

Rapport annuel 2024



Préface	3
Édito	4
Partie 1 : Le Health Data Hub, une infrastructure clé pour la recherche en santé	5
Les quatre grandes missions du Health Data Hub	6
La gouvernance et l'équipe du Health Data Hub	7
Les temps forts et chiffres clés 2024 du Health Data Hub	8
Partie 2 : Les avancées 2024 de la feuille de route pluriannuelle 2023-2025 du Health Data Hub	10
Enjeu 1 : Réduire les délais d'accès aux données de santé et multiplier les projets impactants	11
Simplifier et clarifier les procédures de création d'accès et de réutilisation des données	12
Mettre à disposition des données d'intérêt et de qualité	13
Soutenir les projets	15
Mutualiser les connaissances et les expertises	16
Renforcer l'attractivité du HDH	18
Enjeu 2 : Mettre à disposition les données de la base principale, l'enrichir et faciliter sa réutilisation	19
Développer l'expertise du HDH dans la mise à disposition des données	20
Faciliter l'exploitation de la base principale	21
Contribuer à enrichir la base principale du SNDS	23
Maintenir la plateforme à l'état de l'art fonctionnellement et en matière de sécurité	23
Enjeu 3 : Renforcer les connexions du Health Data Hub aux acteurs de l'écosystème	25
Renforcer le travail partenarial avec les acteurs locaux	26
Renforcer la présence du HDH au niveau européen et international	28
Réunir les conditions d'un partage efficace en termes de standards, de gouvernance d'accès et de valorisation dans un contexte national et européen	31
Réunir les conditions d'un partage efficace en termes de normes de sécurité, et de souveraineté dans un contexte national et européen	33
Enjeu 4 : Être à l'écoute de la société civile et co-construire une culture de la donnée de santé	35
Informer largement sur les enjeux de la réutilisation des données de santé	36
Écouter et impliquer la société civile	37
Développer les partenariats avec les associations d'utilisateurs	38
Partie 3 : Bilan de l'accès aux données	39
Un accès réglementé	40
Trois procédures simplifiées	41
Schématisation du circuit des demandes d'accès aux données de santé	41
L'accès en quelques chiffres	42
Une offre de service destinée aux porteurs de projets	43
Usage des données de santé et transparence	43

Sommaire



Préface

Gérard Raymond

Vice-Président du Health Data Hub

Félicitons-nous d'avoir en France un patrimoine collectif aux atouts considérables dans le domaine de l'utilisation secondaire des données de santé, source de nombreux usages et de bénéfices, notamment sur la surveillance épidémiologique des grandes pathologies, sur la mesure de l'efficacité des soins et des innovations, sur la conception des politiques publiques efficaces et protectrices.

Les données de santé, et les données de manière plus générale, peuvent aussi être mises au service d'une grande partie des composantes de la prévention. En termes de prévention primaire, l'utilisation de la donnée permet, entre autres, de développer les connaissances et de mieux caractériser les facteurs de risque. En termes de prévention secondaire, l'utilisation de la donnée contribue à mieux cibler les populations à risques, par exemple en croisant les facteurs de risques et le niveau de recours effectif aux soins. Elle permet aussi d'améliorer la performance du dépistage à l'aide d'algorithmes d'intelligence artificielle notamment. En termes de prévention tertiaire, enfin, l'utilisation de la donnée permet d'améliorer les parcours de prises en charge en analysant l'historique des patients sur des pathologies de longue durée.

Ces bénéfices ne sont possibles que parce que nous disposons d'un patrimoine collectif qui s'est consolidé progressivement : la création du Sniiram il y a plus de vingt ans, la constitution des grandes cohortes et, plus récemment, des entrepôts hospitaliers. Ce sont autant de briques posées au fil du temps pour construire un édifice auquel nombre d'acteurs participent et dont les patients sont les premiers bénéficiaires. Le Health Data Hub est entré dans la dynamique en 2019, et 5 ans plus tard, nous pouvons conclure qu'il a finalisé la construction de fondations solides et nécessaires au plein épanouissement de l'usage des données de santé pour la recherche. Ce rapport annuel 2024 en livre la démonstration.

Puisque les fondations sont fiables et solides, concentrons-nous désormais sur demain. Les promesses d'une santé numérique bénéfique pour la santé des personnes et le système de soin ne sont plus de l'ordre de l'imaginaire. Mais nous ne construirons un système solidaire seulement si nous parvenons à conserver la dynamique et la force du collectif. Et pour maintenir ce cap, ma devise sera de faire de la seconde vie des données, une narration commune !



Édito

Stéphanie Combes

Directrice du Health Data Hub

2024 aura été une nouvelle année de consolidation pour le Health Data Hub. S'il reste encore des étapes pour atteindre notre ambition – permettre un accès simplifié, sécurisé et utile aux données de santé –, nous en avons franchi plusieurs nécessaires et importantes, tant dans notre organisation que dans notre rôle auprès de l'écosystème. Un des marqueurs les plus visibles est le nombre de projets accompagnés : en un an, ils ont augmenté de plus de 60 %, signe que le HDH s'impose comme un point de passage reconnu pour les acteurs porteurs de projets exigeants et innovants.

Sur la question essentielle des délais d'accès, plusieurs leviers ont été actionnés. L'accélération des extractions issues de la base principale du SNDS, le développement du concentrateur, ainsi que les travaux menés avec la CNIL et le Comité stratégique sur les référentiels simplifiés et les modalités d'accès (contrats type, test de redevance, doctrine de primo accès) ont permis d'avancer. Ces avancées ont été entérinées lors du Comité stratégique des données de santé réuni en février dernier.

L'année 2024 a aussi marqué un tournant en matière d'élargissement des sources disponibles, avec 12 nouvelles bases intégrées au catalogue du SNDS, accompagnées d'un important travail de documentation. Près de 40 projets s'appuient déjà sur ces données. Ce chantier s'intensifiera encore cette année, notamment grâce à l'appel à projets récemment lancé.

Des projets ambitieux ont aussi vu le jour en 2024 : un projet majeur en IA générative appliquée à la santé, réunissant une trentaine de partenaires, a été retenu et lancé. De même, un second projet européen structurant autour de l'intégration de l'IA dans les dispositifs médicaux a démarré en mars 2025.

Nous ne sommes pas encore à destination. Il reste à simplifier les règles, sécuriser le financement des producteurs, renforcer la gouvernance et stabiliser un équilibre entre efficacité opérationnelle et engagements politiques. La feuille de route interministérielle des données de santé, attendue cette année, devrait fixer ce cap. Elle s'inscrira dans les travaux en cours sur la mise en œuvre du règlement européen sur les données de santé, qui doit structurer durablement le paysage.

Merci à toutes celles et ceux – au sein de l'équipe comme dans l'écosystème – qui ont contribué à faire avancer les lignes. Continuons ensemble.

Partie 1

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 l'accélérer l'innovation en favorisant le partage d'expériences et de connaissances.

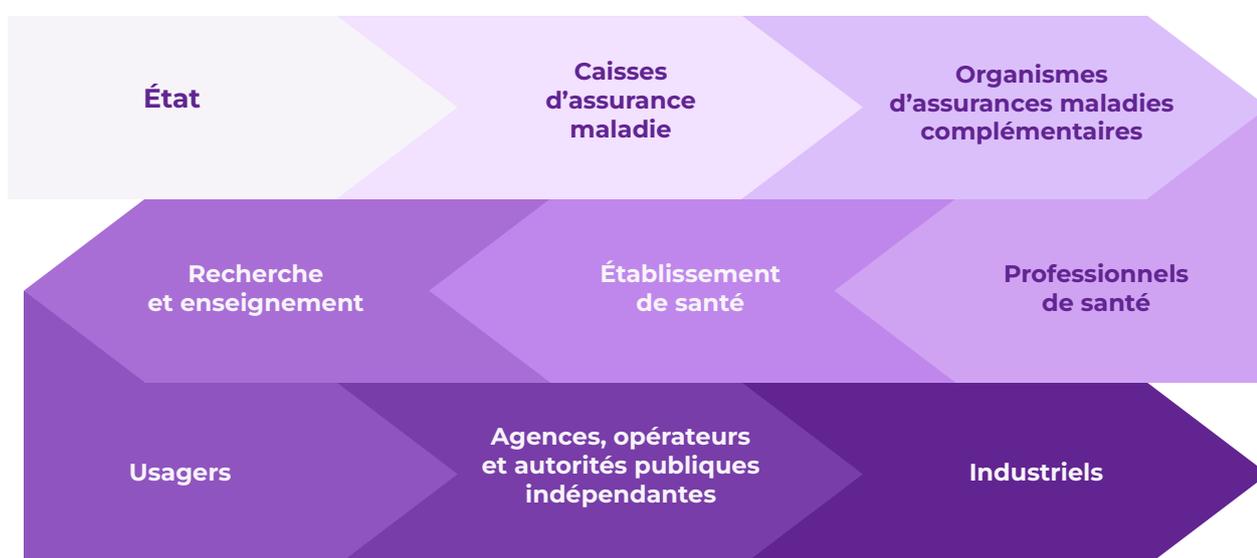
L'offre de service du HDH est complémentaire de celles que peuvent offrir les acteurs publics ou privés qui administrent les bases de données administratives, registres, cohortes, ou enquêtes. Elle répond aux besoins de deux types de projets :

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

De plus, le HDH assure le secrétariat du Comité stratégique des données de santé, et des groupes de travail qu'il porte. Ce comité, présidé par le ministre chargé de la santé, a pour mission de définir les orientations stratégiques, de coordonner les initiatives et de structurer les conditions de production, d'accès et d'usage des données de santé en France.

La gouvernance et l'équipe du Health Data Hub

Le HDH est un groupement d'intérêt public dont le financement est principalement public. Il a ainsi été doté de 25,61 millions d'euros pour l'exercice 2024, 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'assurances maladies 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, des Solidarités et des Familles ainsi que la Direction générale de la recherche et de l'innovation (DGRI), pour le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. 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.

Les temps forts et chiffres clés 2024 du Health Data Hub

**163 PROJETS
INNOVANTS DONT:**

92 projets
impliquant des établissements
de santé

60 projets faisant
appel à l'intelligence artificielle

44 projets
impliquant des industriels

**1 OFFRE DE SERVICE TECHNOLOGIQUE
EN CONSTANT DÉVELOPPEMENT:**

120 projets visant l'utilisation
de la plateforme technologique

1,2 Po de données ingérées
sur la plateforme technologique

33 canaux de transferts de données
vers des institutions partenaires

75 % de projets
plateforme croisant **2 à 7**
bases de données entre elles

303 extractions
de la base principale du SNDS
et **33** appariements directs
et indirects réalisés

1 CONTRIBUTION À L'ENRICHISSEMENT DU PATRIMOINE DE DONNÉES:

12 nouvelles bases de données
inscrites au catalogue du SNDS, pour un total
de **22** bases accessibles via le HDH

18 entrepôts de données de santé
accompagnés dont: **16** entrepôts hospitaliers, **1** entrepôt national
de médecine de ville et **1** entrepôt multicentrique clinique comprenant
des données de la base principale du SNDS

1 ACTEUR DE RÉFÉRENCE POUR LA PRÉFIGURATION DE L'ESPACE EUROPÉEN DES DONNÉES DE SANTÉ (EHDS):

6 participations à **6** consortia lauréats d'appels
à projet européen visant à construire l'EHDS (EHDS Pilot,
SHAIPED, French_HealthData, Quantum, TEF Health, Darwin)

19 bases de données de santé françaises
accompagnées via le projet européen "French_HealthData@EU"

5 cas d'usages et **3 livrables** permettant
de créer un pilote à l'EHDS suite à l'achèvement
du consortium HealthData@EU pilot

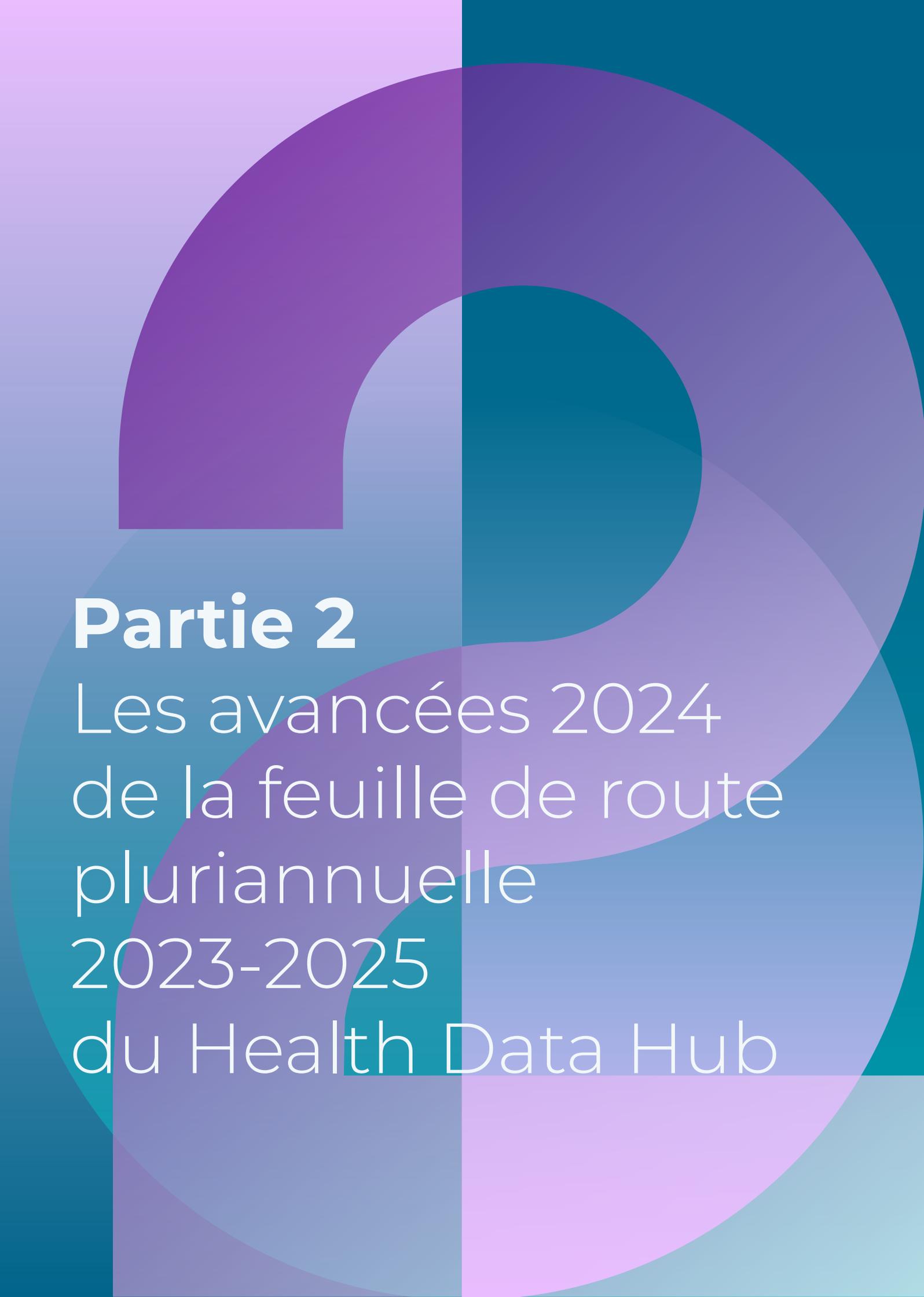
1 OUVERTURE À LA SOCIÉTÉ:

8 webinaires "les Mardis de la donnée de santé",
rassemblant plus de **2000** participants

14 formations et MOOC visionnés par **1584** utilisateurs

5 Data challenges réunissant
2050 participants

1886 fiches au répertoire public
des projets consultées par **91406** visiteurs



Partie 2

Les avancées 2024
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



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

Participation à la consultation de la CNIL sur les référentiels simplifiés

En 2024, le HDH a pris part à la consultation lancée par la CNIL sur les référentiels simplifiés, avec pour objectif de dresser un état des lieux des usages, d'identifier les évolutions souhaitables des textes actuels et de recenser les besoins en matière de nouveaux référentiels dans le champ de la santé.

Contribution au European Data Innovation Board

Le HDH a poursuivi en 2024 son engagement au sein du European Data Innovation Board (EDIB), qui accompagne la mise en œuvre des réglementations européennes telles que le **Data Governance Act** et le **Data Act**. Le HDH y apporte son expertise en matière de gouvernance des données de santé, avec un focus particulier sur l'interopérabilité.

Poursuite des travaux contractuels et tarifaires au sein du Comité stratégique des données de santé

Les travaux sur la tarification des données ont connu une étape importante avec le **lancement d'une phase pilote** visant à tester les grilles de redevances harmonisées élaborées en 2023. Cette expérimentation permet d'évaluer leur appropriation par les établissements et de recueillir des retours pour en améliorer l'efficacité. Un comité de suivi accompagne cette démarche, dont les enseignements seront consolidés en 2025. Dans le prolongement de ces travaux, un groupe de travail dédié à l'élaboration d'un contrat type pour la mise à disposition des données hospitalières a été lancé. Une première version, centrée sur les cas de fourniture de données, a été soumise à des représentants industriels et fait actuellement l'objet d'ajustements avant validation d'un modèle standardisé.

Implication du HDH dans les groupes de travail du Comité stratégique des données de santé

Le HDH est impliqué dans plusieurs groupes de travail du Comité stratégique des données de santé :

- **Le groupe "Catalogue"**, chargé d'identifier les bases à inscrire au catalogue du SNDS et à recommander leur intégration.
- **Le groupe "Financement"**, qui travaille à une tarification harmonisée et transparente pour faciliter l'accès aux données, en particulier pour les acteurs privés.
- **Le groupe "Contrats types"**, qui élabore un modèle national pour simplifier les négociations contractuelles autour des données hospitalières.
- **Le groupe "Standardisation et interopérabilité"**, qui définit les formats et terminologies communs au socle de données hospitalières.
- **Un groupe dédié aux formations**, visant à renforcer les compétences autour de l'usage de la base principale du SNDS.

Ces travaux s'articulent étroitement avec les discussions européennes, notamment dans le cadre de l'action conjointe **TEHDAS2** et de la préfiguration de l'Espace européen des données de santé (**EHDS**).



Mettre à disposition des données d'intérêt et de qualité

Avancées sur le catalogue du SNDS

Le HDH œuvre à la constitution du catalogue du SNDS, qui vise à regrouper une collection de bases de données de santé à fort potentiel scientifique, en complément de la base principale. Cette mission, conduite en lien étroit avec les responsables de données et encadrée par le Comité stratégique des données de santé (Costrat), devient progressivement opérationnelle grâce à la **publication en mai 2024 du référentiel de sécurité du SNDS** – condition préalable.

L'année 2024 a été ponctuée d'avancées sur la mise en œuvre du catalogue : signatures de contrats avec les partenaires, mise en place de canaux de transfert sécurisés, accompagnement des partenaires dans l'élaboration de leurs documents techniques et juridiques, etc.

En parallèle, le catalogue de métadonnées a été aligné au **standard européen HealthDCAT-AP**, dans le cadre du projet EHDS, avec une mise à disposition en open source prévue pour 2025. Le HDH contribue également au projet européen QUANTUM, visant à construire un label de qualité et d'utilité des bases de données de santé à l'échelle de l'Union européenne. Ces deux projets pourront permettre une meilleure identification des bases disponibles par les utilisateurs et de leur pertinence pour la recherche envisagée. Enfin, pour faciliter l'adhésion de nouveaux responsables de bases de données, un **kit documentaire a été finalisé en 2024** et sera publié début 2025. Il s'accompagne de principes clairs sur le partage de données : modalités de primo-accès, redevances, règles contractuelles... Le catalogue du SNDS s'affirme désormais comme un levier structurant pour la recherche, la surveillance et l'innovation en santé.

22

bases intégrées
ou en voie de l'être

38

projets mobilisant
les bases du catalogue

12

nouvelles bases
validées par la CNIL
en 2024

17

partenaires engagés
dans la première vague
de l'AMI catalogue

Zoom : 12 nouvelles bases intègrent le catalogue du SNDS

- 1 Base P4DP
(Collège National des Généralistes Enseignants)**
Platform for Data in Primary Care
- 2 Base Deepsarc
(Centre Léon Bérard)**
Base nationale du réseau de référence sarcomes - BCB Sarcomes
- 3 Base ARAC
(Malakoff Médéric)**
Base d'Analyse des Restes à Charge
- 4 Base PRECISION PREDICT
(Institut Curie)**
Comprendre l'hétérogénéité de réponse aux thérapies ciblées dans les cancers broncho-pulmonaires pour mieux la prédire
- 5 Base REALIGIST
(Institut Curie)**
Données de vie réelle des patients pris en charge pour une tumeur stromale gastro intestinale
- 6 Base ISIS
(Institut de cancérologie de l'Ouest)**
Infertilité Secondaire aux traitements du cancer du Sein
- 7 Base TARPON
(CHU de Bordeaux)**
Traitement Automatique des Résumés de Passages aux Urgences pour un Observatoire National
- 8 Base APSOREN
(CHU de Toulouse)**
Amélioration du Parcours de Soins du patient traumatisé crânien par le développement d'un modèle d'intelligence artificielle (IA) en Réseau de Neurone appliqué à des jeux de données massives
- 9 Base SEDAAR
(Hôpital Fondation A. de Rothschild)**
Screening for Eye Disease using Automated Analysis of Retinal photography
- 10 Base APRIORICS
(CHU de Toulouse)**
Apprentissage Profond Renforcé par l'Immunohistochimie pour la Requalification d'Images de Cancers du Sein
- 11 Base INTEGRA
(CHU de Lille)**
Signature du profil de risque de la mortalité dans une cohorte hospitalière des patients présentant des maladies métaboliques
- 12 Base UroCCR
(CHU de Bordeaux)**
Registre clinique biologique sur le cancer du rein



Soutenir les projets

Publication de guides pour accompagner les porteurs de projets

En 2024, le HDH a poursuivi son engagement en faveur d'un meilleur accompagnement des porteurs de projets, en renforçant son offre d'outils pédagogiques et de soutien réglementaire.

Un guide dédié à l'obligation de transparence dans l'utilisation des données de santé a été publié en français et en anglais. Cette publication a été complétée par une série de webinaires à destination des membres de l'AFCROs, du LEEM et du SNIITEM. Ces actions seront poursuivies et élargies à d'autres publics en 2025.

En parallèle, le HDH a adapté son accompagnement au **nouveau référentiel de sécurité du SNDS**, entré en vigueur en mai 2024. Ce changement majeur permet désormais aux porteurs de projets de bénéficier de l'homologation de la plateforme technologique du HDH, sans démarche propre à leur espace d'analyse. En revanche, la réalisation d'une Analyse d'Impact relative à la Protection des

Données (AIPD) reste obligatoire. Le HDH a ainsi renforcé son modèle d'AIPD et enrichi son programme d'ateliers pour accompagner les porteurs dans la construction de ce document, en insistant notamment sur l'information des personnes concernées et l'exercice de leurs droits. Toujours dans cette logique de simplification, des **outils pratiques** ont été publiés en lien avec les méthodologies de référence MR-007 et MR-008, qui permettent un accès simplifié à la base principale du SNDS sans autorisation spécifique de la CNIL. En collaboration avec la CNIL et la CNAM, le HDH a mis à disposition des **"starter kits"** pour guider pas à pas les porteurs dans la constitution de leur dossier.

Un **guide complémentaire** a été publié pour détailler les démarches à suivre en cas de modification d'un projet. Il permet aux porteurs d'identifier si les changements envisagés sont substantiels et nécessitent une nouvelle évaluation par le CESREES ou une autorisation de la CNIL, en tenant compte de leur impact sur les données ou les droits des personnes concernées.

En contrepartie de l'ouverture et de l'accès facilité aux données du SNDS, un ensemble de conditions doivent être respectées. Elles sont fixées par **le code de la santé publique**.

Concernant l'obligation de transparence, des actions sont réalisées à différents stades d'avancement du projet. D'abord, avant sa mise en œuvre, via la complétion d'un formulaire dédié qui a vocation à alimenter une fiche projet restituée dans **le répertoire public** des projets. Mais aussi, après sa mise en œuvre, via **la transmission des résultats** et de la méthodologie qui en sont issus, pour enrichir la fiche projet concernée. Tous les porteurs de projets ayant conduit des études mobilisant des données de santé depuis la loi du 26 janvier 2016 de modernisation de notre système de santé sont concernés. Quel que soit le cadre réglementaire dans lequel s'inscrivent ces projets (procédure simplifiée, demande d'autorisation, accès permanent, etc.), il demeure de leur responsabilité de s'acquitter de cette obligation. Cette démarche de transparence permet d'apporter aux citoyens des garanties sur la bonne utilisation de leurs données de santé.

Accompagnement technique des porteurs de projets

Un chantier d'amélioration des supports d'accompagnement a été lancé afin de mieux répondre aux besoins des utilisateurs de la plateforme technologique proposée par le HDH. Jusqu'alors, l'assistance était assurée de manière ponctuelle par les référents données, qui intervenaient pour reproduire les erreurs, proposer des solutions ou partager de bonnes pratiques sur les outils disponibles (Jupyter, RStudio, Spark). L'objectif est désormais de structurer cet appui et de le rendre plus accessible. Dans cette dynamique, un premier guide exécutable a été mis à disposition sur la plateforme. Il permet d'illustrer concrètement l'usage des outils, notamment pour l'accès aux données et l'exécution de traitements, facilitant ainsi la prise en main dès l'arrivée des utilisateurs.

De plus, en 2024, le HDH a continué de mettre à disposition son expertise en data ingénierie, data science, standardisation et analyse de la base principale du SNDS pour accompagner huit projets stratégiques portés par des institutions publiques. Enfin, en 2024, le HDH a lancé une nouvelle offre dédiée à l'ouverture de bases de données de santé anonymisées. Cet accompagnement couvre toutes les étapes : vérification du caractère anonyme, documentation, interopérabilité, choix de licence, et attribution d'un identifiant DOI pour garantir la citation des données partagées. Trois bases ont ainsi été publiées sur data.gouv.fr, dont deux bases synthétiques et une base issue d'un data challenge.

3
bases
de données
mises à disposition
en open data

Mutualiser les connaissances et les expertises

Poursuite du programme Data challenges en santé

En 2024, deux nouveaux Data Challenges ont été organisés. Le premier, **DigiLut, porté par l'hôpital Foch**, portait sur la détection automatique du rejet de greffe pulmonaire. Plus de 250 équipes internationales ont participé, produisant plus de 1 200 modèles testés. Les trois lauréats ont partagé une dotation de 25 000 €, et les modèles ont été mis à disposition en open source.

Le second, **Cytologia, mené par le Groupe Francophone d'Hématologie Cellulaire et Algoscope**, visait à améliorer le diagnostic hématologique par intelligence artificielle. Près de 75 000 images issues de 23 centres hospitaliers francophones ont été utilisées pour entraîner des modèles de classification. Ce projet a été sélectionné dans le cadre des défis "Convergence IA" lors du **Sommet pour l'action sur l'intelligence artificielle de février 2025**.

Deux Data Challenges sélectionnés lors d'appels à projets antérieurs seront lancés en 2025 : **Cardi-Hack**, sur la prédiction du pronostic des jeunes patients atteints de cardiomyopathie hypertrophique, et **Dat-Hub**, sur le diagnostic des syndromes parkinsoniens via l'imagerie Datscan.

Cinq nouveaux projets ont été par ailleurs retenus en 2024 dans le cadre de l'appel à projets "Data Challenges en santé", pour une ouverture des compétitions prévue en 2026. Ils portent sur des thématiques variées allant de la prédiction de l'insuffisance rénale aiguë à la classification de myocardites, en passant par l'analyse d'œuvres d'art brut dans une perspective de santé mentale, la détection de lésions orales ou encore la prédiction de la réponse à l'immunothérapie dans les cancers de la vessie.

Dans une logique d'ouverture de la science, les données utilisées dans les Data Challenges sont mises en **open data sur data.gouv.fr**. En 2024, la Société Française d'Allergologie a publié une base de données anonymisées issue du Data Challenge **Allergen Chip**, rassemblant plus de 4 000 bilans allergologiques et leurs données cliniques.

5 Data Challenges
organisés et **1 800**
participants dans
le monde depuis le
début de l'initiative

Participation à la feuille de route des Administrateurs Ministériels des Données, des Algorithmes et des Codes sources (AMDAC)

Le HDH a contribué à l'élaboration de la feuille de route de l'AMDAC du ministère chargé de la santé. Il a participé à la production de documents de référence sur l'open source et a vocation à continuer son implication en 2025 sur les sujets relatifs à la mise en place de procédures de pseudonymisation et d'anonymisation, la structuration des entrepôts de données de santé, et à la promotion de l'accessibilité des jeux de données prioritaires. Ces actions s'intègrent dans l'engagement que le HDH poursuit en matière de science ouverte. Parmi les actions phares de l'année 2024, la **mise en œuvre de la Bibliothèque Ouverte d'Algorithmes en Santé (BOAS)**, référentiel en ligne d'algorithmes en open source. En 2024, 23 ressources y étaient accessibles, dont 16 dédiées à la base principale du SNDS (extractions, ciblage, indicateurs) et sept autres liées à des bases tabulaires ou d'imagerie issues notamment

des Data Challenges. Chaque algorithme est accompagné d'une fiche descriptive et accessible via une plateforme ouverte telle que GitLab.

23 ressources référéncées dans la BOAS

Lancement du Programme de Formation sur les Données de Santé (PFDS)

Lauréat d'un appel à projets France 2030, le projet PFDS est porté par l'université Grenoble Alpes. L'objectif est de former les professionnels de santé à l'exploitation des données numériques via des jeux de données pédagogiques.

Le HDH contribue à plusieurs axes du programme :

- L'enrichissement d'un jeu de données synthétiques en médecine de ville à partir du projet P4DP.
- La création d'une version synthétique du socle de données commun des entrepôts hospitaliers.
- Le développement d'un générateur de données d'imagerie médicale pour les usages IA.

Ces ressources seront mutualisées à l'échelle nationale pour améliorer la formation en santé numérique.

Lancement de PARTAGES: IA générative en santé

Le projet "PARTAGES", issu d'une candidature du HDH et porté par un consortium d'une trentaine de partenaires incluant des laboratoires de recherches (du CNRS, de l'INRIA, ou de diverses universités), des établissements de santé, et des entreprises deep tech est l'un des lauréats de l'appel à projets France 2030 sur l'IA générative.

32 acteurs français mobilisés dans le cadre du projet PARTAGES

Le projet mobilise **32 acteurs de l'écosystème français, combinant expertise en IA et représentation diversifiée du secteur hospitalier.**

Parmi eux, des groupes hospitaliers tels que l'AP-HP et le GCS HOURAA, des centres de lutte contre le cancer comme le Centre Léon Bérard, l'Institut Curie, ainsi que les CHU de Toulouse, Bordeaux, Amiens-Picardie, Brest, Lille, Nancy, Rouen et la Réunion. Les groupes d'établissements ELSAN et Ramsay Santé ainsi que deux hôpitaux privés (Foch, Saint-Joseph et Marie Lannelongue) font également partie des acteurs impliqués dans ce projet.

Côté recherche, ce sont les laboratoires de recherche spécialisés en traitement automatique du langage (TAL), tels que LIA (Avignon), LIG (Grenoble), LIS (Aix-Marseille), LS2N (Nantes), BPH (Bordeaux) qui se sont ralliés au projet. Associés aux équipes de l'AP-HP et à la start-up Mistral AI, ils travaillent à la conception d'un modèle de langage adapté au domaine médical. L'évaluation du modèle et la fiabilisation des données seront assurées par le LIMICS, LISN, le département de santé numérique du CHU de Rouen (DéSaN) et la start-up reciTAL.

Le projet PARTAGES repose sur quatre étapes :

1. Le développement d'un LLM médical open source en français, destiné à servir de socle aux futures applications en santé.

- Le codage automatique des actes médicaux (DIM), facilitant la facturation hospitalière.

2. La création d'une base de comptes rendus médicaux fictifs annotés, permettant d'entraîner les modèles sur des cas d'usage spécifiques en toute simplicité et sans recourir à des démarches réglementaires strictes étant donné qu'aucune donnée personnelle n'est utilisée.

- Le résumé automatique des comptes rendus médicaux, permettant aux soignants de gagner du temps.

3. Le déploiement de sept modèles spécialisés, développés en collaboration avec GreyC, l'AP-HP, le LIMICS, le Centre Léon Bérard, l'UNESS et le centre de recherche BPH (Inserm/Université de Bordeaux), couvrant des cas d'usage stratégiques :

- La génération de cas cliniques, pour la formation des futurs professionnels de santé.

- L'identification de biomarqueurs, afin d'accélérer la recherche en oncologie.

- L'analyse de la réponse aux traitements, optimisant les stratégies thérapeutiques.

- La détection automatique d'infections, contribuant à la lutte contre l'antibiorésistance.

- La pseudonymisation des comptes rendus, pour favoriser leur réutilisation ultérieure pour la recherche.

4. La mise en place d'une plateforme nationale d'évaluation fédérée, développée par le CNRS, permettant de tester, comparer et valider les modèles.

L'ensemble des modèles et des outils développés seront mis en open source, afin d'être réutilisés et enrichis par l'ensemble de l'écosystème de la santé. En structurant un écosystème souverain, sécurisé et collaboratif, PARTAGES vise à accélérer l'innovation en IA générative dans la santé, au bénéfice des professionnels et des patients.

Renforcer l'attractivité du Health Data Hub

Sélection de onze lauréats dans le cadre d'appels à projets

En 2024, **cinq nouveaux projets** ont été sélectionnés dans le cadre de **l'appel à projets conjoint HDH - Green Data for Health**, lancé pour promouvoir la recherche croisant données de santé et données environnementales. Ces projets portent notamment sur les liens entre vagues de chaleur et hospitalisations, l'impact de l'exposome sur le développement des enfants, ou encore la prédiction des consultations pour asthme ou BPCO en Nouvelle-Aquitaine.

En complément, une formation citoyenne a été déployée avec l'INERIS, et des travaux menés au sein du CNIG ont permis de produire deux livrables de référence sur les possibilités de géolocalisation des données de santé et les besoins croisés en santé-environnement.

Lancé en partenariat avec **Roche et MEDIPATH**, un nouveau programme visant à valoriser les données industrielles a permis, quant à lui, de sélectionner **six projets** mobilisant ces données sur les cancers du sein et du poumon.

Deux projets exploitent la base MEDIPATH sur le cancer du poumon non à petites cellules pour améliorer le dépistage ou étudier les parcours patients. Deux autres

mobilisent les bases mises à disposition par Roche pour étudier les schémas thérapeutiques ou l'efficacité des traitements. Enfin, deux projets méthodologiques visent à générer des données de synthèse ou à standardiser les bases au format OMOP-CDM pour faciliter leur réutilisation.

En complément de ces appels à projet, une stratégie proactive a été menée en 2024 pour renforcer la visibilité du HDH auprès des chercheurs et acteurs publics comme privés. Près de 100 rendez-vous ont permis de présenter l'offre du HDH, donnant lieu à l'identification de 23 projets concrets.

Benchmark de structures nationales et internationales pour nourrir la réflexion sur la mise en place d'une fondation

Afin d'explorer la création d'un mécanisme de financement mixte, le HDH a étudié en 2024 plus de 45 structures françaises et internationales proches des données de santé. L'objectif : identifier des modèles inspirants permettant de mobiliser des financements privés ou hybrides au service de la recherche d'intérêt général. Une phase d'entretiens est prévue en 2025.

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

Ensemble, ils ont lancé un appel à projets conjoint **"La donnée pour la recherche et l'innovation en santé-environnement"**.

Cette initiative s'inscrit dans le cadre du 4^e Plan National Santé-Environnement (PNSE4) afin de répondre aux défis croissants liés à la santé publique et à l'environnement.

Les résultats des projets seront cruciaux pour éclairer l'action publique et déterminer les mesures à prendre face aux enjeux environnementaux et sanitaires.

La première vague de cet appel à projets a porté sur les thématiques du bruit, des bassins industriels et des perturbateurs endocriniens. Les projets mobilisés dans la seconde vague se structurent sur les thématiques relatives au changement climatique et à la santé, au cumul d'expositions environnementales, le tout dans une approche "Une seule santé" (One Health).

Actuellement, neuf projets d'excellence en recherche, reliant données de santé et facteurs environnementaux, bénéficient d'un accompagnement humain et technique de la part du HDH dans leur mise en œuvre pour une durée de 18 mois. Une nouvelle phase du programme est en cours de préparation.

163
projets
accompagnés par le HDH



Enjeu 2

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





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

Mise à disposition des données de la base principale du SNDS

En 2024, grâce à la dynamique de collaboration étroite, la CNAM a pu réaliser 154 projets, dans lesquels le HDH a contribué à environ 300 extractions de données, dont une vingtaine d'appariements directs. Également, depuis la signature de **la convention de cotraitance** en décembre 2023, le HDH bénéficie d'une autonomie accrue, rendue possible par le partage d'expertise de la CNAM sur les phases de préparation des données.

Cette **montée en compétence** couvre désormais les étapes de préparation de la donnée sur la plateforme SNDS de la CNAM et de conventionnement, pour un total de 24 projets accompagnés à fin 2024 (dont huit appariements indirects et six ciblage de population). Cette nouvelle organisation s'inscrit

dans un partenariat renforcé avec la CNAM, qui continue de piloter plusieurs chantiers d'optimisation des processus. Des ateliers communs ont été mis en place pour améliorer la maîtrise des délais d'accès aux données.

Ce travail conjoint permettra dès 2025 un suivi plus précis grâce à des indicateurs adaptés aux différentes typologies de projets, au bénéfice de l'ensemble de l'écosystème.

Ouverture d'une partie des services du concentrateur

En 2024, le concentrateur a été mis en production et une partie des services ont été ouverts pour **industrialiser les appariements de bases diverses à la base principale**, avec ou sans reconstitution du NIR, dans un contexte de hausse des demandes des utilisateurs pour ce type de croisement.

300
extractions
de données

Le SI concentrateur, une solution pour simplifier les appariements entre responsables de données.

En 2024, le HDH a poursuivi le déploiement de son système d'information "concentrateur", conçu pour **faciliter l'appariement de données entre plusieurs responsables de traitement** dans le respect des plus hauts standards de sécurité. Il a ainsi fait l'objet d'une homologation dédiée, en fin d'année. Accessible via un portail web dédié, cette solution permet aux responsables de données de suivre **l'état d'avancement des opérations d'appariement, de manière centralisée, sécurisée et transparente**. Elle intègre également une gestion simplifiée des projets multicentriques, notamment grâce à des mécanismes de dédoublement automatique.

Deux modalités d'appariement sont proposées :

- L'appariement direct via le NIR, pour les bases de données contenant un triplet NIR ouvrant droit, sexe et date de naissance.
- L'appariement avec reconstitution du NIR, à partir des informations d'état civil (nom, prénom, date de naissance, sexe).

La sécurité des données directement identifiantes est au cœur de l'architecture technique du concentrateur :

- Toutes les données sont chiffrées en transit et au repos,
- L'authentification à double facteur est obligatoire pour accéder à l'outil,
- Le concentrateur est strictement cloisonné par rapport à la plateforme technologique du HDH,
- Les données identifiantes ne peuvent jamais être consultées dans l'outil: elles sont uniquement transmises pour traitement pseudonymisé.

Ce système constitue un jalon important pour l'industrialisation des appariements et la simplification des démarches des porteurs de projets, tout en assurant un haut niveau de protection des données.

Faciliter l'exploitation de la base principale

Le partage de documentation et de programmes d'exploitation

En 2024, le HDH a poursuivi ses efforts pour faciliter l'utilisation de la base principale du SNDS, tant par le **développement de ressources documentaires que par l'enrichissement de la Bibliothèque Ouverte d'Algorithmes en Santé (BOAS)**. Dix nouveaux projets ont ainsi été retenus dans le cadre de l'Appel à Manifestation d'Intérêt ouvert en continu, couvrant des thématiques variées: oncologie, périnatalité, traumatologie, greffe, comportements de santé et inégalités sociales. Ces projets ont comme objet de produire et valider des algorithmes de ciblage robustes, mobilisables dans d'autres travaux de recherche.

En parallèle, la **documentation en ligne du SNDS** a été renforcée. Huit nouvelles fiches ont été publiées, fruit de collaborations avec la société HEVA, la DREES et des experts métier. Elles apportent des éclairages sur des sujets concrets comme la télémédecine, les médicaments en accès précoce, la valorisation des séjours MCO et HAD, ou encore la gestion des valeurs manquantes. Un outil de visualisation a également été mis à disposition pour faciliter l'exploration de la base. Ces ressources contribuent à rendre les données plus accessibles et opérationnelles pour les porteurs de projet.

A ce jour, la documentation collaborative du SNDS comprend :

- Plus de 70 fiches thématiques sur la base principale du SNDS, dont certaines intègrent des extraits de code réutilisables.
- Un guide d'initiation destiné aux nouveaux utilisateurs souhaitant manipuler ces données.
- De nombreuses ressources partagées par nos partenaires, telles que des supports de formation, des guides et diverses documentations.
- Des renvois vers des ressources externes, incluant des open data, des programmes et des données synthétiques.
- Des scripts de standardisation de la base principale du SNDS au format OMOP-CDM, accompagnés de leur documentation, ainsi qu'une page dédiée aux standards de données de santé.
- Cette documentation enregistre près de 10 000 visites par mois.

70
fiches
thématiques
sur la base
principale
du SNDS

10 000
visites
par mois sur la
documentation
collaborative
du SNDS

25
projets
accompagnés
dans le cadre
de l'appel
à manifestation
d'intérêt relatif
la Bibliothèque
Ouverte
d'Algorithmes
en Santé

La conception et la diffusion de formations sur la base principale du SNDS

Le HDH a poursuivi en 2024 la publication de **modules en ligne dédiés à la base principale du SNDS**. Après la mise en ligne de trois premiers MOOC (Massive Open Online Course) en 2023, un quatrième module a été publié cette année, consacré aux bases complémentaires du SNDS, notamment les causes médicales de décès (CépiDc), la cartographie des pathologies et les concepts clés à maîtriser. Une introduction aux bonnes pratiques de gestion des données y est également proposée. Pour élargir la diffusion de ces connaissances, **une version anglaise** de l'ensemble des modules est désormais disponible.

D'une manière générale, l'ouverture élargie de l'accès à la base principale du SNDS implique de mobiliser des compétences de plus en plus spécifiques, tant pour comprendre la structure des données que pour les exploiter de manière rigoureuse.

Pour structurer cette dynamique et proposer une stratégie en matière de formations, un **groupe de travail dédié** a été mis en place en juin 2024 au sein du Comité stratégique des données de santé. Réunissant plusieurs partenaires du HDH (dont la CNAM, l'Inserm, l'ATIH, l'AP-HP et le CNRS), ce groupe a travaillé à la définition d'une **grille de compétences** listant les notions clés à maîtriser. Il a également identifié les priorités en matière de formations devant venir compléter l'offre existante.

Dans ce cadre, un premier module a été développé en lien avec la CNAM et l'ATIH. Il présente les nomenclatures qui structurent la base principale du SNDS et a été mis à disposition dès 2024. D'autres modules, axés sur des cas d'usage concrets, viendront compléter cette offre dès 2025.

La mise au format OMOP des données pour des projets à dimension européenne et internationale

En 2024, les travaux de standardisation de la base principale du SNDS au format OMOP-CDM se sont poursuivis avec la mise à disposition, dans la BOAS, d'une **nouvelle version de scripts** permettant la conversion des données de la base principale SNDS au format OMOP v5.3 couvrant la période 2015-2021.

Ce travail a été réalisé dans le cadre du **projet Perséphone**, porté par le HDH et lauréat de l'appel à projets lancé par EHDEN.

Les MOOC
ont été consultés
près de
600 fois
en 2024

Relever les défis des projets transfrontaliers – L'exemple du projet Coagulopathie

Dans le cadre du **projet pilote européen pour l'Espace Européen des Données de Santé (EHDS)**, le HDH a participé à une étude coordonnée par l'Agence européenne des médicaments (EMA) sur l'histoire naturelle des troubles de la **coagulation chez les patients atteints de la COVID-19 ou vaccinés** pendant la période Omicron. Cette initiative, impliquant quatre pays européens (Croatie, Danemark, Finlande et France), visait à tester les conditions de faisabilité d'une étude pan-européenne reposant sur le format standard OMOP-CDM.

Le HDH a mobilisé les données de la base principale du SNDS, en lien avec la CNAM, pour extraire et échantillonner la population cible (10 millions de bénéficiaires).

Les données ont ensuite été transférées sur la plateforme technologique du HDH, ingérées, validées puis converties au format OMOP-CDM v5.4. Une attention particulière a été portée à l'alignement sémantique des vocabulaires (ex. CIM-10), indispensable pour garantir l'interopérabilité. Enfin, un environnement technique sécurisé, basé sur Spark et Hive, a été mis en place avec une allocation dynamique de ressources pour traiter efficacement ces volumes massifs.

Ce projet illustre la capacité du HDH à s'inscrire concrètement dans des coopérations européennes en santé, à la fois sur le plan technique, réglementaire et organisationnel.

Contribuer à enrichir la base principale du SNDS

Poursuite des travaux sur la base principale du SNDS

Plusieurs initiatives menées de manière continue visent à **faciliter l'utilisation de la base principale du SNDS**, en l'enrichissant par appariement avec d'autres sources de données. Ces dynamiques s'appuient sur des partenariats stratégiques (comme celui avec Green Data for Health sur la santé-environnement) et sur la mise en place progressive du système d'information "concentrateur".

Sous l'égide du Comité stratégique

des données de santé, un travail structuré a été engagé pour élargir le périmètre de la base principale. Après une première phase centrée sur **un socle commun de données issues des entrepôts hospitaliers**, les travaux se sont poursuivis en 2024 avec une ambition étendue : intégrer de nouvelles sources comme les données de santé scolaire, de santé au travail et de l'autonomie. Parmi les pistes explorées, l'intégration à venir des certificats de santé de l'enfant (PMI) dans le contexte de la dématérialisation en cours via le programme

numérique en santé porté par la DNS. En parallèle, des échanges ont été lancés avec les ministères et organismes concernés pour cartographier les systèmes existants et définir les modalités d'intégration: DEEP et DGESCO pour la santé scolaire, DGT pour la santé au travail.

L'objectif est de renforcer progressivement la couverture de la base principale, et ainsi **améliorer la capacité d'analyse et de pilotage** des politiques publiques de santé.

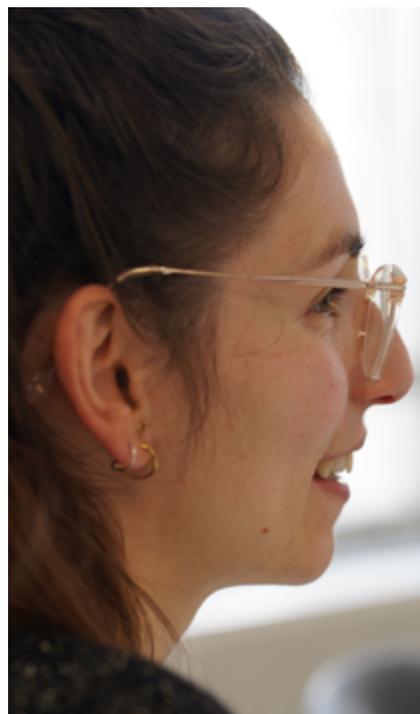
Maintenir la plateforme à l'état de l'art fonctionnellement et en matière de sécurité

L'offre technologique

Le HDH a poursuivi l'amélioration continue de sa plateforme pour répondre aux besoins croissants des utilisateurs. Pour permettre des analyses à grande échelle, **l'outil de calcul Spark** a été optimisé en 2024 : configuration par défaut améliorée, lancement des moteurs de calcul réduit à quatre minutes, prise en compte de la mémoire RAM dans la répartition des ressources et support multi-utilisateurs. Ces évolutions ont permis **d'industrialiser les traitements de données**, dont la réduction du temps d'exécution d'un traitement Python sur 70 000 fichiers DICOM, passé de 15 jours à 15 minutes. Par ailleurs, l'introduction de l'outil DBT-Spark a facilité la transformation de grands volumes de données au format OMOP.

En 2024, un **système simplifié** a été mis en place pour importer rapidement des paquets Python ou R, ainsi que des bibliothèques hébergées sur des dépôts Git, tout en intégrant des contrôles de sécurité. Cette nouvelle fonctionnalité permet aussi d'utiliser des **modèles pré-entraînés en intelligence artificielle** dans un environnement sécurisé, renforçant l'autonomie des utilisateurs et la rapidité des mises à disposition.

En 2024, le HDH a poursuivi **l'industrialisation des processus d'ingestion et de validation des données** sur sa plateforme technologique. Les validations automatisées via **l'outil Dataguard** ont été renforcées pour améliorer la performance des traitements, en particulier sur les variables sensibles. Le traitement distribué via Spark a été optimisé pour gérer plusieurs pro-



jets simultanément, avec des gains d'efficacité notables. Par ailleurs, les données sont désormais converties au format Parquet, facilitant le traitement de grands volumes et ouvrant la voie à une pseudonymisation automatisée, prévue pour 2025.

Des travaux ont également été engagés pour **améliorer la détection et la gestion des données sensibles** comme les dates ou lieux, en lien avec l'évolution de la doctrine du HDH. Ces efforts s'étendent aussi aux données non tabulaires, telles que l'imagerie médicale, qui présentent des défis techniques spécifiques.

Enfin, un **outil de suppression standardisée des identifiants FINESS** a été développé pour limiter les risques de réidentification dans les extractions de la base principale du SNDS. Ce traitement garantit la conformité aux règles encadrant l'usage des données de santé, notamment vis-à-vis des industriels de santé.

Implémentation des exigences eIDAS et mise à jour de l'homologation de la plateforme technologique au regard du nouveau référentiel de sécurité SNDS

À la suite de l'entrée en vigueur du nouveau référentiel de sécurité du SNDS, le HDH a procédé à la **ré-homologation de sa plateforme technologique**. Parmi les évolutions requises, il a mis en place un dispositif de vérification de l'identité des utilisateurs, conforme au référentiel de l'ANSSI sur les moyens d'identification électronique, lui-même aligné sur les exigences du règlement européen eIDAS.

Désormais, avant tout accès à la plateforme, chaque utilisateur doit suivre une procédure de vérification d'identité à distance, opérée par un prestataire certifié par l'ANSSI, mandaté par le HDH.

Le cadre de sécurité de la plateforme technologique du HDH a été conçu selon les principes de défense en profondeur, une approche visant à superposer plusieurs niveaux de protection afin de sécuriser les données face aux menaces potentielles.

Cette stratégie repose notamment sur le contrôle des accès, le chiffrement des données, la surveillance des activités et la gestion des incidents. Son élaboration s'est inscrite dans une démarche menée 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 chargé de la santé.

La sécurité de la plateforme technologique est, en outre, testée et revue périodiquement par le biais d'audits techniques menés par des prestataires spécialisés en sécurité des systèmes d'information indépendants qualifiés par l'ANSSI, dont quatre ont déjà été réalisés avec succès depuis sa mise en œuvre.

Sur cette plateforme sécurisée, le HDH met à disposition de chaque

porteur de projet un "espace d'analyse", qui lui permet notamment de recevoir, stocker, traiter et visualiser les données personnelles de santé nécessaires à ses études, sans toutefois pouvoir les exporter, de façon à garantir leur confidentialité.

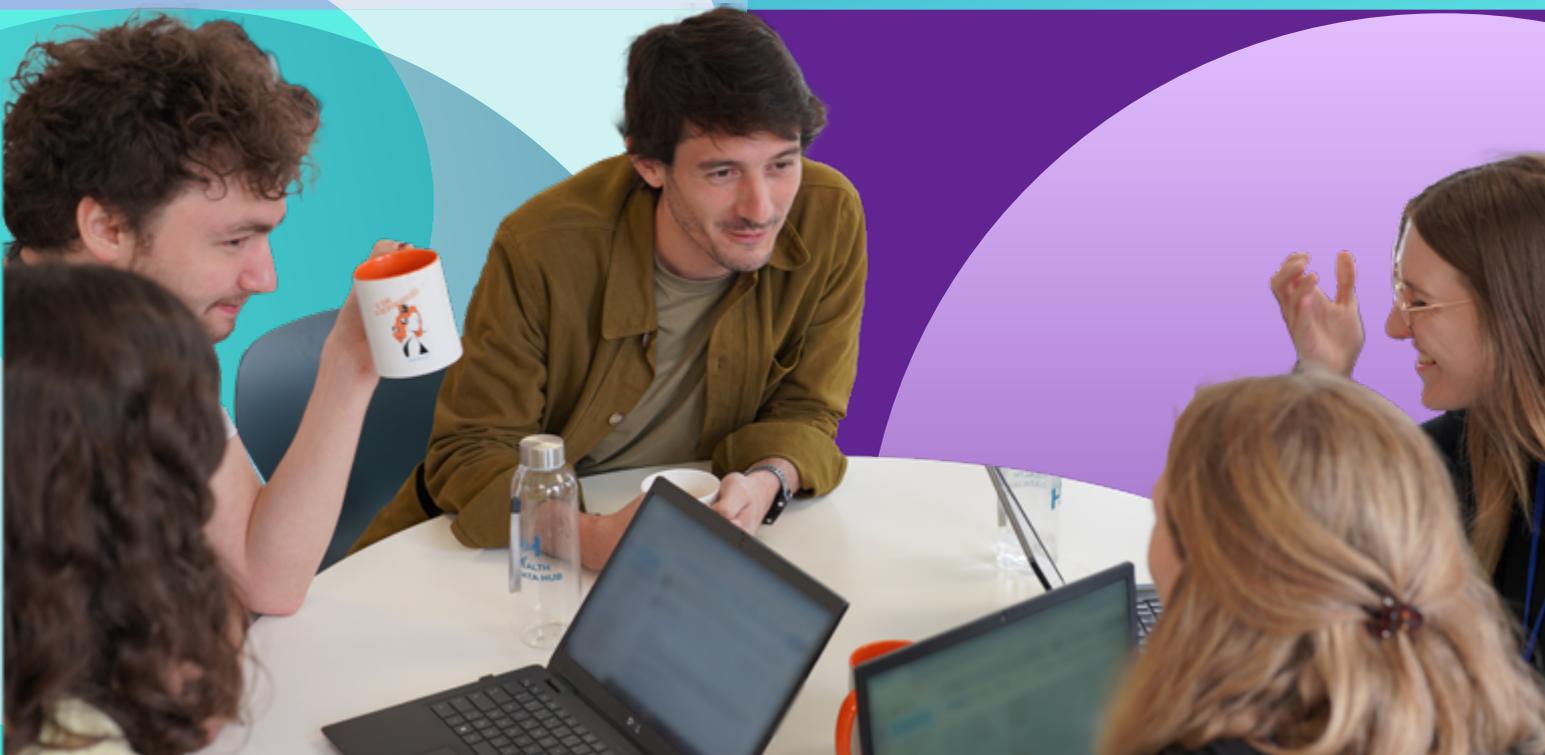
Les utilisateurs qui pourront traiter les données mises à disposition doivent suivre une procédure d'habilitation. Depuis le nouveau référentiel de sécurité du SNDS, l'identification et l'authentification des utilisateurs est aussi un prérequis avant qu'ils puissent accéder aux données du SNDS. Ces mesures doivent être mises en œuvre dans les conditions prévues par le règlement 910/2014 du 23 juillet 2014 sur l'identification électronique et les services de confiance pour les transactions électroniques au sein du marché intérieur (eIDAS).





Enjeu 3

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



Renforcer le travail partenarial avec les acteurs locaux

Lancement d'un appel à projets pour dynamiser l'utilisation des données d'entrepôts hospitaliers

Afin d'accélérer le déploiement des entrepôts de données de santé à l'échelle nationale, un appel à projets doté de 75 millions d'euros a été lancé dans le cadre du programme France 2030 pour **soutenir la création et le développement des EDS hospitaliers**.

En 2024, le HDH a poursuivi son accompagnement auprès des lauréats de l'appel à projets, notamment en organisant plusieurs **sessions de formation collective**. Ces formations ont couvert des thématiques variées telles que le cadre réglementaire présenté par la CNIL, l'usage d'un kit pédagogique facilitant la mise en place d'un EDS, ainsi que les bonnes pratiques de gestion de projets numériques en collaboration avec la Direction Interministérielle du Numérique (DINUM).

En parallèle, le HDH, en partenariat avec l'ANAP, a réalisé un **panorama des éditeurs de solutions logicielles pour les EDS**.

En complément de l'action stratégique menée dans le cadre de France 2030 pour soutenir la constitution et le déploiement des entrepôts au sein des établissements, le ministère chargé de la santé a souhaité lancer un appel à projets complémentaire, doté d'un financement de 3,5 millions d'euros, visant à soutenir l'utilisation de ces infrastructures.

La première vague de l'**appel à projets DATAE** (Données de santé et Applications), opéré par le HDH sous l'égide de la DGOS, s'est clôturée au deuxième trimestre 2024. Parmi les candidatures, 15 projets issus de neuf établissements de santé ont été sélectionnés. Ces projets couvrent des thématiques variées, telles que l'amélioration des diagnostics, l'optimisation des prises en charge ou la compréhension des maladies.

15 projets lauréats de l'appel à projets DATAE

Poursuite de la mise en oeuvre du projet d'entrepôt multicentrique EMC2

En 2021, l'EMA a retenu la candidature du HDH et de ses partenaires dans le cadre d'un **appel d'offres sur des données en vie réelle**. Ce dernier visait à faciliter l'accès à des bases de données individuelles hospitalières européennes, structurées à l'échelle du patient et conformes au format OMOP-CDM. Pour répondre à cette demande, 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 Paris Saint-Joseph, ainsi qu'à la plateforme de pharmaco-épidémiologie de Bordeaux.

Après l'autorisation CNIL de l'entrepôt EMC2 en 2023, l'année 2024 a été consacrée à l'élaboration d'une **stratégie de standardisation des données** (notamment via l'alignement des terminologies) et au cadrage des modalités contractuelles et de gouvernance entre partenaires. Les travaux se poursuivront en 2025.

D'autres projets multicentriques ont également progressé :

- Le **programme UNIBASE**, mené avec Unicancer, se poursuit avec l'accompagnement de huit projets en oncologie issus de deux vagues de sélection.
- Le **projet French PArADIsE**, une étude multicentrique coordonnée par le CHRU de Nancy et menée avec les HCL et les CHU de Dijon et Besançon dont l'objectif est d'améliorer le diagnostic

des patients admis aux urgences pour dyspnée aiguë en exploitant dix ans de dossiers patients informatisés. Ce projet a obtenu son autorisation CNIL, posé ses bases juridiques et développé un kit d'outils pour la pseudonymisation et la gestion des données dans un contexte multicentrique.

Sélection de huit lauréats dans le cadre de l'appel à projets P4DP

Le projet Platform For Data in Primary Care (P4DP) vise à constituer un **entrepôt national de données de médecine générale**, à partir d'informations cliniques collectées en soins primaires. Porté par le Collège National des Généralistes Enseignants (CNGE), en partenariat avec le Collège de la Médecine Générale (CMG) et le HDH, ce projet marque une avancée majeure pour la recherche en santé à partir des données de ville.

Le 8 octobre 2024, la CNIL a autorisé la mise en place de l'entrepôt P4DP, permettant ainsi le démarrage de la collecte des données auprès de 2 000 cabinets médicaux répartis sur le territoire. Une **convention** a également été signée entre le CNGE et le HDH pour permettre l'inscription de la base au catalogue du SNDS.



En parallèle, pour démontrer le potentiel scientifique et clinique de ces données, un **appel à projets dédié à la médecine générale** a été lancé en juin 2024 à l'occasion de Campus Live organisé par PariSanté Campus. Huit projets lauréats ont été sélectionnés, portant sur des enjeux variés: suivi de prescriptions, infections, pathologies chroniques, recours aux soins ou encore usage de la téléconsultation.

Ces projets s'appuient sur les données P4DP, souvent croisées avec celles de la base principale du SNDS, et sont accompagnés par le HDH avec un financement de 100 000 € par projet et un soutien technique sur 24 mois, en lien avec le CNGE et le CMG.

Animation des liens avec les acteurs clés de l'écosystème

En 2024, des partenariats structurants ont été approfondis :

- En **santé-environnement**, cinq nouveaux projets ont été soutenus dans le cadre de l'AAP conjoint

avec Green Data for Health, et un webinaire citoyen a permis de sensibiliser le public à ces enjeux.

- En **génomique**, une réflexion a été engagée pour mieux articuler les travaux avec le Collecteur Analyseur de Données (CAD), tandis que plusieurs webinaires ont contribué à la diffusion des usages auprès des acteurs du Plan France Médecine Génomique 2025.
- Avec les **organismes de recherche** (CNRS, Inserm), des actions de sensibilisation et un appui à plusieurs PEPR ont été menés.

Le HDH a également intensifié ses échanges avec :

- Les acteurs de **l'intelligence artificielle**, à travers une future formation citoyenne, des partenariats avec les clusters IA et des expérimentations autour du traitement automatique du langage naturel notamment via le projet Partages.
- Les **professionnels de santé**, notamment à l'occasion d'une audition par l'Académie de médecine et de travaux avec les syndicats d'internes.

- Les **start-up et industriels**, via l'événement "Start-up Connect", co-organisé avec PariSanté Campus, une participation active aux travaux du CSF Industries et Technologies de Santé et le nouvel appel à manifestation d'intérêt visant à stimuler la valorisation des données industrielles.

- Les **agences sanitaires**, avec notamment le soutien à Santé Publique France dans le cadre du projet Orchidée.

Enfin, le HDH a poursuivi sa collaboration avec **PariSanté Campus**, notamment lors de l'événement annuel Campus Live, pour diffuser une culture partagée de la donnée de santé.

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

HealthData@EU Pilot

Prévu pour deux ans, ce projet a livré ses résultats fin 2024 :

- **Déploiement de la plateforme** dans plusieurs pays, dont la France.
- Développement du standard de métadonnées **Health DCAT-AP**.
- Création d'un **formulaire européen unique de demande d'accès**, en collaboration avec Findata (Finlande).
- Démonstration de faisabilité via **cinq cas d'usage** (IA, santé publique, génomique...).
- Événements européens réunissant les partenaires autour des livrables clés.

Le HDH a notamment collaboré avec l'Allemagne sur l'interopérabilité technique et a contribué aux discussions sur les standards de métadonnées et la gouvernance du futur EHDS.

L'objectif du projet est de **développer une première version de l'infrastructure transfrontalière HealthData@EU**, en préfiguration du cadre établi par le règlement sur l'Espace européen des données de santé, publié au Journal Officiel de l'Union européenne

le 5 mars 2025. Cette plateforme permettra l'échange sécurisé d'informations entre les différents nœuds de l'infrastructure que constituent les plateformes de données nationales, les infrastructures de recherche européennes et les agences européennes. Afin de tester concrètement les solutions développées dans le cadre de ce projet, cinq cas d'usage ont été déployés, couvrant des thématiques variées et impliquant des acteurs européens de premier plan :

- Surveillance de la résistance antimicrobienne – pilotée par le Centre européen de prévention et de contrôle des maladies (ECDC).
- Identification des troubles de la coagulation liés à la Covid-19 – menée par l'Agence européenne des médicaments (EMA).
- Analyse des tests, hospitalisations et vaccinations des populations vulnérables – portée par Sciensano.
- Anticipation des parcours de soins en cardiométabolique par intelligence artificielle – développée par l'Université d'Helsinki et l'Université de Bordeaux.
- Identification des signatures génomiques des cancers colorectaux – coordonnée par ELIXIR.

Lancé en octobre 2022 sous la coordination du HDH, le **projet HealthData@EU Pilot** rassemble 17 acteurs européens majeurs, dont les plateformes nationales de données de santé de plusieurs États membres (Finlande, Danemark, Allemagne, Belgique, Hongrie, Croatie et Espagne) ainsi que de la Norvège, deux agences européennes – l'Agence européenne des médicaments (EMA) et le Centre européen de prévention et de contrôle des maladies (ECDC) – ainsi que des infrastructures de recherche européennes telles que BBMRI-ERIC, Elixir et EBRAINS.

Il préfigure l'Espace européen des données de santé (EHDS), en posant les bases d'une infrastructure technique sécurisée pour l'échange de données de santé entre pays.

17
acteurs européens
engagés
dans le projet
HealthData@EU
Pilot

TEHDAS2

Le HDH participe activement à TEHDAS2, action conjointe européenne, destinée à **formuler les recommandations** pour la mise en œuvre de l'EHDS. En 2024, elle a contribué à :

- Une étude comparative sur les politiques tarifaires d'accès dans 20 pays.
- Des travaux sur la sécurisation des traitements de données (pseudonymisation, anonymisation, environnements sécurisés).

Les livrables seront intégrés à la consultation publique prévue à l'automne 2025.

French_HealthData@EU

Lancé en janvier 2024 pour trois ans et doté d'un financement de 4,6 millions d'euros, le projet French_HealthData@EU mené par le HDH vise à **renforcer la position de la France dans le futur EHDS**. Il mobilise un consortium de près de 20 partenaires, sous la coordination du HDH, avec deux axes prioritaires qui pourront déboucher cette année :

- Déploiement du catalogue de métadonnées chez les partenaires selon les standards européens.
- Amélioration de la qualité des données dans les bases françaises.

Projet QUANTUM

Dans le cadre du projet européen QUANTUM également lancé en janvier 2024 pour trois ans par la Commission Européenne, le HDH contribue à la **définition d'un label européen de qualité et d'utilité des données de santé**, outil clé du futur EHDS. En 2024, le HDH a co-rédigé les premières spécifications de ce label, aux côtés de ses huit partenaires français et de 35 institutions européennes.

Présidence d'un groupe technique au sein de la Community of Practice

Le HDH assure la présidence du groupe de travail "cross border gateways" de la Community of Practice, **forum européen d'échange de bonnes pratiques** pour les futurs acteurs de l'EHDS. Il y pilote les travaux techniques liés à la connexion des plateformes nationales à la plateforme centrale de la Commission européenne.

Projet SHAIPEP

Le HDH coordonne le projet européen SHAIPEP, dédié à **l'usage de l'intelligence artificielle dans le développement de dispositifs médicaux**, en s'appuyant sur les organismes d'accès aux données. Ce consortium de 29 partenaires issus de 11 pays a été sélectionné en juillet 2024 et a démarré officiellement en mars 2025. Trois cas d'usage sont prévus (maladies chroniques rénales, oncologie, cardiologie). Ce projet est entièrement financé par la Commission européenne à hauteur de 4 millions d'euros.

Les partenaires français sont le ministère chargé de la santé, la Haute Autorité de Santé, les centres de lutte contre le cancer Léon Bérard et l'Institut Curie, l'Inserm, EIT Health France, Implicity, Therapanacea, Therapixel et le HDH.

DARWIN EU®

En 2023, le HDH a rejoint le réseau DARWIN EU®, piloté par l'Agence européenne des médicaments (EMA), pour faciliter l'accès aux données de la base principale du SNDS dans le cadre de la **surveillance post-AMM**.

Les premières demandes CNIL ont été déposées fin 2024: deux études portent sur la prévalence et l'incidence de pathologies, ainsi que sur l'usage des médicaments en France (autorisations CNIL obtenues) ; deux autres explorent la caractérisation des patients avec une condition médicale incidente ou initiant un traitement, avec un examen en cours par le CESREES.

Ces études font l'objet de décisions uniques : une fois l'accès accordé, l'EMA pourra réutiliser les données pour mener au fil de l'eau les traitements lui permettant de répondre à ses missions tout au long du cycle de vie des médicaments concernés. En revanche, l'analyse des données dans ce contexte requiert un important effort préalable de standardisation des données pour les mettre au format international OMOP, compatible avec les scripts utilisés dans le réseau.

TEF-Health

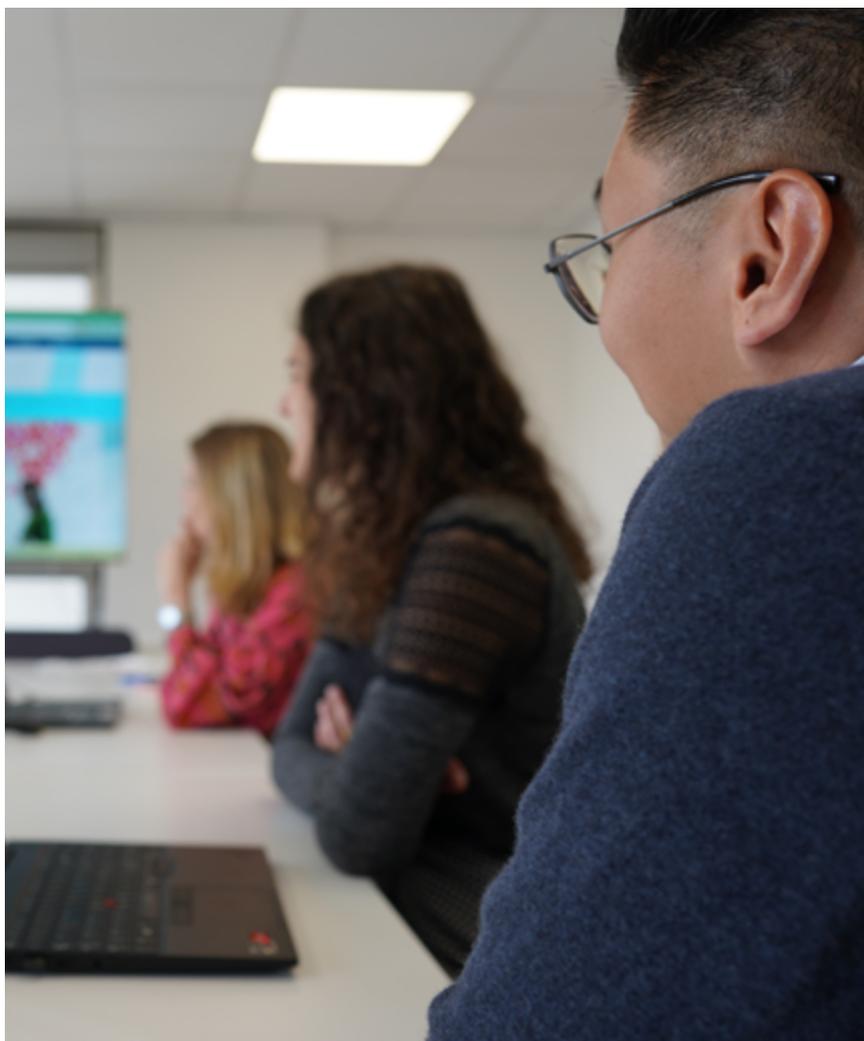
Le HDH poursuit son implication dans le projet TEF-Health, qui vise à développer une **infrastructure de tests pour les dispositifs médicaux intégrant de l'IA**. En 2024, plusieurs échanges ont été amorcés avec des start-up françaises en vue de cas d'usage concrets.

Projet Orchidée

Le HDH coordonne un volet du projet européen Orchidée, porté par Santé publique France, qui vise à produire des **indicateurs de surveillance à partir de données hospitalières**. En 2024, ses actions ont porté sur la cartographie des capacités, la standardisation des données et le cadrage réglementaire.

35
partenaires
européens
contribuent
au projet
QUANTUM

29
partenaires
issus de 11 pays
européens
réunis autour
du projet SHAIPEd



Lancé en octobre 2024, le **projet Orchidée**, mené par Santé publique France et regroupant 28 partenaires nationaux, a été sélectionné dans le cadre d'un appel à projets de la Commission européenne pour améliorer et renforcer les systèmes de surveillance à l'échelle nationale. L'objectif est de constituer un réseau d'établissements de santé produisant en routine des indicateurs de surveillance sur des thématiques clés de santé publique en réutilisant les données hospitalières existantes.

En tant que leader de l'un des work packages du projet, le HDH est chargé de :

- Recueillir et évaluer les capacités et besoins des établissements partenaires.
- Assurer la mise à disposition de documents de spécifications et d'outils techniques pour permettre aux partenaires d'implémenter les indicateurs et de les transmettre à Santé publique France de manière homogène.
- Industrialiser et coordonner l'implémentation des indicateurs auprès des partenaires.
- Constituer et animer un Comité technique (COTECH) abordant les dimensions techniques et opérationnelles de l'implémentation des indicateurs.
- Réaliser le cadrage réglementaire du réseau Orchidée.

Au cours des premiers mois du projet, les efforts ont été axés sur plusieurs actions clés :

- Cartographie des capacités existantes : une enquête a été élaborée et diffusée auprès des établissements membres du consortium afin d'identifier leurs besoins pour les premières phases d'implémentation.
- Standardisation des données : le HDH a coordonné la définition d'une description standardisée des variables nécessaires au calcul des premiers indicateurs, avec un focus sur les infections respiratoires aiguës.
- Analyse réglementaire : une étude a été initiée pour déterminer le cadre juridique le plus adapté à la production des indicateurs du réseau Orchidée. Pour garantir l'adhésion des partenaires et aligner la démarche avec les pratiques des établissements, des discussions sont en cours avec le bureau des DPO des CHU.

En 2025, le projet entrera dans sa phase opérationnelle.

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

55
contributions
pour l'élaboration
de la "stratégie
interministérielle
pour construire
notre patrimoine
national
des données
de santé"

Contribution à l'élaboration de la "stratégie interministérielle pour construire notre patrimoine national des données de santé"

En janvier 2024, les travaux sur la "stratégie interministérielle pour construire notre patrimoine national des données de santé" ont été lancés, faisant suite à la remise au gouvernement du rapport [Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire des données de santé](#), piloté par Jérôme Marchand Arvier. Cette démarche vise à définir une **trajectoire commune, cohérente et ambitieuse** pour le développement des bases de données et l'usage secondaire des données de santé. Cette stratégie a également comme objectif d'accompagner les acteurs dans la préparation de l'entrée en vigueur du **règlement de l'EHDS**, dont les dispositions sur l'usage secondaire entreront en application en 2028.

Le HDH a participé à l'élaboration de cette stratégie en étroite collaboration avec les ministères chargés de la Santé, de la Recherche, de l'Économie, ainsi qu'avec la CNAM et les agences concernées. Le 30 septembre 2024, une proposition de stratégie articulée autour de quatre piliers a été présentée publi-

quement et soumise à consultation. Cette consultation a recueilli **55 contributions**, qui serviront de base aux travaux finaux pour aboutir à une stratégie interministérielle consolidée, dont la publication est prévue courant 2025.

Contribution et animation des travaux du Comité stratégique des données de santé

En 2024, le HDH a piloté ou contribué à quatre groupes de travail du Comité stratégique des données de santé, avec pour objectifs de clarifier les règles d'accès, accélérer les procédures et harmoniser les pratiques :

- **Catalogue du SNDS** : élaboration de principes d'accès privilégiés pour les contributeurs et définition d'une grille tarifaire harmonisée destinée aux projets portés par les acteurs privés. Deux notes de référence ont été produites : l'une sur le primo-accès, l'autre sur les principes de tarification.
- **Contrat type** : lancement d'un modèle de contrat de mise à disposition de données hospitalières, pour réduire les délais de négociation. Une première version a été soumise aux parties prenantes, dont un collège d'industriels.



- **Tarification** : phase pilote de tests en conditions réelles des grilles de redevances menée avec plusieurs partenaires. Elle a permis de consolider les grilles et les guides associés.

- **Standardisation du socle commun** : coordination de la convergence vers le modèle OMOP pour les entrepôts hospitaliers, sur cinq thématiques clés (identité, PMSI, biologie, médicaments, données cliniques), avec une réflexion engagée sur le format FHIR.

Le Comité stratégique des données de santé a également lancé de nouveaux travaux structurants avec les parties prenantes :

- Application du **référentiel de sécurité** dans le cadre de l'élargissement du SNDS,
- Renforcement de la **formation des acteurs** sur la base principale,
- Réflexion sur un **financement pérenne** des entrepôts de données de santé.

Des travaux menés à l'échelle nationale et européenne sur les enjeux de standardisation

Coordonné par le HDH, un **groupe de travail** a été lancé en janvier 2024 par le Comité stratégique des données de santé afin **de spécifier le format et la sémantique** permettant de représenter de manière standardisée les 51 items du socle commun aux entrepôts de données hospitaliers. Les membres du groupe se sont penchés sur les cinq grandes familles de variables du socle (données démographiques, PMSI, biologie, médicaments, examens cliniques) pour identifier les éléments de standardisation les plus pertinents. Les arbitrages et recommandations formulés ont été consignés dans une première série de livrables, qui pourront désormais être testés en conditions réelles dans les entrepôts de données hospitaliers.

En parallèle, et en lien avec les travaux de l'Agence du Numérique en Santé (ANS), le HDH a engagé des actions pour **aligner les terminologies médicales utilisées**, notamment dans le cadre de la standardisation de la base principale du SNDS et du déploiement de l'entrepôt EMC2. Ce travail, essentiel à l'interopérabilité des référentiels sémantiques et à la réutilisation efficace des données, s'est accompagné d'une réflexion sur une stratégie globale d'alignement. L'ensemble de ces travaux sera mis en open source en 2025.

Enfin, pour favoriser la montée en compétence de l'écosystème, le HDH a publié début 2025 **une série de fiches informatives sur dix standards de données de santé**, nationaux et internationaux. Réalisées avec le cabinet Veltys et des experts du domaine, ces fiches visent à mieux faire connaître les caractéristiques des principaux standards syntaxiques (structures de données) et sémantiques (terminologies de référence) utilisés dans le secteur.

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



Accompagner les acteurs concernés par la refonte du référentiel du SNDS

Historiquement limité à la seule base principale, le SNDS a été élargi par la loi relative à l'organisation et à la transformation du système de santé (OTSS) de 2019. Il inclut désormais **d'autres sources de données**, qui peuvent être hébergées sur des systèmes d'information décentralisés, répartis sur le territoire, et non plus uniquement centralisés.

Cet élargissement a conduit à la **publication d'une nouvelle version du référentiel de sécurité** le 6 mai 2024 qui a vocation à assurer une protection accrue des données du SNDS. Le périmètre du SNDS étant encore méconnu, la communication sur le référentiel de sécurité est essentielle pour assurer la diffusion de ces règles de sécurité. A ce titre, le HDH prépare, en collaboration avec le ministère chargé de la santé, un guide pédagogique qui permettra d'informer largement sur ce nouveau référentiel. Ce guide sera publié et mis à disposition de tous courant 2025.

Par ailleurs, sa mise en œuvre, encore récente pour de nombreux

acteurs, suscite des interrogations. Ainsi, un groupe de travail a été proposé sous l'égide du Comité stratégique des données de santé. Sa mission consiste à **recenser les points de blocages ou les incompréhensions liés au référentiel de sécurité** et, le cas échéant, à identifier les besoins d'accompagnement des acteurs concernés. Il est lancé en 2025.

Préparation à une nouvelle solution d'hébergement : anticipation, partenariats et stratégie

Dans le cadre de sa préparation à la migration vers une nouvelle solution d'hébergement cloud, le HDH a poursuivi en 2024 **ses travaux de veille et de coordination avec les acteurs publics et industriels**. Ces démarches s'inscrivent dans un contexte de transition stratégique, marqué par la loi du 21 mai 2024 sur la sécurisation et la régulation de l'espace numérique (SREN) et les avancées attendues sur les offres labellisées SecNumCloud. Le HDH a ainsi mené plusieurs échanges avec l'ANSSI, la DINUM et des prestataires potentiels, en vue **d'anticiper la disponibilité**

d'une offre nationale ou européenne conforme à ses exigences de sécurité.

Le HDH reste également mobilisé aux côtés de la DINUM, dans le cadre du programme de démonstrateurs cloud de confiance, en particulier sur le défi de sécurité dont il avait contribué à définir les contours dès 2023. Parallèlement, dans le prolongement du rapport piloté par Jérôme Marchand-Arvier publié en 2023, le HDH s'est vu confier **l'instruction d'une solution intercalaire visant à héberger une copie de la base principale du SNDS** sur une infrastructure provisoire. Cette solution, intégrée à la stratégie interministérielle de construction du patrimoine national des données de santé, doit renforcer l'autonomie du HDH et alléger la charge de la CNAM dans les opérations d'extraction.

Une procédure de marché public sera engagée en 2025 pour identifier le prestataire chargé d'opérer cette solution, avec pour objectif de garantir un hébergement sécurisé et une mise à disposition efficace des données pour les projets de recherche utilisant la base principale du SNDS.





Enjeu 4

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



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

Afin de permettre une large diffusion de l'information sur la seconde vie des données de santé, le HDH, via notamment sa direction citoyenne, a poursuivi en 2024 ses actions initiées précédemment. Il s'agit notamment de sa newsletter citoyenne des données de santé réalisée en partenariat avec France Assos Santé, et de son engagement dans le projet SN@SU, lauréat de l'appel à manifestation d'intérêt Compétences et Métiers d'avenir en Santé numérique.

Par ailleurs, ont été rédigés et publiés plusieurs articles, parmi lesquels :

- **Jacques Lucas, Caroline Guillot et Anne Schweighofer, 2024, Comment parler des données de santé aux citoyens ? Le défi de l'information créative et participative des structures de santé**, LEH, Innovation et communication en santé.
- **Elisa Régnier, Caroline Guillot, Stéphanie Combes, Intelligence artificielle, et données de santé pour la recherche : les défis en termes de sensibilisation et de formation**, Ecole et intelligence artificielle : je t'aime, moi non plus?, Administration et Éducation, n°183.
- **Stéphanie Combes, Caroline Guillot, Faire de la seconde vie des données un projet commun**, Think Tank Craps, 40 réflexions pour penser le digital et la prévention en santé.
- **Caroline Guillot, Améliorer les connaissances des personnes sur la seconde vie des données de santé et l'IA**, IA en santé, entre promesses et prudence. Tous pharmaciens, les Cahiers, n° 23.
- **Stéphanie Combes, Caroline Guillot, Citoyens et seconde vie des données de santé : les enjeux en matière d'information et de droits**, Entre gouvernance des données et intelligence artificielle : quelle place pour la poursuite de l'intérêt général ? Actes du colloque de clôture des travaux de la Chaire Justice sociale et IA.

En 2024, deux autres projets majeurs d'information ont été lancés avec France Assos Santé. Le premier a consisté à produire et mettre en ligne une **campagne de communication pédagogique** sur les réseaux sociaux, comprenant des vidéos et des supports d'apprentissage. Elle a permis de diffuser des messages clés et simples sur la seconde vie des données. Un second projet, visant également à **éveiller les consciences des citoyens à propos de l'utilisation des données de santé pour la recherche**, a porté sur le thème des bénéfices. Ce dernier constitue un enjeu fort, et un défi méthodologique et sociologique : il vise à explorer les nouvelles possibilités de produire des narrations sur les bénéfices liés à la réutilisation des données de santé. Ce travail se poursuivra en 2025.

Les Mardis de la donnée de santé

Le HDH a réalisé plusieurs programmes de formations citoyennes dès 2023, dont les modules sont accessibles et téléchargeables en ligne depuis son site internet. En 2024, cette démarche s'est poursuivie avec de nouveaux modules, tels que :

- Un programme sur les **entrepôts de données de santé**, constitué de cinq modules. En partenariat avec la CNIL, la DREES, FAS, et la participation de l'AP-HP.
- Un programme sur les **données de santé-environnement**, s'articulant en 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), France Assos Santé, et OpenData France.
- Un programme sur les **données génomiques** pour le soin et la recherche, en deux modules. En partenariat avec des experts de : SeqOIA, Plan France Médecine Génomique 2025, Collecteur Analyseur de Données, Auragen, et du Service de génétique médicale du CHU de Montpellier.

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.

Écouter et impliquer la société civile

Le HDH a poursuivi ses actions d'écoute, d'ouverture et de dialogue avec la société et en particulier à travers ces quatre initiatives de 2024 :

- **Organisation et animation d'un débat sur l'IA en santé, au CESE réunissant des lycéens et des membres de la CNIL, de la DNS, de la DGOS et du CNPEN.** Un événement en partenariat avec l'Espace éthique d'Ile de France et France Assos Santé. Cette action est lauréate des prix de la participation citoyenne, remise au Sénat.
- **Lancement de l'initiative des "Petits déjeuner du numérique en santé", avec des étudiants,** en partenariat avec PariSanté Campus ; et animation de sa première édition consacrée aux métiers de demain avec le numérique, la e-santé et l'IA, ainsi que les entrepôts de données de santé.
- **Organisation et animation d'un colloque ouvert principalement aux associations de patients sur l'utilisation des données de santé pour la recherche, avec un focus sur la prévention et le dépistage.** Un événement en partenariat avec la Fédération Française des Diabétiques, et l'Espace éthique Ile-de-France.
- **Lancement du groupe de travail "Adhésion patient" du Comité stratégique de filière (CSF) "industries et technologies de santé" 2023-2026,** visant à mieux sensibiliser la société civile, et renforcer le dialogue. Un groupe de travail co-animé par le HDH et l'association Seintinelles, avec la participation d'autres acteurs : des représentants du LEEM, d'Amgen, de la Fondation Arcad et Medtronic.

Un ensemble d'autres travaux ont été lancés et pour certains finalisés, dans le but d'identifier les cibles potentielles à couvrir pour dialoguer. Il s'agit entre autres de la réalisation de deux cartographies, l'une sur les "espaces/lieux citoyens", et l'autre sur les acteurs de la médiation numérique.



Des Français mieux sensibilisés aux réglementations et à l'usage des données de santé...

- **85 %** ont entendu parler du RGPD, **35 %** savent de quoi il s'agit avec précision.
- **83 %** savent ou se doutent que les données de santé sont soumises à une réglementation et à une sécurisation plus stricte que les autres types de données personnelles.
- **66 %** savent que leurs données anonymisées de santé peuvent être utilisées.
- **58 %** savent que leurs données pseudonymisées de santé peuvent être utilisées. Des français qui souhaitent approfondir leur connaissances sur l'utilisation secondaire des données de santé...
- **32 %** estiment être suffisamment informés sur les projets de recherche analysant les données de santé.
- **56 %** souhaitent être mieux informés sur l'utilisation de leurs données de santé. Ils préfèrent majoritairement obtenir ces informations via le site Ameli ou Mon Espace Santé.
- **53 %** affirment connaître leurs droits concernant leurs données de santé. Une grande partie de la population ignore encore les droits dont ils disposent, tels que le droit d'accès, d'opposition ou d'effacement des données.

Développer les partenariats avec les associations d'usagers

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'usagers 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 disproportionnée par rapport 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.

En 2024, cette expérimentation s'est transformée en service effectif, appelé **“Guichet Asso des requêtes”** (à la base principale du SNDS) en partenariat avec France Assos Santé. A ainsi été produite et mise en ligne **une documentation dédiée** comprenant notamment un livret pédagogique et une fiche de recueil du besoin. Des premières associations ont ainsi été impliquées dans ce nouveau service. Par exemple, le “groupe de travail : Assurance Maladie” de France Assos Santé a formulé une première requête. Elle vise à dénombrer le nombre de personnes en ALD qui, arrivant en fin de droits pour leurs indemnités journalières, basculent vers le régime d'invalidité.

AFM Téléthon - Dénombrer et décrire les patients atteints de maladies neuromusculaires bénéficiant de certains traitements

Les maladies neuromusculaires sont des pathologies du muscle ou de sa commande nerveuse. L'unité motrice (constituée par le muscle, la jonction nerf/muscle et le nerf) fonctionne mal et le muscle ne peut pas se contracter normalement. Le symptôme le plus courant est une faiblesse musculaire, qui peut varier en termes de moment d'apparition, d'étendue sur les muscles (un ou plusieurs groupes musculaires), de durée (temporaire ou permanente) et de progression (évolutive ou stable). Ces maladies peuvent causer d'autres symptômes : des problèmes cardiaques, des difficultés d'apprentissage, une atteinte de l'œil et/ou de l'audition, etc. Suivant leur nature, ces maladies se déclenchent à des âges différents de la vie, mais nombre d'entre elles sont infantiles.

Le projet vise à construire un **tableau de bord** identifiant les médicaments prescrits aux per-

sonnes atteintes de certaines maladies neuromusculaires. Celui-ci permettra à l'AFM de produire de la connaissance sur l'accès aux médicaments intervenant dans ces maladies. 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. Dans un premier temps, la preuve de concept a été développée sur trois médicaments. Pour chacun d'eux, le travail a consisté à identifier le nombre de patients sur la période 2018-2021 selon l'âge, le sexe, la répartition géographique et la comédication (ou change-

ment de traitement). Les codes de médicaments utilisés pour l'identification des traitements ont été récupérés grâce au serveur multi-terminologies de l'ANS. Cette approche est en cours de généralisation aux autres médicaments intervenant dans les traitements des pathologies d'intérêt.

La preuve de concept a permis d'estimer le nombre de patients pour trois traitements (Zolgensma, Evrysdi et Spinraza) suivant les caractéristiques démographiques de ces derniers, dans l'amyotrophie spinale. Les résultats peuvent être corroborés par l'expérience terrain de l'association et l'existence d'un registre de qualité dans cette maladie. Le travail entrepris permettra d'affiner la méthodologie et les éléments pertinents pour l'AFM Téléthon.



Partie 3

Bilan de l'accès
aux données

Un accès réglementé



Lorsque l'on traite des données personnelles de santé, une autorisation de la CNIL peut être requise. 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.

Par ailleurs, 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.

Dans cette dynamique, le HDH a consolidé, en 2024, son offre d'accompagnement pour les porteurs de projets en proposant de nouveaux outils pédagogiques. Dès le printemps, un [guide dédié à l'obligation de transparence](#) associée à l'utilisation secondaire des données de santé a été publié en français et en anglais. Cette publication a été complétée par une série de webinaires réunissant plus de 150 participants. Dans le prolongement de ces actions, le HDH met à jour et enrichit régulièrement [le répertoire public des projets](#), en intégrant les premiers résultats de ces derniers ainsi que les nouveaux projets mobilisant des données de santé.

Ainsi, 25 fiches ont été enrichies par des résultats au tout début de l'année 2025.

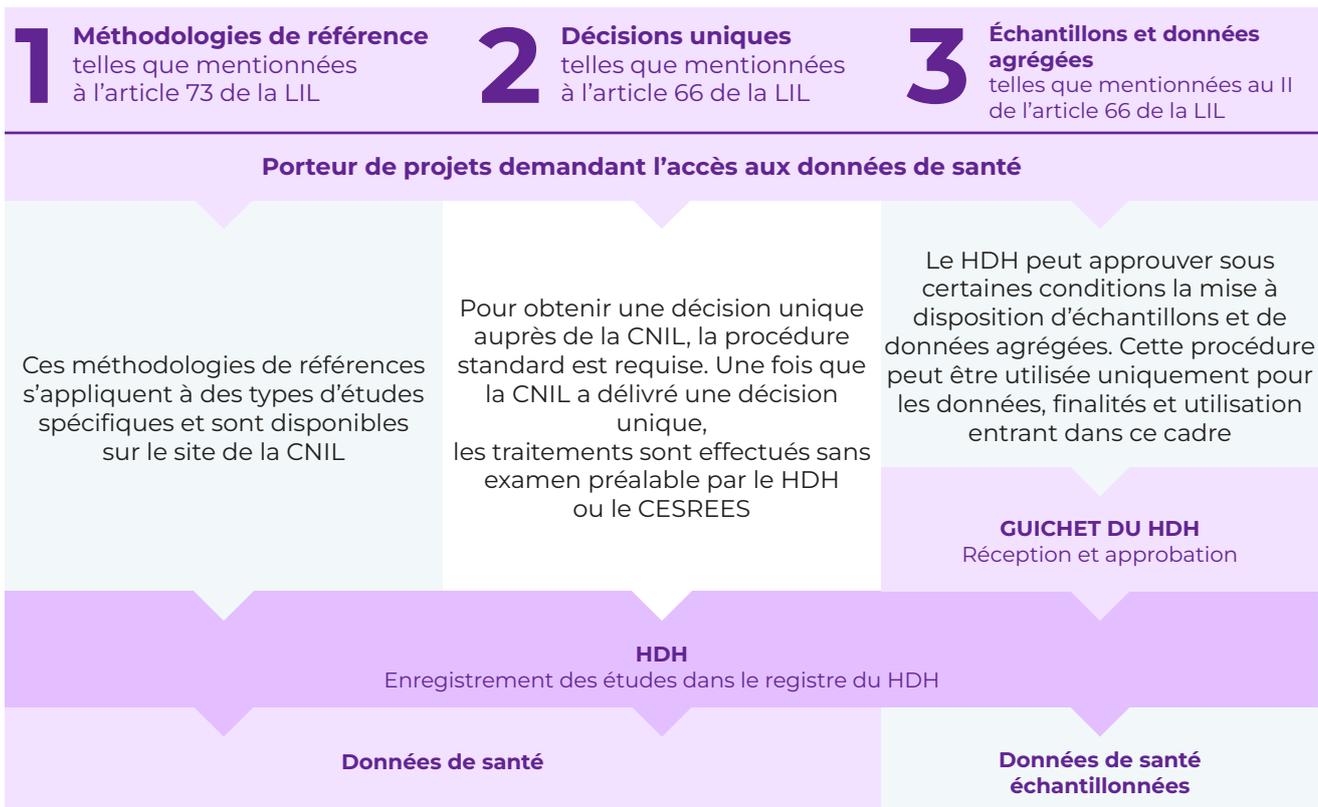
Dans une logique de simplification, des **outils pratiques** ont été développés autour des méthodologies de référence MR-007 et MR-008, facilitant l'accès à la base principale du SNDS après avis favorable du CESREES. En partenariat avec la CNIL et la CNAM, le HDH a publié des **"starter kits"**, guidant les porteurs dans la constitution de leur dossier. Trois webinaires ont permis de présenter ces référentiels et la documentation pédagogique associée à près de 850 participants.

Mi année, un [guide complémentaire](#) a été publié pour détailler les procédures en cas de modification d'un projet, permettant aux porteurs d'évaluer si ces évolutions nécessitent une nouvelle évaluation par le CESREES ou une autorisation de la CNIL. Enfin, un **guide d'aide à la rédaction des protocoles de recherche**, destiné à clarifier les exigences du CESREES en matière de méthodologie, d'objectifs et de sources de données, a été élaboré et paraîtra en 2025.

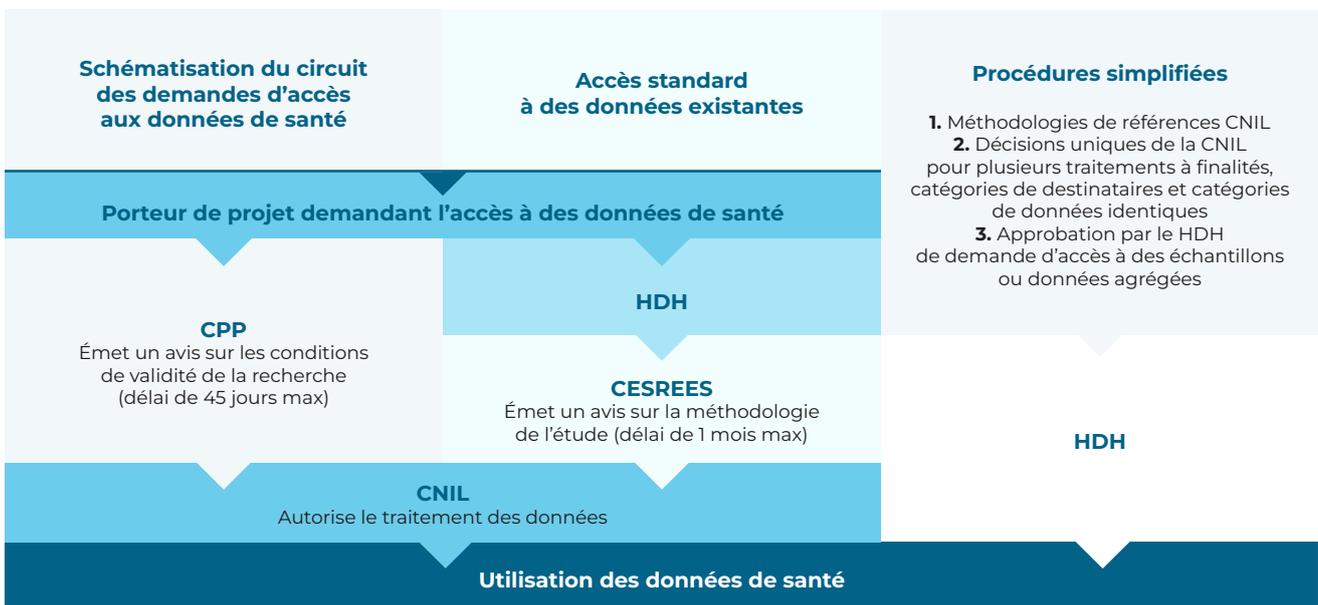
284 projets

ont été déposés au guichet
du HDH en vue
d'une autorisation CNIL dans le cadre
des MR-007 ou MR-008

Trois procédures simplifiées



Schématisation du circuit des demandes d'accès aux données de santé



L'accès en quelques chiffres

Indicateur	Valeur
Nombre total de projets déposés au guichet en 2024	302
Nombre de projets déposés au guichet en vue d'une autorisation de la CNIL	273
- dont transmis au CESREES	234
- dont réorientés MR ou CPP	18
- dont non complets	21
Nombre de projets déposés en vue d'une demande d'accès à l'ESND	18
Nombre de projets enregistrés relevant d'une MR (avec avis CESREES)	11
- dont MR-007	6
- dont MR-008	5
Nombre de projets enregistrés relevant d'une MR (hors avis CESREES)	1584
- dont MR-004	1491
- dont MR-005	17
- dont MR-006	65



Une offre de service 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 2024, le HDH a poursuivi l'enrichissement de son catalogue de formations en proposant deux formations co-construites avec la CNIL : l'une sur les modalités réglementaires d'accès à la base principale du SNDS, l'autre sur les procé-

dures d'accès simplifiées MR-007 et MR-008. Egalement, a été proposée une version anglaise des quatre modules du MOOC dédié à la base principale du SNDS.

Dans une logique d'enrichissement continue de son offre de services, d'autres formations viendront compléter ce catalogue en 2025.

Usage des données de santé et transparence

Le répertoire public des projets, prévu par la loi, s'est enrichi en 2024 de près de 2000 projets supplémentaires. Au total, plus de 10 000 projets mobilisant des données personnelles de santé y étaient répertoriés fin 2024. Pour chacun d'eux, une fiche présente le bénéfice attendu, les données mobili-

sées, la méthode, les coordonnées du porteur de projet et de son DPO, ainsi qu'un calendrier de mise en œuvre. Plusieurs travaux ont été menés en 2024 :

- La refonte des formulaires alimentant le répertoire public afin de faciliter la mise en œuvre des obli-

gations de transparence incombant à tout porteur de projet mobilisant des données du SNDS élargi.

- La collecte des premiers résultats des projets qui ont pu être publiés dans le répertoire public des projets début 2025.



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