à cinq étudiants de participer à la Summer School AI4health ; l'accueil et la

présentation du HDH à une délégation de représentants de la fédération d'entreprises japonaises Keidanren lors de leur visite à Paris ; l'accueil de représentants du National Cancer Center et l'organisation d'une réunion de travail avec les représentants d'Unicancer, SeqOIA et Auragen. Les travaux d'instruction de nouvelles opportunités partenariales entamés en 2023 se poursuivront donc en 2024 avec l'organisation de délégations et d'échanges avec des acteurs du Royaume-Uni.

Poursuite des appels à projets internationaux

Suite à l'accord de coopération internationale signé en juillet 2022 entre le Fonds de recherche Québec-Santé (FRQS) et le HDH, les deux partenaires ont sélectionné en 2023 **deux nouveaux projets** grâce à un AAP conjoint ouvert, en 2023, pour la deuxième année consécutive. Lors de cette seconde vague de l'AAP "IA et Santé" visant à soutenir des collaborations scientifiques franco-québécoises, le comité de sélection a accordé une attention particulière aux projets dont les résultats seront partagés en open source, afin que les innovations en IA en santé puissent bénéficier à l'ensemble de la recherche. Les deux nouveaux projets sélectionnés ont été dévoilés lors d'un webinaire organisé le 28 novembre. Ils bénéficieront d'un financement et d'un accompagnement sur 18 mois.

9 Il s'agissait des principaux producteurs, utilisateurs et régulateurs de l'utilisation secondaire des données de santé à l'échelle nationale en Israël et au Japon (ministères de la santé, les caisses nationales de l'Assurance Maladie, les principaux hôpitaux, centres et instituts de recherche en santé).

10 Opéré par l'agence japonaise de surveillance de sécurité des médicaments et des dispositifs médicaux, le réseau MID-NET rassemble les données cliniques médico-administratives de près d'une trentaine d'hôpitaux et organismes de santé à travers le Japon, représentant près de 6 millions de patients.

Zoom sur les projets sélectionnés

Prise en charge des patientes atteintes d'un cancer du sein à l'ère des grands modèles de fondation.

Porté par l'École de technologie supérieure de Montréal, l'université Paris-Saclay et l'Institut Gustave Roussy, ce projet mobilise les données de 10 000 patientes et 5 millions d'extraits d'images de lames de tissus afin de consolider l'apprentissage de modèles

d'IA destinés au diagnostic et au suivi du cancer du sein. Dans une démarche de science ouverte, les chercheurs prévoient notamment de publier leur code en open source.

Livre blanc des données synthétiques en santé.

Le Centre hospitalier de l'Université de Montréal (CHUM) et le CHU de Nantes souhaitent mobiliser des données synthétiques

générées à partir des données de plus de 700 patients québécois et français afin d'étudier les facteurs de rupture d'anévrisme. L'objectif sera de proposer un livre blanc sur la génération via l'IA de données synthétiques interopérables en santé pour mieux accompagner l'ensemble des professionnels de santé.

Réunir les conditions d'un partage efficace en termes de standards, gouvernance d'accès et de valorisation dans un contexte national et européen

Identification des acteurs pour enrichir les connaissances en matière d'accès aux données de santé dans le monde

Dans le cadre du projet pilote HealthData@EU, le HDH a lancé en décembre 2022 un AMI pour la mise en place d'un conseil scientifique consultatif. Plus de 250 candidatures provenant de plus de 20 pays différents ont été reçues, et 49 membres sélectionnés (institutions de recherche et d'enseignement, de l'industrie, des autorités publiques aux organisations de patients et de citoyens). Les activités des différents groupes de travail du conseil ont débuté en avril 2023 et se poursuivront tout au long du projet, contribuant de manière significative à façonner l'avenir de l'EHDS.

Contribution aux groupes de travail du projet HealthData@EU : formulaire de demande d'accès aux données de santé européen, et identification de standards de données pertinents

Dans le cadre du projet pilote HealthData@EU le HDH contribue notamment au Working Package (WP) 7 "conformité réglementaire et juridique", et au WP8 "interopérabilité et qualité". Le WP7 a développé en 2023 un des livrables clés

du projet : **un formulaire unique de demande d'accès aux données**. L'objectif était de concevoir un formulaire facile d'utilisation pour les demandeurs, respectant les exigences du règlement EHDS et du RGPD. Pour simplifier la procédure, une demande unique en une étape a été préconisée, afin de faciliter le processus tant pour les demandeurs que pour les organismes nationaux d'accès aux données. Le travail sur le formulaire a été mené par la plateforme de données de santé finlandaise, Findata. À partir d'une cartographie des processus de demandes d'accès existants, et des préconisations du projet de règlement EHDS, un groupe de travail a élaboré ce formulaire tenant compte des besoins et expériences de plusieurs pays européens dont l'Allemagne, le Danemark, la Finlande, la France et la Norvège. Le WP7 a également produit un rapport sur les conditions générales d'utilisation des données et les mesures de sécurité, mettant l'accent sur le parcours d'accès et de mise à disposition des données observé dans le cadre des cinq cas d'usage du projet.

Le HDH contribue aussi aux travaux du WP8 sur l'interopérabilité et la qualité des données, mené par l'infrastructure de recherche européenne BBMRI-ERIC. Celui-ci a pour objectif **d'élaborer des lignes directrices** concernant : les standards à utiliser pour permettre l'interopérabilité sémantique, les critères de qualité des données pour permettre une interprétation fondée des résultats, et les prérequis de sécurité pour le transfert de données. Ces recommandations sont formulées à partir de l'observation des difficultés rencontrées et des solutions adoptées dans le cadre des cas d'usage du projet. Des recommandations préliminaires ont été présentées au consortium en mars 2023, et les recommandations finales sont attendues pour fin 2024.

Travaux sur l'impact des règlements européens sur la gouvernance d'accès aux données de santé en France et sur l'IA Act

En 2020, la Commission européenne a publié la stratégie "Façonner l'avenir numérique de l'Europe" comportant, comme axe clé, **la mise en œuvre d'une stratégie européenne pour les données** qui devait se matérialiser par une série d'actes législatifs. Depuis, plusieurs réglementations ont été adoptées (Data Governance Act, NIS 2, Règlement sur les services numériques, Règlement sur les marchés numériques) ou sont en cours de l'être (EHDS, IA Act, Data Act). Ces réglementations vont largement modifier le cadre juridique de la réutilisation des données de santé. Pour anticiper ces réglementations, le HDH a initié une collaboration avec la Chaire droit et éthique de la santé numérique de l'Institut catholique de Lille avec **la mise en place d'un post-doctorat** sur les enjeux normatifs pour le partage de données de santé à l'échelle européenne.

En 2023, le post-doctorat a abouti à **des analyses approfondies de l'ensemble de ces textes** qui permettent de mettre en lumière les évolutions du cadre réglementaire et leurs impacts sur l'accès aux données. Ces travaux permettent d'alimenter les réflexions menées au sein du HDH, et contribuent ainsi à **l'amélioration continue de l'accompagnement des projets et de la diffusion des connaissances** sur l'accès aux données à l'écosystème.

Travaux de mise à jour du référentiel de sécurité du SNDS, en lien avec le ministère du Travail, de la Santé et des solidarités

Alors que le SNDS réunissait historiquement les données de la base principale du SNDS (Sniiram, PMSI, etc.), la loi relative à l'organisation et la transformation du système de santé (OTSS) de 2019 l'a élargi à de nouvelles données hébergées sur des systèmes d'information éparpillés sur le territoire, et non plus nécessairement centralisés. Or, le référentiel de sécurité encore en vigueur, pris en application des anciennes dispositions, ne s'applique pas à ces systèmes. Ainsi, de multiples systèmes d'information hébergent et mettent à disposition des données du SNDS sans qu'une protection efficace en matière de sécurité (hébergement, transmission, pseudonymisation, contrôle, etc.) ne soit prévue par les textes. Ainsi, durant l'année 2023, le HDH a participé aux **travaux de mise à jour du référentiel de sécurité du SNDS** en lien avec la DREES. Cette mise à jour est importante pour renforcer la robustesse des systèmes d'information hébergeant des données du SNDS et ainsi développer la confiance des citoyens concernant l'usage de leurs données. La publication du référentiel interviendra en 2024.

Réunir les conditions d'un partage efficace en termes de normes de sécurité, de souveraineté dans un contexte national et européen

Travaux sur les exigences de sécurité communes dans le cadre de l'EHDS

Dans le cadre des travaux relatifs à la conception et au déploiement de l'infrastructure technique européenne HealthData@EU (cf. supra), le Working Package 5 du projet pilote HealthData@EU s'intéresse aux exigences de sécurité communes. Dans ce contexte, le HDH et le Health Data Lab allemand ont conduit en septembre 2023 **une analyse des risques portant sur la disponibilité, l'intégrité, la confidentialité et la traçabilité de l'infrastructure**. Après avoir identifié ces risques et menaces, tels que l'impossibilité des utilisateurs de données d'accéder au catalogue central de métadonnées, plusieurs partenaires du projet (BBMRI, CIPH, ECDC, EHelse, Findata, HDL, HDH, IACS, Orphanet et Sciensano) ont évalué les impacts de chaque risque et la probabilité d'occurrence de chaque menace identifiée. Cette analyse des risques servira de base à la suite des travaux de développement et de déploiement de l'infrastructure en 2024 en collaboration rapprochée avec les services centraux de la Commission européenne.

Travaux autour du démonstrateur de la Direction Interministérielle du Numérique

En 2023, le HDH a poursuivi ses travaux avec la DINUM, notamment sur la mise en place d'un démonstrateur Cloud de confiance dont les objectifs sont de collecter les besoins des administrations en termes de cloud, d'analyser les offres de services disponibles au regard des besoins identifiés et de construire des défis techniques à destination des fournisseurs, afin de les alimenter de retours factuels sur les services prioritaires à mettre en oeuvre au profit des administrations et opérateurs de l'État. Le HDH a ainsi contribué à la définition des défis techniques, notamment du premier défi portant sur les aspects de sécurité, en amont de l'AAP France 2030 lancé au premier trimestre 2024.

Ces travaux ont également bénéficié au rapport d'expertise technique portant sur la faisabilité d'héberger l'entrepôt de données de santé multi-centrique EMC2, dont le HDH est responsable de traitement, sur une plateforme souveraine. Piloté par la DNS, la DINUM et l'Agence du numérique en santé (ANS), l'audit a montré qu'à ce jour, aucun des hébergeurs étudiés n'est en capacité de fournir les conditions de sécurité nécessaires à l'hébergement de données de santé selon des exigences similaires à celles auxquelles le HDH est actuellement soumis. Suite à ces conclusions, la CNIL a autorisé l'entrepôt EMC2 à être hébergé sur la plateforme technologique du HDH. Les exigences prioritaires du HDH et d'EMC2 continueront d'être intégrées aux défis techniques développés dans le cadre du démonstrateur Cloud de confiance en 2024.

Informar les citoyens sur les enjeux des données de santé consiste à relever de multiples défis : informer largement (le plus de citoyens possibles), sensibiliser de façon pédagogique au fait que les données de santé sont utilisées pour la recherche et que les citoyens ont des droits, et faire connaître plus en profondeur comment cela fonctionne. C'est pourquoi une newsletter citoyenne, en partenariat avec France Assos Santé (FAS), a été lancée en 2022, et s'est poursuivie en 2023. Un nombre d'abonnés croissant peut ainsi être tenu informé des études portées par les associations, et de l'ensemble des événements et projets citoyens en relation avec la seconde vie des données.

ENJEU

4

Être à l'écoute de la société civile et co-construire une culture de la donnée de santé



Informier largement sur les enjeux de la réutilisation des données de santé

Le défi de l'information est aussi de **sensibiliser**, en allant au plus près des patients et de la société civile ; mais aussi de **concilier information légale et pédagogique**. C'est en ce sens qu'une expérimentation d'affiches dans les salles d'attente d'un établissement de soins a été lancée en 2022, et s'est clôturée en 2023. Le CHU de Rouen Normandie dispose d'un entrepôt de données de santé, EDSaN, pour lequel a déjà été réalisé l'information collective notamment grâce au travail des coordinateurs de cet entrepôt, du délégué à la protection des données (DPO) et des usagers du système de santé de l'établissement. Avec la participation du Datalab Normandie, la Région Normandie, France Assos Santé, l'école de design Designlr, le CHU et le HDH ont opté pour une démarche complémentaire : une campagne de communication via des affiches pour permettre aux patients de prendre pleinement conscience des bénéfices de l'utilisation des données pour la recherche et à terme, de mieux comprendre leurs droits relatifs aux données de santé et les exercer de façon éclairée.

Pour compléter les initiatives de sensibilisation autour de l'usage des données de santé, **deux rapports ont été publiés** en 2023 : l'un sur [l'état d'avancement des projets accompagnés par le HDH](#), et le second, issu d'un groupe de travail du conseil scientifique consultatif du HDH, sur la [balance bénéfiques/risques de l'usage des données](#).

Enfin, pour permettre aux personnes de connaître la seconde vie des données avec plus de précisions encore, le HDH a réalisé **plusieurs programmes de formations** en 2023. Il a aussi lancé la première édition du nouveau **diplôme universitaire (DU) sur la réutilisation des données** pour la recherche en santé, avec les universités de Bordeaux, Aix-Marseille et Paris-Cité. Il est également partenaire de SN@SU, lauréat de l'AMI Compétences et Métiers d'avenir en Santé numérique.

Les Mardis de la donnée de santé

Le HDH a réalisé plusieurs programmes de formations "citoyens" mis en ligne en 2023, chacun composé de plusieurs modules accessibles et téléchargeables en ligne depuis son site internet :

- **Programme 1 : système national des données de santé**, huit modules. En partenariat avec la CNAM, la CNIL, la DREES et FAS.
- **Programme 2 : entrepôts de données de santé**, cinq modules. En partenariat avec la CNAM, la CNIL, la DREES, FAS, et la participation de l'AP-HP.
- **Programme 3 : données de santé-environnement**, deux modules. En partenariat avec Ecolab, le regard expert de l'INERIS, et la participation de Santé Publique France, la Société Francophone de Santé et Environnement (SFS), l'Institut national de la recherche agronomique (INRAE), Générations futures, l'Office français de la biodiversité (OFB), FAS, Open Data France.

Chacun de ces programmes et modules est dispensé pendant "Les Mardis de la donnée de santé" : des webinaires qui se déroulent certains mardis, au moins une fois par mois, entre 12h et 12h45. En 2023, l'ensemble du premier programme a ainsi été traité pendant ces temps.

Plus de
2000
connexions pour
les 8 premiers
webinaires



Écouter et impliquer la société civile

Le HDH repose sur un principe fondateur d'écoute, d'ouverture et de dialogue avec la société civile qu'il impulse dans chacun de ses projets via, notamment une direction dédiée. En 2023, quatre méthodes d'écoute ont été mises en œuvre : étude mixte, conférence de consensus, concertation et débat lycéen. Chacune d'elles portant sur :

- **La place des citoyens dans le futur espace européen des données de santé**, rapport d'étude mixte - qualitative et quantitative, TEHDaS, le 31 mars.
- **Les modalités de sensibilisation des citoyens à la seconde vie des données**, rapport des résultats de la conférence de consensus citoyenne, les 1^{er} et 2 avril.
- **La place pour les usagers dans les entrepôts de données de santé**, colloque de restitution de la concertation, le 11 juillet.
- **Les questions éthiques de l'IA en santé et des données pour la recherche**, lancement effectif du premier atelier du débat lycéen, le 29 novembre.

Aussi, deux groupes de dialogues se sont mis en place en 2023 :

- **Un groupe d'associations de patients**, ayant produit un glossaire citoyen des types de bases de données (cohorte, entrepôts de données de santé, panel, registre...).
- **Un groupe d'associations d'étudiants** ayant produit une formation vulgarisée sur l'IA en santé et ayant construit une campagne de sensibilisation à la seconde vie des données. Un groupe composé de : l'Union Nationale

des Étudiants en Chirurgie Dentaire (UNECD), l'Association Nationale des Étudiants en Pharmacie de France (ANEPF), l'Associations Nationales des Étudiants en Médecine de France (ANEMF), l'Association Nationale des Étudiants Sages-Femmes (ANESF), la Fédération Nationale des Étudiant.e-s en Kinésithérapie (FNEK), la Fédération Nationale des Étudiants en Sciences Infirmières (FNESI) et la Fédération des Associations Générales Étudiantes (FAGE).

Conférence de consensus sur les données de santé : initiative lauréate de la 8e édition des Trophées de la participation et de la concertation

La conférence de consensus sur les données de santé pour la recherche est une initiative du HDH et de FAS, avec le soutien de PSC, et l'accompagnement du cabinet spécialisé en concertation publique, État d'Esprit Stratis, s'inscrit aussi dans l'engagement du HDH au Partenariat pour un gouvernement ouvert (PGO), initiative portée par le ministère de la transformation et de la fonction publique visant à favoriser la transparence, la participation et la collaboration avec la société civile.

Cette conférence a réuni en avril 2023, **soixante-dix citoyens venus de la France entière** pour leur permettre de se saisir pendant un week-end de la question suivante : "Comment sensibiliser la société au partage des données de santé ?". Ils ont été invités à porter une réflexion globale sur le sujet : pouvait-il être discuté, et si oui, par quels moyens ?

Pendant ce week-end, les participants ont prouvé que les citoyens sont capables de discuter de ce sujet complexe avec des mots simples. Ils se sont positionnés rapidement en faveur d'un **élargissement du partage des données de santé**, l'associant à un vecteur d'amélioration pour la santé des personnes. Aussi, selon eux, dédramatiser le sujet du partage des données de santé est essentiel. À l'issue des deux jours, les participants ont fait des propositions de communication pour parler à la société de ce sujet, qui ont été publiées, et serviront d'inspiration et d'orientation aux futures actions de sensibilisation qui seront mises en place par le HDH et ses partenaires. La conférence a été lauréate de la 8e édition des Trophées de la participation et de la concertation.

Développer les partenariats avec les associations d'utilisateurs

Pour répondre à une attente forte des associations, le HDH a lancé une expérimentation de "requête à la demande". Lorsque les données ne sont pas en open data pour dénombrer et caractériser, par exemple, une population de personnes atteintes par une pathologie donnée, les associations d'utilisateurs du système de santé souhaitent accéder à quelques chiffres complémentaires ou plus adaptés. La demande d'accès standard, nécessitant le dépôt d'un protocole d'étude, serait à ce titre totalement disproportionnée au besoin. Le service proposé par le HDH consiste à **fournir aux demandeurs des indicateurs tels qu'un dénombrement de patients ou d'actes médicaux**. Cela leur est utile en soi, mais c'est aussi une façon pédagogique de les initier à l'usage en les informant de ce qu'est le SNDS, des conditions d'accès, en rencontrant des data scientists et en comprenant mieux leur travail, etc. Ainsi, en 2023, **deux requêtes ont été finalisées** auprès de l'AFM-Téléthon et de l'Afa Crohn RCH France (Afa). Aussi, a été initiée la mise en place du guichet des requêtes à la demande, au sein de FAS.

Projet MICI - Afa Crohn RCH France, mieux connaître les populations atteintes de maladies inflammatoires chroniques de l'intestin

Les maladies inflammatoires chroniques de l'intestin (MICI) comprennent la maladie de Crohn et la rectocolite hémorragique (RCH). Cela se caractérise par une inflammation de la paroi du tube digestif. Au quotidien, les patients peuvent avoir des douleurs abdominales, des diarrhées, des douleurs anales, des pertes d'appétit, des nausées et vomissements, etc. Ils vivent chaque jour avec ces symptômes qui handicapent leur quotidien. Les MICI évoluent par périodes de poussées, avec des durées et intensités variables. Le diagnostic de ces pathologies peut prendre plusieurs années. Les facteurs qui en sont responsables demeurent encore mal compris à ce jour. En Europe, 3 millions de personnes seraient concernées, bien que ces chiffres varient selon les publications scientifiques. Il existe actuellement des données en open data pour caractériser les personnes atteintes de cette pathologie (nombre de personnes concernées selon le sexe, les lieux de vie, etc.) mais elles sont souvent partielles.

Le projet vise à construire un tableau de bord recensant différents indicateurs et caractéristiques des personnes atteintes de MICI, ainsi que leur prise en charge. Ces données permettront à l'Afa d'en savoir davantage sur les patients concernés, d'apprécier l'évolution de la pathologie dans le temps et de conduire différents types d'actions, dont de la prévention. Ce projet s'inscrit dans une démarche open science, avec pour objectif la

diffusion d'une information accessible à tous. Les codes utilisés pour l'estimation des indicateurs et les éléments de chiffrage obtenus seront en accès libre et pourront compléter la documentation ouverte relative à la base principale du SNDS.

L'apport du HDH a consisté à construire ce tableau de bord. Des experts ont notamment utilisé l'algorithme de repérage de la CNAM pour identifier les patients atteints de MICI, dont la description est disponible en ligne dans la cartographie des pathologies et des dépenses. Des indicateurs sur les patients, notamment leurs caractéristiques démographiques (âge, sexe, localisation), leurs traitements et leurs prises en charge hospitalières ont été calculés et restitués. Des échanges entre l'association, des cliniciens et les experts du HDH ont permis d'affiner ces indicateurs.

Le premier tableau de bord a été restitué à l'Afa. Il comprend notamment une déclinaison du nombre de patients pris en charge pour des MICI en 2019, par sexe, classe d'âge, répartition géographique ; du nombre d'hospitalisations longues et de jours ; du volume des principales chirurgies digestives et de la part des personnes ayant reçu au moins un acte en lien avec une pose de stomie digestive ; du nombre moyen de jours d'arrêt dans l'année (même s'il n'est pas possible de relier ces arrêts à leur cause) ; et du recours aux principales classes de médicaments. Un support pédagogique comprenant les clés de lecture a été fourni. Il permettra de faciliter la réutilisation et la valorisation de ces éléments. Il est convenu de mettre ces indicateurs à jour tous les trois ans. Ces données viennent compléter des indicateurs déjà recueillis dans le cadre de l'observatoire de l'Afa et issus principalement de la CNAM.

Accompagner les personnes concernées dans l'exercice des droits

Un nouvel outil est en cours de développement par le HDH, pour faciliter l'exercice des droits des citoyens sur leurs données de santé. Il associe **un formulaire numérique pour exercer ses droits** à un portail pédagogique d'information sur les données de santé et la recherche. Cet outil innovant vient répondre aux besoins d'information des acteurs de l'écosystème et de la société civile sur les données de santé. Ainsi, en 2023, grâce notamment aux retours d'un groupe de travail, une première **version du portail patient et du formulaire national de gestion des droits** permettant un exercice des droits facilité pour les personnes concernées ont été lancées.

Insights patients pour co-construire un portail d'information

En 2022, d'après le baromètre de connaissances sur les données de santé du HDH, la moitié seulement des Français déclarait être au fait de leurs droits concernant leurs données personnelles, et seulement un sur cinq les connaissait en détail. De plus, 77 % des personnes interrogées estimaient être insuffisamment informées sur les projets de recherche utilisant les données de santé, et une majorité d'entre elles exprimait le désir d'en apprendre davantage notamment sur l'utilisation des données pour la recherche. Afin de répondre à cette demande, **deux comités ont été constitués**. Le comité de pilotage composé de représentants de FAS, de la DREES, de la CNIL, et du HDH, et le comité citoyen composé de représentants d'usagers du système de santé et d'associations d'étudiants se réunissent depuis janvier 2023.

Les comités se sont entendus sur l'importance de **construire une information pédagogique** en mettant particulièrement l'accent sur le fait que les données de santé peuvent servir à la recherche. C'est en sensibilisant le public de cette façon que les droits seront compris et que les citoyens pourront les faire valoir s'ils le souhaitent. Parmi les autres questions considérées comme prioritaires par les comités qui se sont réunis récemment se trouvent : quelles règles doivent être respectées pour l'utilisation de ces données ? Quelles données sont concernées par cet outil et quels droits y sont associés ? Quels projets sont actuellement en cours ? Toutes ces questions seront abordées de façon accessible et pédagogique dans **un portail d'information citoyen**. Son objectif est de fournir toutes les réponses nécessaires pour prendre des décisions éclairées lors de l'exercice des droits ou simplement lors de la recherche d'informations. Il vise à éviter que les choix relatifs à l'exercice des droits ne se fassent par défaut ou par manque d'information.

L'utilisateur de ce portail d'information pourra être dirigé vers un espace dédié pour faire valoir ses droits. Il s'agira d'un formulaire permettant de faire une demande d'exercice des droits relatifs aux données personnelles de santé de la part d'un citoyen. Ce formulaire sera envoyé au DPO de la base concernée. Une procédure technique sécurisée permettra de faire valoir les droits tels que demandés dans le formulaire. Un email sera envoyé au demandeur pour le prévenir que cela a été réalisé.

3

PARTIE

BILAN DE L'ACCÈS AUX DONNÉES



Un accès réglementé

Il est nécessaire d'obtenir une autorisation auprès de la CNIL pour traiter des données personnelles de santé. En amont de cette autorisation, la procédure prévoit l'intervention de deux institutions : le HDH est le point d'entrée unique des projets mobilisant des données de santé, et le CESREES rend un avis sur la finalité, la pertinence et la qualité scientifique du projet.

Afin de faciliter l'accès aux données et pour que les porteurs de projets puissent réaliser leur traitement dans les meilleurs délais, le temps imparti à chacune de ces institutions est encadré. Le HDH dispose ainsi de sept jours pour transmettre le projet complet au CESREES. Ce dernier a ensuite un mois pour rendre son avis éclairé. Si l'avis est favorable, la CNIL dispose de deux mois, éventuellement reconductibles, pour émettre son autorisation. En cas de non-respect de ces délais, les avis ou décisions sont réputés tacitement favorables ou acquis.

En 2023, ce sont
303
 projets qui ont
 été déposés au
 guichet du HDH

Il existe des procédures d'accès dites simplifiées, dans le cadre desquelles l'autorisation de la CNIL n'est pas requise. C'est notamment le cas si le traitement envisagé est conforme à une méthodologie de référence (dite MR). Afin de promouvoir l'accès aux données et favoriser leur usage, le HDH contribue au déploiement de ce type de procédures, en étroite collaboration avec la CNIL. En 2023, il s'est ainsi prononcé sur **deux nouvelles méthodologies de référence** visant à faciliter l'accès aux données de la base principale du SNDS, l'une à destination des acteurs publics (MR-007), l'autre à destination des acteurs privés (MR-008) dès lors que les projets respectent l'ensemble des conditions fixées par les textes. Suite à la publication de ces référentiels en juillet 2023, le HDH a construit, en lien avec la CNIL et la CNAM, des outils opérationnels et pédagogiques afin de guider pas à pas les acteurs dans la mise en œuvre de leur projet.

3 PROCÉDURES SIMPLIFIÉES

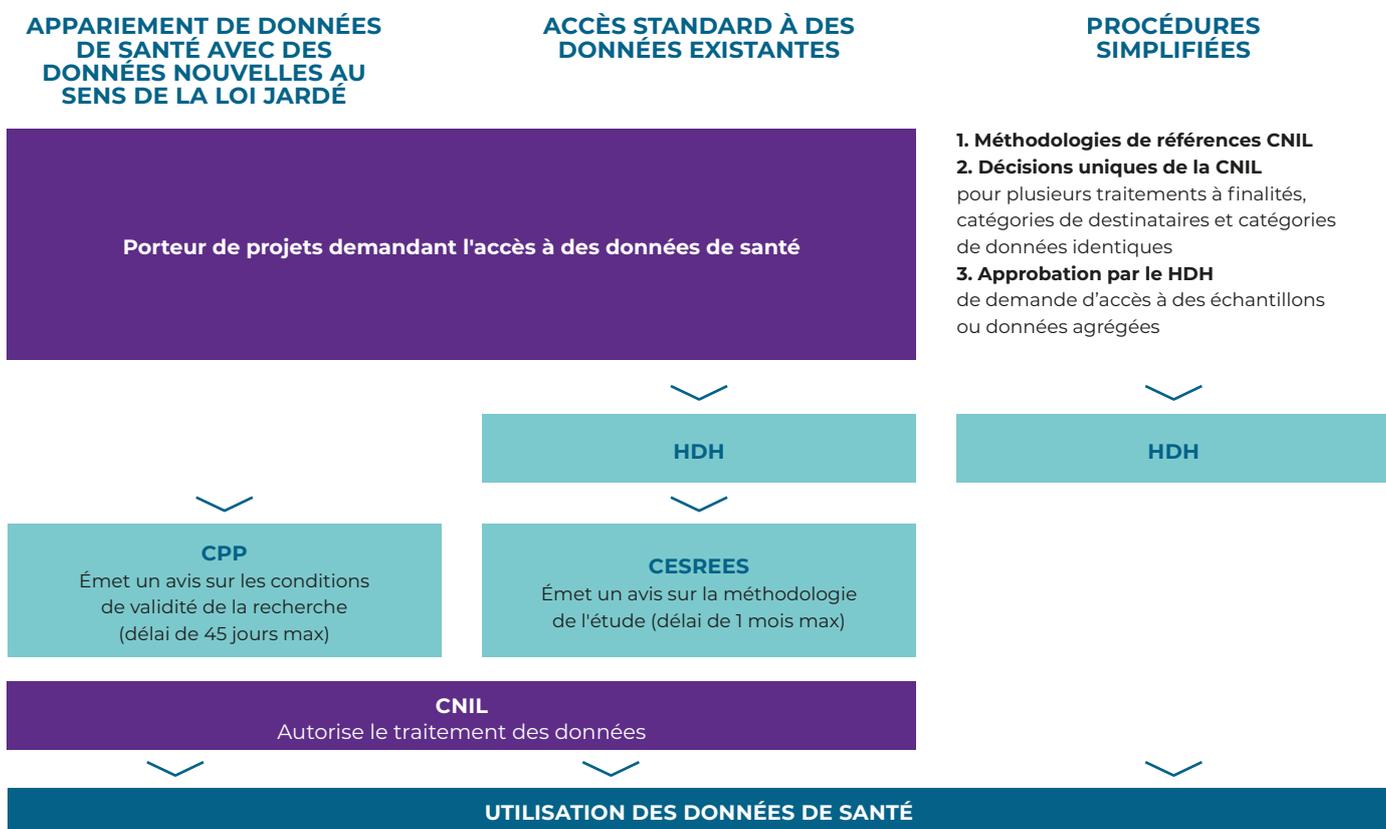
1 MÉTHODOLOGIES DE RÉFÉRENCE TELLES QUE MENTIONNÉES À L'ARTICLE 73 DE LA LIL

2 DÉCISIONS UNIQUES TELLES QUE MENTIONNÉES DANS L'ARTICLE 66 DE LA LIL

3 ÉCHANTILLONS ET DONNÉES AGRÉGÉES TELS QUE MENTIONNÉS AU II DE L'ARTICLE 66 DE LA LIL



SCHÉMATISATION DU CIRCUIT DES DEMANDES D'ACCÈS AUX DONNÉES DE SANTÉ



L'accès en quelques chiffres (chiffres 2023)

INDICATEUR	VALEUR
Nombre total de projets déposés au guichet en 2023	303
Nombre de projets déposés au guichet en vue d'une autorisation de la CNIL	296
dont transmis au CESREES	256
dont réorientés MR ou CPP	34
dont non complets	6
Nombre de demandes d'accès à l'ESND	7
Nombre de projets relevant d'une MR	1449
MR-004	1344
MR-005	32
MR-006	73

Une offre de services destinée aux porteurs de projets

Favoriser l'utilisation des données de santé nécessite de documenter ce qu'elles contiennent et les possibilités qu'elles offrent, rendre lisibles les procédures d'accès, et favoriser le partage des connaissances. C'est pourquoi le HDH élabore en permanence **des outils opérationnels et pédagogiques** destinés aux porteurs de projets. Ces outils les accompagnent à toutes les étapes de leur projet et sont améliorés chaque année. Ainsi, en 2023, le HDH a poursuivi l'enrichissement de son **catalogue de formations** en proposant quatre formations en e-learning dédiées à la base principale du SNDS. Une large documentation présentant les nouvelles méthodologies de références (MR-007 et MR-008), et des formations ont été constituées, dont une co-construite avec la CNIL.

Usage des données de santé et transparence

Le répertoire public des projets, prévu par la loi, s'est enrichi en 2023 de près de 2 000 projets supplémentaires. Au total, plus de 7 500 projets mobilisant des données personnelles de santé y étaient répertoriés fin 2023. Pour chacun d'eux, une fiche présente le bénéfice attendu, les données mobilisées, la méthode, les coordonnées du porteur de projet et de son DPO, ainsi qu'un calendrier de mise en œuvre. Plusieurs travaux ont été initiés en 2023, avec des applications concrètes en 2024 :

- **La refonte des formulaires** alimentant le répertoire public afin de faciliter la mise en œuvre des obligations de transparence incombant à tout porteur de projet mobilisant des données du SNDS élargi (mise à disposition de formulaires par démarche réglementaire, et un formulaire dédié au partage des résultats et méthodes issus de ces études).
- **La préparation de communications** rappelant les obligations de transparence associée à l'utilisation des données de santé .

Par ailleurs, dans une logique d'optimisation des informations recueillies par les différents acteurs intervenants dans le processus réglementaire, des travaux ont été menés en 2023 en lien avec la CNIL. Cette démarche de simplification s'inscrit dans le cadre de l'évolution prochaine du formulaire de demande d'autorisation de la Commission. Ces travaux se poursuivront en 2024 pour aboutir à une révision des documents composant l'actuel kit de démarrage afin de les simplifier et éviter les éventuelles incohérences en rationalisant les informations demandées. Cette réflexion associera le CESREES et la CNAM.



L'accès aisé et unifié, transparent et sécurisé,
aux données de santé pour améliorer la qualité
des soins et l'accompagnement des patients.

9 rue Georges Pitard, 75015 Paris

E-mail : hdh@health-data-hub.fr

Édition : Health Data Hub

Conception et réalisation : Parimage

Mai 2024

