



# RAPPORT ANNUEL 2021



HEALTH  
DATA HUB

# SOMMAIRE

Préface de Bruno Maquart	3
Édito de Stéphanie Combes	4

## 1

<b>LE HEALTH DATA HUB, UNE INFRASTRUCTURE CLÉ POUR LA RECHERCHE EN SANTÉ</b>	<b>5</b>
<b>QUATRE MISSIONS AU SERVICE DE L'INTÉRÊT GÉNÉRAL</b>	<b>6</b>
<b>L'ORGANISATION DU HEALTH DATA HUB</b>	<b>9</b>
<b>CINQ PRIORITÉS POUR 2021</b>	<b>11</b>

## 2

<b>RETOUR SUR LA FEUILLE DE ROUTE STRATÉGIQUE DU HEALTH DATA HUB</b>	<b>15</b>
<b>QUATRE ENJEUX INSCRITS DANS LA FEUILLE DE ROUTE STRATÉGIQUE 2019 - 2022</b>	<b>16</b>
<b>ENJEU 1 : DÉCLOISONNER LE PATRIMOINE DE DONNÉES DE SANTÉ ET AMÉLIORER SA QUALITÉ</b>	<b>17</b>
<b>ENJEU 2 : RENFORCER L'USAGE DES DONNÉES PAR UNE OFFRE DE SERVICE ATTRACTIVE</b>	<b>29</b>
<b>ENJEU 3 : POSITIONNER LA FRANCE COMME UN LEADER DANS L'USAGE DES DONNÉES DE SANTÉ</b>	<b>43</b>
<b>ENJEU 4 : GARANTIR LA PARTICIPATION DE LA SOCIÉTÉ CIVILE ET PROMOUVOIR LES USAGES DES DONNÉES DE SANTÉ</b>	<b>52</b>

## 3

<b>BILAN DE L'ACCÈS AUX DONNÉES</b>	<b>56</b>
<b>ACCÈS AUX DONNÉES ET PERSPECTIVES</b>	<b>57</b>



**Bruno Maquart,**  
Président  
du Health Data Hub

## PRÉFACE

Comprenant près d'une soixantaine de collaborateurs, avec l'objectif d'arriver à une centaine à la fin de 2022, l'équipe du Health Data Hub est aujourd'hui dimensionnée pour remplir les missions qui lui ont été confiées par la loi et pour conduire les projets attendus, en mobilisant des acteurs multiples. Le Hub peut s'enorgueillir de ses premiers succès.

En effet, les projets accompagnés par le Hub commencent à produire des résultats. On pourrait citer le projet Bacthub de l'Inserm et de l'APHP, visant à créer une base de données unique sur les infections bactériennes dans le sang, de même que le projet Nhance d'Inria et de l'AP-HP, qui repose sur une base d'images échographiques du ventre non nominatives, ou encore le projet Innerve, porté par Quantmetry, dont l'objectif est de développer un logiciel de diagnostic précoce des neuropathies des petites fibres. Certains d'entre eux ont déjà donné lieu à publication.

En deux ans, le Health Data Hub a ainsi trouvé sa place dans l'écosystème de la donnée de santé, au service de ceux qui assurent sa collecte comme de ceux qui en sont les utilisateurs. Quelques étapes restent à franchir, notamment pour mettre à la disposition de tous la base principale de l'Assurance Maladie, sachant que le Hub est co-responsable de son traitement depuis la publication d'un décret en juin dernier. C'est main dans la main avec la Caisse nationale d'assurance maladie que le Hub travaille depuis plusieurs mois, permettant à ses équipes de gagner en compétence. J'en remercie la Cnam, plus engagée que jamais à répondre, avec le Hub, aux attentes de la communauté.

Le Health Data Hub a récemment gagné ses galons au niveau international. L'Europe, désireuse de favoriser la circulation des données entre ses États membres, multiplie les initiatives. Le Health Data Hub, avec ses partenaires nationaux, contribue pleinement à cette dynamique : c'est ainsi qu'il a réuni un consortium, composé d'une quinzaine d'acteurs européens majeurs, dont les plateformes nationales de données de santé de plusieurs États membres, apte à répondre à l'appel à candidatures de la Commission européenne en vue de la constitution d'un espace européen des données de santé.

Enfin, le Health Data Hub est un projet résolument citoyen. Les données de santé suscitent d'importants débats, rendant indispensable la parole des citoyens. Il importe, parallèlement, de rappeler la nécessité qu'il y a de favoriser, dans le respect du droit des personnes, l'utilisation d'un patrimoine commun, permettant de relever le défi d'offrir à toutes et à tous la médecine la plus performante, tirant le meilleur parti des nouvelles technologies.

Bonne lecture à toutes et à tous.



**En deux ans,  
le Health Data Hub  
a trouvé sa place dans  
l'écosystème de la  
donnée de santé**





**Stéphanie Combes,**  
Directrice  
du Health Data Hub

## ÉDITO

Comment améliorer les dépistages et faire en sorte que les patients soient pris en charge le plus tôt possible ? Comment leur proposer les meilleurs traitements tout au long de leur vie ? Comment appuyer les professionnels de santé, dans un contexte clinique de plus en plus complexe ou pour faire face à une crise sanitaire ? **Grâce à l'intelligence artificielle, les données de santé ouvrent aujourd'hui la voie à de nouvelles pistes et des avancées majeures.**

Elles sont même devenues incontournables. Celles issues des implants cardiaques peuvent prévenir les insuffisances cardiaques ; celles provenant des mammographies peuvent accélérer le dépistage du cancer du sein ; celles retraçant les parcours de soin peuvent améliorer la prise en charge des maladies rares, etc. En France, nous avons la chance de disposer de bases de données extrêmement riches, toutes ces innovations sont donc à notre portée.

C'est pour relever ce défi que le Health Data Hub a été créé il y a maintenant deux ans : pour **pouvoir accéder de manière simple, unique, transparente et sécurisée, à un catalogue de bases de données de santé.** Grâce à sa plateforme technologique, les porteurs de projet d'intérêt public qui le souhaitent peuvent en effet mobiliser des sources de données volumineuses, les croiser et les traiter, en utilisant une importante puissance de calcul pour faire notamment tourner des algorithmes de recherche complexes. Le Health Data Hub est là pour soutenir leurs projets et les aider à trouver les solutions de demain afin d'améliorer la santé de tous les citoyens.



### **2021, une année de consolidation du Health Data Hub**



**Fin 2021, le Health Data Hub et ses 56 collaborateurs accompagnaient déjà 55 projets.** Au-delà de ces cas d'usage, il a multiplié les partenariats avec les acteurs clé de l'écosystème pour industrialiser le dispositif et soutenir l'innovation. De nouveaux appels à projets ont été lancés, des travaux stratégiques ont été entamés et une feuille de route scientifique se dessine. A la suite du décret officialisant l'extension du SNDS, le 30 juin dernier, le Health Data Hub et ses partenaires ont travaillé sans relâche pour réunir toutes les conditions nécessaires à la mise en œuvre du partage de la base principale et d'une première version du catalogue de données via la plateforme.

Résolument ouvert à la société civile comme à l'international, le Health Data Hub s'est engagé, avec ses homologues européens et les citoyens, dans les projets qui préfigurent le futur espace européen des données de santé. L'inauguration en décembre dernier de PariSanté Campus, dont le Health Data Hub est membre fondateur, témoigne de son engagement et de son ambition, au service de la souveraineté sanitaire de notre pays.





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



# QUATRE MISSIONS AU SERVICE DE L'INTÉRÊT GÉNÉRAL

Créé par la loi du 24 juillet 2019 relative à l'organisation et la transformation du système de santé, le HDH est une structure publique permettant aux acteurs d'accéder facilement à des données non nominatives de santé, hébergées sur une plateforme sécurisée qui respecte les droits des citoyens. 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 met en œuvre 4 missions.

**1 Offrir un guichet unique** pour accompagner les porteurs de projets dans leurs démarches administratives. Ces projets sont instruits par le Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (CESREES) qui émet son avis avant leur soumission auprès de la Cnil pour autorisation ;

**2 Ouvrir l'accès à un catalogue de bases de données partenaires.** Construit de manière progressive et itérative, le contenu de ce catalogue est porté par un arrêté pris après avis de la CNIL en cours d'instruction au moment de la publication de ce rapport. Les bases du catalogue seront répliquées dans la plateforme technologique afin de les croiser avec d'autres bases de données, en accord avec les acteurs à l'origine de ces données ;

**3 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 ;

**4 Animer l'écosystème pour accélérer l'innovation** en favorisant le partage d'expériences et de connaissances.

**L'offre de service du HDH** est complémentaire de celle que peuvent offrir les acteurs publics ou privés administrant des bases de données administratives, registres, cohortes, ou enquêtes. Elle répond à un besoin exprimé par deux catégories de projets :

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

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

## **Au service de l'intérêt général**

Les nouveaux outils comme l'intelligence artificielle ouvrent la voie à d'importants progrès dans l'accompagnement des patients, l'évaluation et le choix des traitements ou encore la gestion du système de santé. Et ce, grâce au traitement et au croisement d'un grand volume de données de qualité. En voici quelques exemples.

**Accélérer et améliorer la prise en charge des patients en ophtalmologie** grâce au projet [SEDAAR](#), porté par le Dr Rabih Hage MD, Ophtalmologue, Hôpital Fondation A. de Rothschild

On estime à 1,7 million de personnes vivant en France atteintes de handicaps visuels. Ces troubles sont essentiellement liés à des pathologies de la rétine. Le diagnostic est en grande majorité établi par les ophtalmologues grâce à divers examens basés sur l'imagerie. Il y a un enjeu à automatiser en partie ces diagnostics car d'une part, plus tôt le diagnostic est posé, plus on a de chances de traiter efficacement les pathologies, et d'autre part, les spécialistes se raréfient.

**Prévenir les effets indésirables graves pendant le parcours de soin** grâce au projet [HUGOSHARE](#), porté par le Pr Marc Cuggia, GCS Hôpitaux Universitaires Grand Ouest (HUGO)

Chaque année en France, le mauvais usage des médicaments est responsable de 10000 décès, principalement chez les plus de 65 ans. C'est en analysant les trajectoires thérapeutiques médicamenteuses complètes des patients (en ville et à l'hôpital) qu'il sera possible, à l'avenir, de prévenir les accidents liés à la prise ou à l'interruption de traitements, d'améliorer les parcours de soins des patients les plus fragiles et de développer un complément aux outils d'aide à la prescription pour les professionnels de santé pour un bon usage du médicament.



**Mieux connaître les déterminants du reste à charge pour les patients** grâce au projet [ARAC](#), porté par Malakoff Humanis

En France, les connaissances sur le reste à charge final des patients, après remboursement de l'Assurance maladie et des complémentaires santé sont encore limitées. Il existe en effet assez peu de données sur les écarts de ces restes à charge entre les personnes (selon le contexte de recours aux soins des personnes, de leur couverture assurance complémentaire, de leur situation professionnelle ou géographique, etc.). Grâce aux résultats issus des croisements de données de l'Assurance maladie et des assurés mutualistes, il sera possible de repérer les difficultés d'accès aux soins et de contribuer à la lutte contre les renoncements aux soins.

**Automatiser des activités des professionnels de santé pour libérer du temps médical** grâce au projet [TAMIS](#), porté par le Pr Yaneck Gottesman, Telecom Sudparis

Les prises de sang nécessitent, dans 10% à 15% des cas, une analyse complémentaire et manuelle réalisée par un hématologue afin de caractériser finement certaines cellules sanguines. En exploitant des images des frottis sanguins, il est possible d'entraîner des algorithmes qui détectent et caractérisent plus finement les différentes cellules d'intérêt. De cette manière, du temps médical sera libéré pour l'hématologue, au bénéfice du patient, et ce dernier obtiendra des résultats plus rapidement.

**Appuyer les professionnels de santé dans le diagnostic de pathologies** grâce au projet [DeepMap](#), porté par Damae Medical

Chaque année dans le monde, ce sont 10 millions de carcinomes basocellulaires qui sont diagnostiqués. Leur diagnostic, important pour éviter des interventions soit inutiles soit trop tardives, est possible par biopsie ou par des techniques d'imagerie. L'exploitation des images collectées par l'intelligence artificielle permettra de construire des outils qui accompagneront le dermatologue dans la pose du diagnostic, le guidage d'un acte chirurgical moins invasif et impactant pour le patient et le suivi de son traitement.



# L'ORGANISATION DU HEALTH DATA HUB

## L'équipe du Health Data Hub

**L'équipe du HDH est composée d'une soixantaine de personnes animées par l'envie commune d'aider les acteurs de la recherche et de l'innovation à mener à bien leurs projets.**

## La gouvernance

**Le HDH est un Groupement d'Intérêt Public** dont l'Assemblée générale réunit 56 membres répartis en 9 collèges.



Un représentant de chaque collège siège au Conseil d'Administration, sauf pour l'État qui compte deux représentants : la DREES pour le ministère des Solidarités et de la Santé et la DGRI pour le ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation. La CNAM, l'UNOCAM, l'INSERM, le CHU de Limoges, l'UNPS, France Asso Santé, Santé Publique France et le SNITEM en sont membres.

**L'Assemblée générale et le Conseil d'administration sont présidés par Bruno Maquart**, président d'Universcience. Gérard Raymond, président de France Asso Santé est vice-président du HDH, preuve de l'importance des associations de patients dans la gouvernance.

**En janvier 2021, le HDH s'est doté d'un conseil scientifique**, composé de 28 experts aux compétences complémentaires : médecins, pharmaciens, spécialistes de l'intelligence artificielle, juristes et spécialistes de l'éthique, généticiens, épidémiologistes, économistes de la santé, etc. Ce conseil est présidé par le Pr Alain Livartowski, oncologue, spécialiste des tumeurs thoraciques et adjoint au directeur des data de l'institut Curie. Il émet des avis visant à éclairer les décisions de la direction, dans un contexte international hautement compétitif. Il travaille en lien étroit avec la direction scientifique et la direction médicale du HDH. Le 26 novembre dernier, 7 membres internationaux sont venus rejoindre le conseil afin d'aider le HDH à identifier des partenariats hors de nos frontières ou de nouveaux outils et services. Ils contribuent à diffuser les savoirs de façon plus large au sein de la communauté scientifique mondiale et à intégrer le HDH dans le paysage européen des données de santé.

**Le financement du HDH est principalement public.** Sur la période 2019-2022, il se monte en moyenne à 20 millions d'euros par an. Il provient en grande partie du fond de transformation de l'action publique (FTAP) et de l'Objectif National des Dépenses de l'Assurance Maladie (ONDAM).

**Au 31 décembre 2021, le HDH comptait 56 collaborateurs** travaillant dans des équipes métier qui accompagnent les projets de bout en bout (guichet, fabrique, data, plateforme et produit), dans des directions stratégiques (scientifique, médicale et citoyenne) ou des fonctions support.

## CINQ PRIORITÉS POUR 2021

**En 2021, le Health Data Hub a poursuivi ses actions pour étendre ses relations avec les associations et les citoyens, développer ses partenariats avec des organismes divers, publics comme privés, ou encore augmenter son empreinte européenne et internationale.**

Ces actions ont été organisées autour de cinq priorités :

**Un accompagnement des projets de bout en bout.** Du montage des dossiers jusqu'à la mise à disposition des données, en passant par leur ciblage et la gestion des appariements, les équipes du Health Data Hub se sont mobilisées pour permettre à chaque projet d'aboutir, en lien étroit avec des partenaires tels que la CNAM et en cible la CNAV. Un kit de démarrage a été diffusé.

**Le Grand public au cœur de notre action.** À la recherche d'une relation de confiance et d'un dialogue ouvert avec la société civile, le HDH a privilégié en 2021 une information plus accessible, notamment pour ce qui concerne l'exercice des droits de chaque patient. Ses actions de communication ont été guidées par les besoins des citoyens, grâce à des groupes de travail, formés notamment de représentants d'usagers, et des consultations citoyennes, assurées par la direction citoyenne. Pour ce faire, une stratégie partenariale avec des associations de patients a été mise en place, comme pour l'étude VivreCovid19 réalisée par France Assos Santé en collaboration avec le HDH.

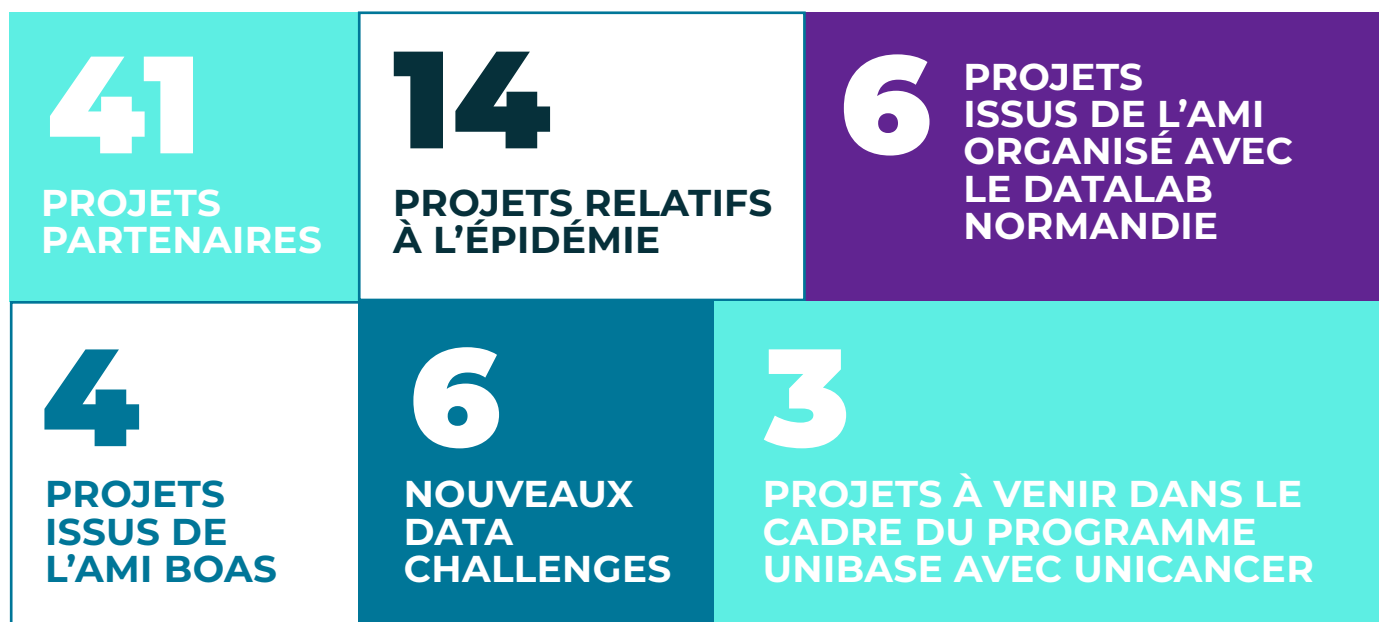
**Des partenariats stratégiques élargis.** Pour prendre toute sa place dans l'écosystème des données de santé, le HDH a noué des partenariats avec plusieurs grands acteurs de la santé, au service des patients. Pour mettre en œuvre un catalogue de plus large portée, une vingtaine de partenaires ont été identifiés parmi les acteurs de l'écosystème des registres, des bases nationales, des cohortes, des établissements de santé, des agences sanitaires, etc. Certains de ces partenariats permettront de préfigurer l'articulation entre le HDH et des initiatives locales ou thématiques. C'est notamment le cas avec la CNAM, l'INSERM, l'AP-HP, Unicancer, le Datalab Normandie ou le Groupement de coopération sanitaire des hôpitaux du Grand Ouest.

**Une infrastructure technologique innovante et sûre.** Développée depuis 2019, la plateforme technologique du Health Data Hub a été mise en service dès 2020. Cette plateforme constitue un défi technologique car elle doit répondre à d'importantes exigences de sécurité. Elle doit également proposer une offre logicielle permettant de répondre à des projets innovants parfois non réalisables à ce jour en France. C'est notamment dans ce cadre que les études de réversibilité se sont poursuivies, en collaboration avec nos partenaires comme la DINUM. La gouvernance de la cybersécurité implique la réalisation d'audits réguliers, le dernier ayant été conduit cet été, après un audit mené par l'ANSSI en février 2021.

**La structuration interne du HDH.** La structure du HDH se développe à grande vitesse pour répondre aux enjeux inscrits dans la feuille de route stratégique 2020-2022. En 2021, cette démarche a continué, notamment avec un renforcement important des équipes.

## LES TEMPS FORTS ET CHIFFRES CLÉS DU HDH

### 61 PROJETS INNOVANTS...



### ... QUI PROGRESSEDENT



## UN RAYONNEMENT INTERNATIONAL

**7**

MEMBRES  
INTERNATIONAUX AU  
CONSEIL SCIENTIFIQUE  
CONSULTATIF

**3**

ÉVÈNEMENTS ORGANISÉS  
PAR LE HDH :

**700**

participants à l'évènement "IA et la  
médecine : promesses et limites"  
co-organisé avec le MIT en mai 2021

**350**

participants à la 3<sup>e</sup> édition du colloque  
"données de santé et intelligence  
collective" organisé par le HDH autour  
de 17 intervenants en décembre 2021

**428**

participants à l'école d'hiver  
« AI4HEALTH » en janvier 2022

**15**

PARTENAIRES AU  
SEIN DU CONSORTIUM  
CANDIDAT À EHDS

**3**

PARTICIPATIONS  
À DES APPELS  
À PROJETS  
EUROPÉENS

**1**

E-CONSULTATION  
EUROPÉENNE

## UN PROJET RÉSOLUMENT CITOYEN

**6**

PARTENARIATS AVEC  
DES ASSOCIATIONS




2

# RETOUR SUR LA FEUILLE DE ROUTE STRATÉGIQUE DU HEALTH DATA HUB



# QUATRE ENJEUX INSCRITS DANS LA FEUILLE DE ROUTE STRATÉGIQUE 2020 - 2022

**En janvier 2020, le HDH a publié sa feuille de route stratégique. Elle s'articule autour de quatre grands enjeux stratégiques :**



**Décloisonner le patrimoine de données de santé** et améliorer sa qualité ;

**Développer une offre de services performante** pour les utilisateurs de données de santé ;

**Placer la France comme un leader** de l'utilisation des données de santé dans l'écosystème européen et international ;

**Garantir la participation de la société civile** à l'utilisation des données de santé.







## ENJEU 1

# DÉCLOISONNER LE PATRIMOINE DE DONNÉES DE SANTÉ ET AMÉLIORER SA QUALITÉ

L'une des missions du HDH est de mettre à la disposition des porteurs de projets un catalogue de bases de données. Ce catalogue réunit une collection de bases de données pseudonymisées que le HDH est habilité à mettre à leur disposition. Des échanges sont en cours avec une vingtaine de partenaires pressentis tels que Santé Publique France, l'AP-HP, l'Inserm, l'Université de Bordeaux, Unicancer, etc.

### ► Action 1.1 – Inscrire au catalogue des jeux de données prioritaires

La loi du 24 juillet 2019, relative à l'organisation et à la transformation du système de santé, a étendu le Système National des Données de Santé (SNDS). Il s'agit désormais, au-delà des données médico-administratives, de toutes les données de santé associées à un financement public, c'est-à-dire celles de registres, de cohortes de recherche ou d'entrepôts de données hospitalières, etc. La grande base de données médico-administratives, rassemblant historiquement le SNIIRAM, le PMSI et le CépiDc, est désormais nommée "base principale" du SNDS.

Si le HDH est chargé par la loi de réunir les données du SNDS, de les organiser et de les mettre à disposition, cela ne signifie pas qu'elles sont ou seront réunies au sein d'un unique fichier ni même d'une plateforme.

Le HDH est habilité à les mettre à la disposition des projets, soit de manière ponctuelle - les données précises requises par un projet et effacées ensuite -, soit de manière plus pérenne afin de mutualiser les efforts de déplacement de données pour plusieurs projets.

Quoi qu'il en soit, les bases de données source ne sont pas modifiées, elles sont simplement

copiées, et ce dans le respect de la réglementation. Lors de ces copies, les informations personnelles directement identifiantes sont supprimées et les personnes concernées informées.

Cette copie de bases de données, qui a pour but de servir un grand nombre de projets, est très encadrée. Les bases de données concernées sont listées sur un arrêté pris par le Ministère des Solidarités et de la Santé après avis de la CNIL.

Cette démarche présente un intérêt car déplacer ces données demande des ressources aussi bien du côté de l'acteur qui administre la base de données que du HDH qui la réceptionne. En s'organisant pour effectuer cette copie une seule fois accompagnée de mises à jour, les acteurs mutualisent ces efforts. Que ce soit pour mettre en place les modalités techniques de transfert et les modalités d'information des personnes concernées ou pour enrichir la base de données avec d'autres sources dans la plateforme du HDH.

L'autre intérêt de cette démarche pour les acteurs à l'origine des données est de pouvoir déléguer au HDH la contractualisation avec les utilisateurs, la préparation, le croisement et la mise à disposition des données dans des environnements informatiques dédiés par projets et hautement sécurisés, mais aussi la maintenance des environnements.

Toutes les bases de données ne peuvent cependant pas être concernées par ce dispositif. Elles sont donc sélectionnées, et la collection retenue compose ce que l'on appelle le « catalogue » du HDH. Ce dernier se construit donc progressivement, en lien avec les acteurs à l'origine ou ceux qui administrent les bases de données concernées, sous l'égide du Comité stratégique du Système National des Données de Santé, créé en 2021 par le décret SNDS.

# 18

## BASES PRIORITAIRES POUR LA 1<sup>ÈRE</sup> VERSION DU CATALOGUE



### **La publication du décret, une étape de plus de mise en œuvre de la loi de 2019**

Avec le décret relatif au SNDS du 29 juin 2021, les pouvoirs publics mettent en place le fonctionnement de la gouvernance du SNDS et le HDH en tant que co-responsable pour le traitement de sa base principale, avec l'Assurance Maladie. C'est également ce décret qui nomme le HDH comme responsable du traitement du catalogue du SNDS.

Le décret modifie en outre la liste des organismes, établissements ou services bénéficiant d'un accès permanent aux données du SNDS, en raison de leurs missions de service public. Il prévoit les conditions d'information des personnes et aménage l'application des droits prévus par le RGPD, dans le contexte du SNDS. Ainsi, le droit d'opposition ne s'applique ni à la constitution de la base principale du SNDS ni à la mise à disposition de ces données pour les organismes bénéficiant d'un accès permanent. Néanmoins,

les personnes peuvent s'adresser au directeur du HDH ou au directeur de l'organisme gestionnaire d'assurance maladie obligatoire dont elles relèvent pour exercer leur droit d'accès, droit de rectification et droit à la limitation pour les données de la base principale.

Afin de structurer et rendre plus accessibles les données du SNDS, le décret a décrit la distinction entre « base principale » et « catalogue ». Il a également créé un comité stratégique, notamment chargé d'identifier les bases de données existantes qui ont vocation à s'inscrire dans le catalogue, pour appuyer le ministre dans la fixation des orientations générales du développement du SNDS.

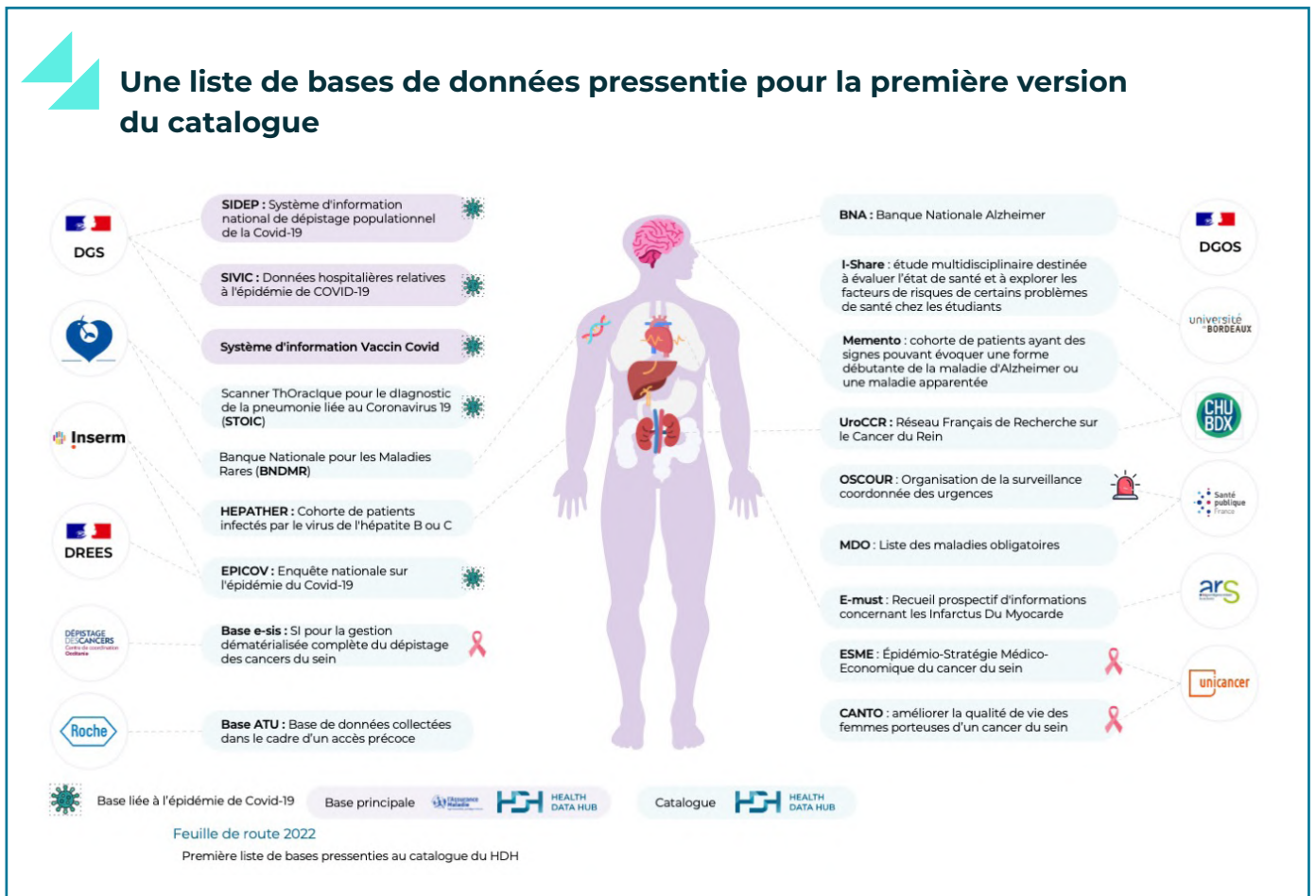
La liste des bases de données inscrites au catalogue doit en effet être publiée dans un arrêté du ministre chargé de la santé. Cet arrêté devra préciser également les flux prévus dans la base principale, au-delà des flux historiques. Ainsi, un certain nombre de systèmes d'information incluant des données relatives à l'épidémie pourraient rejoindre demain la base principale. Cet arrêté est encore en cours d'instruction par la CNIL avec un travail d'échanges entre la CNIL, la DREES et les responsables de traitement HDH et Cnam.

Une fois ces textes publiés, le dispositif sera complété par une demande d'autorisation auprès de la CNIL pour utiliser l'environnement technologique du HDH, déjà utilisé pour les projets ponctuels, à des fins d'hébergement pérenne de ces différentes sources de données.

En pratique, cela signifie que les équipes du HDH pourront également être en mesure de produire les extractions de la base principale à destination des porteurs de projets habilités. Cette capacité permettra de venir compléter l'offre de service de l'Assurance Maladie et facilitera la mise à disposition de ces données dans l'environnement technologique du HDH lorsque cela est souhaité

par les porteurs de projets. Il en sera de même pour les autres bases du catalogue puisque, en application du décret relatif au SNDS, le HDH est désormais responsable du stockage et de la mise à disposition des données de la base principale et de l'ensemble des bases de données du catalogue ainsi que des opérations de pseudonymisation supplémentaires pour les données dont il assure la mise à disposition, après que la Cnam a assuré la pseudonymisation initiale permettant l'appariement potentiel des bases entre elles et avec la copie de la base principale dont le HDH est destinataire.

En attendant que les derniers textes soient publiés, le HDH et la CNAM travaillent depuis plusieurs mois activement pour déployer cette nouvelle activité, à la fois sur les aspects techniques de préparation du transfert, mais également pour permettre la montée en compétence des ingénieurs du HDH. Ces derniers seront en effet amenés à réaliser demain des activités intensives en savoir-faire, aujourd'hui menées principalement par l'Assurance Maladie.



## **Des engagements vis-à-vis des responsables de données**

Dans le cadre du partage de ces données, le HDH a pris en 2020 plusieurs engagements vis-à-vis de ceux qui les collectent, c'est-à-dire les responsables de données.

- ▶ Il les appuie dans la collecte, la standardisation et la documentation des données. Au-delà de sa contribution financière, il peut mettre à leur disposition des compétences pour les aider à préparer et transférer ces données.
- ▶ Il leur fournit un hébergement sécurisé et les accompagne dans la mise en conformité avec la RGPD.




- ▶ Il valorise leur travail et leur expertise. Pour ce faire, il renforce la visibilité des bases de recherche, ce qui augmente les citations mentionnant leur réutilisation et il les alerte à chaque fois que des résultats obtenus grâce à leurs données sont publiés. Enfin, il favorise, voire accompagne, les partenariats entre les responsables de données et les utilisateurs et leur offre la possibilité de tarifier leurs prestations.

## Notre offre de service à destination des responsables de données

### Contrat et modalités de service

OFFRE	PRINCIPE	STATUT
 <b>Convention type</b>	<p>Une convention type, prête à l'emploi, précise les grands principes de partenariat entre le Health Data Hub et le Responsable de données pour la mise au catalogue de la base. Cette convention type est actuellement stabilisée et partagée avec de nombreux Responsables de données. La valorisation économique de l'accès aux données est toutefois toujours en cours d'instruction.</p>	 <b>Finalisé</b>
 <b>Modalité technique de partage</b>	<p>Le Health Data Hub propose un transfert sécurisé à trois niveaux : l'établissement d'un réseau virtuel privé (VPN), la mise en place d'un flux de données chiffrées (SFTP) et le chiffrement des données transmises par le Responsable des données. Le canal de transfert est en cours de mise en place auprès de quelques partenaires.</p>	 <b>Finalisé</b>
 <b>Open sciences</b>	<p>Le Health Data Hub a pour objectif de favoriser le partage, sous licence ouverte, des connaissances et ressources sur la base principale du SNDS et les autres bases de données. Différents contributeurs et utilisateurs sont à l'origine de ce partage.</p>	 <b>Finalisé</b>
 <b>Publicité de la base via le méta catalogue</b>	<p>Le Health Data Hub s'engage à mettre en avant l'équipe du Responsable de données, sa gouvernance, ses travaux et à améliorer sa visibilité auprès de la communauté scientifique. Pour cela il diffuse une fiche d'identité relative à la base de données. Un catalogue de métadonnées plus détaillé, qui pourra s'enrichir progressivement, est en cours de déploiement.</p>	 <b>Finalisé</b>
 <b>Tarification</b>	<p>La possibilité d'une rétribution financière est à l'étude afin de compenser une partie de l'effort de collecte. Mais elle doit s'inscrire dans le respect d'une réglementation complexe.</p> <p>En 2021, une première analyse fine des différents textes susceptibles d'interférer avec la mise en place d'un modèle de redevance a été finalisée. En 2022, une proposition de modèle sera élaborée, en collaboration avec les instances du HDH.</p>	 <b>En cours</b>

## Accompagnement réglementaire

OFFRE	PRINCIPE	STATUT
 <b>Conformité réglementaire de la base</b>	Le Health Data Hub aide les acteurs qui le souhaitent à vérifier la conformité de la base de données. Pour cela, il leur a proposé un kit d'auto-évaluation.	 <b>En cours</b>
 <b>Information patient</b>	Des mentions d'information type concernant la transmission des données au Health Data Hub (catalogue et appariement à la base principale du SNDS) ont été finalisées et partagées avec de nombreux Responsables de données. Trois profils ont été identifiés pour adapter les communications. Ces mentions doivent également être adaptées aux canaux de communication existants du Responsable de données pour faciliter leur diffusion. En parallèle, le Health Data Hub délivrera un second niveau d'information sur son site internet.	 <b>Finalisé</b>
 <b>Procédure du droit d'opposition</b>	Le Health Data Hub garantit l'exercice des droits des personnes, notamment en publiant sur son site internet le Répertoire Public et la procédure expliquant les modalités d'exercice de leurs droits. Une procédure du droit d'opposition a été finalisée et partagée avec les Responsables de données concernés. Un cahier des charges est en cours de finalisation pour mettre en œuvre un formulaire rendant plus pratique l'exercice des droits.	 <b>Finalisé</b>
 <b>Mise en relation avec le CESREES</b>	Les Responsables de données alimentant le catalogue du Health Data Hub sont associés à l'instruction du CESREES. Cela permet de bénéficier de leur expertise pour juger de la faisabilité, de la méthodologie, de la pertinence ou encore du caractère éthique d'une étude mobilisant les données dont ils sont à l'origine. Le secrétariat de ce comité, assuré par le Health Data Hub, organise cette mise en relation.	 <b>En cours</b>

## Zoom Covid

L'épidémie de Covid19 a démontré le caractère essentiel du partage des données de santé. Pleinement mobilisé dès le printemps 2020, le HDH a poursuivi ses actions en 2021 en acceptant d'office tout projet relatif à l'épidémie, quand bien même les services du HDH étaient encore en phase de test.

L'arrêté relatif à l'état d'urgence sanitaire entré en vigueur le 21 avril 2020, complétant l'arrêté du 23 mars 2020, permettant au HDH de réunir et mettre à disposition des données relatives à l'épidémie à des fins de

recherche est entré dans le droit commun le 31 mai 2021. Il restera valable jusqu'à ce que l'arrêté portant les bases du catalogue du HDH soit publié.

Tout au long de 2021, 4 nouveaux projets relatifs à l'épidémie ont été déposés auprès du HDH, ce qui porte à 14 le nombre de projets actuellement accompagnés en lien avec la pandémie.

Ces projets sont variés et permettent d'étudier différentes facettes du phénomène. Ils ont pour but d'améliorer les modèles

de prévision de l'évolution de l'épidémie, d'analyser les inégalités sociales de santé ou encore d'étudier les parcours de soin de certaines populations particulièrement affectées en raison de leurs comorbidités. Le délai d'obtention des autorisations réglementaires et le délai de réalisation n'ont pas permis à ces projets d'aboutir en 2021 mais certains pourront produire des résultats dès le premier semestre 2022. En effet, tous ont reçu l'avis favorable du CESREES et plus d'un tiers ont été autorisés par la CNIL.

### ► Action 1.2 – Installer une procédure nationale automatisée pour faciliter la sélection et la remontée des données au catalogue et leur appariement

#### Mise en œuvre de l'offre de service pour faciliter la gestion des appariements

Les données étant dispersées sur le territoire, il est essentiel de pouvoir les croiser pour les enrichir mutuellement ou pour obtenir la masse critique nécessaire à certains projets.

Idéalement, pour croiser ces données, il faut disposer d'une variable pivot commune. Dans le monde des données de santé, cette variable peut être l'identifiant individuel d'une personne (NIR), lorsqu'il est disponible. Mais cet identifiant étant soumis à un cadre réglementaire strict, il doit

être transformé pour le protéger et éviter que la personne concernée puisse être identifiée, tout en permettant les travaux de recherche.

Pour qu'il garde son rôle de pivot, il doit cependant toujours être transformé de la même manière. Il existe donc une procédure de transformation qui implique différentes entités de l'Assurance maladie, permettant de s'assurer que les mêmes acteurs ne détiennent pas le NIR initial et le résultat final de sa transformation.

Mobiliser cette procédure demeure complexe et un certain nombre d'actions doivent être réalisées à la main. Dans un contexte où les projets se multiplient et où les demandes d'appariement vont aller croissant, il est nécessaire d'industrialiser cette procédure. C'est pourquoi le HDH et la CNAM travaillent sur un outil, appelé "concentrateur", qui permettra d'automatiser certaines

étapes et de rendre la procédure plus accessible à ses nombreux bénéficiaires. Il est prévu d'élargir ses fonctionnalités à l'automatisation des croisements impliquant une reconstitution du NIR par la CNAV.

À noter qu'au-delà d'une réalisation plus aisée des appariements, cet outil permettra aussi de manipuler le NIR en toute sécurité pour que le droit des personnes concernées puisse être mis en œuvre et que celles qui le souhaitent puissent, par exemple, s'opposer à la réutilisation de certaines de leurs données à des fins de recherche. Cela prendra la forme d'une interface accessible à tous les citoyens.

Un premier cahier des charges de cet outil a été conçu en 2021 ; sa mise en œuvre pratique est attendue pour 2022.

### ▶ **Action 1.3 – Créer une « cohorte socle » rassemblant des données cliniques et médico-administratives finement localisées sur un échantillon représentatif de la population**

Un certain nombre de bases préexistantes ont été identifiées pour être partagées au sein du catalogue du Health Data Hub. Néanmoins, l'articulation entre les établissements de santé, leurs entrepôts de données de santé et le Health Data Hub revêt également un potentiel très important pour la constitution de bases de données multicentriques thématiques ou pour la remontée de données à la demande. Plusieurs actions ont donc été menées en 2021 pour explorer ces possibilités.

### **Un partenariat avec Unicancer autour du programme UNIBASE**

Avec plus de 380 000 nouveaux cas par an, les cancers représentent en France la première cause de décès chez l'homme et la deuxième chez la femme. Pour accélérer la recherche et améliorer la survie de ces malades, il est aujourd'hui impératif de mettre en relation les données disponibles et de les exploiter dans un environnement technologique sécurisé, permettant notamment d'utiliser l'intelligence artificielle.

C'est dans ce but que le HDH a signé en juillet 2021 un partenariat avec Unicancer, acteur majeur de l'innovation dans la lutte contre le cancer. Le programme UNIBASE est au cœur de ce partenariat. Il vise à créer en trois ans une collection de bases de référence en cancérologie, à partir des données des Centres de lutte contre le cancer (CLCC) répartis sur tout le territoire. Ces données seront mises à la disposition des acteurs de la recherche pour qu'ils puissent déployer des études multicentriques collaboratives et relever des défis scientifiques impossibles à résoudre sans recourir à des grands volumes de données issues de la vie réelle. Le fait de centraliser les données, de croiser des sources multiples et de combiner données cliniques, données d'imagerie et données omiques contribue à faire d'UNIBASE un programme unique en cancérologie.

Pour lui donner corps, le HDH et Unicancer ont lancé, le 25 novembre 2021, un appel à manifestation d'intérêt (AMI) ouvert aux 10 CLCC disposant d'entrepôts de données Consore, le moteur de recherche pour le bigdata en cancérologie. Trois projets pilotes, en phase avec le Plan cancer 2021-2030, seront sélectionnés et déployés à partir de 2022 sur la plateforme technologique du HDH, avec un accompagnement et un financement du HDH.



Forts de cette expérience, Unicancer et le HDH souhaitent, par la suite, ouvrir UNIBASE aux CLCC n'utilisant pas Consore, ainsi qu'à d'autres structures hospitalières. Des travaux seront menés pour que les données soient interopérables et modélisées conformément à des standards internationaux.

## 3 PROJETS PILOTES SERONT RETENUS DANS LE CADRE DU PROGRAMME UNIBASE

### Un programme partenarial pour les établissements de santé souhaitant réaliser des projets de recherche multicentriques en mutualisant leurs données

En 2021, le HDH a lancé un programme partenarial ouvert à tous les établissements de santé qui souhaitent renforcer leur stratégie liée à la donnée, et explorer avec le HDH des solutions pour mutualiser et enrichir leurs données hospitalières. Concrètement, il s'agit de réaliser des projets de recherche exploitant des bases de données de santé mutualisées entre les différents centres hospitaliers. Le HDH leur apporte son appui pour animer ces projets ainsi que son appui financier, et met à leur disposition sa

plateforme technologique pour croiser les données consolidées avec les données issues de la base principale du SNDS et réaliser le cas d'usage identifié. Ce programme vise aussi à renforcer la capacité de ces acteurs à travailler en réseau (pour pouvoir par exemple répondre ensemble à des opportunités stratégiques supranationales) et à développer leur expertise grâce au partage d'expérience.

### Le Health Data Hub lauréat d'un appel d'offre de l'Agence européenne du médicament

Fin 2021, l'Agence européenne du Médicament a retenu la candidature déposée par le HDH et ses partenaires dans le cadre d'un appel d'offre sur des données en vie réelle. Objectif : pouvoir accéder à des bases de données individuelles hospitalières, provenant des pays de l'UE, à la maille patient et au format OMOP-CDM.

Pour y répondre, le Health Data Hub s'est associé à quatre établissements de santé : les HCL, le Centre Léon Bérard, le CHRU de Nancy et l'hôpital Saint Joseph, ainsi qu'à la plateforme de Pharmaco-épidémiologie de Bordeaux, l'ANS et Capgemini. Ensemble, ils vont consolider une base de données cliniques hospitalières, enrichies par des données médico-administratives de la base principale du SNDS et transformées dans un modèle commun international permettant de les rendre interopérables. Avec ces données, l'Agence européenne du médicament pourra instruire différentes problématiques relatives aux effets secondaires des médicaments. Pour l'ensemble des partenaires, ce projet est l'occasion d'élaborer des méthodes de travail communes permettant de renouveler l'expérience toutes les fois où cela sera pertinent, en réponse à des enjeux nationaux ou internationaux.

## 17 NOMBRE D'ÉTABLISSEMENTS AVEC LEQUEL LE HDH TRAVAILLE

## **Le HDH membre du consortium pour l'Observatoire de médecine de ville.**

Le HDH est membre du consortium porté par le Collège national des généralistes enseignants (CNGE) qui travaille à la constitution d'un « entrepôt de données issues de la médecine de ville », c'est-à-dire des professionnels de santé libéraux et des soins réalisés en dehors de l'hôpital ou des établissements d'hébergement médico-social. Ce consortium regroupe les start-ups Loamics et MyDataModels, ainsi que le CHU de Rouen, l'Université Nice Côte-d'Azur et l'Université de Normandie, avec le soutien du CNAM et du Ministère des Solidarités et de la Santé.

En 2021, il a franchi la première étape de son appel à manifestation d'intérêt pour la stratégie d'accélération « santé numérique ».

Ce projet présente un intérêt majeur parce qu'il permettra de :

- ▶ Offrir une vision complète des parcours de soin, y compris en amont et en aval d'un éventuel passage à l'hôpital, grâce au complément qu'apporteront les données de consultation en ville aux données du SNDS ;
- ▶ Détecter des signaux précurseurs de résurgence d'une épidémie, grâce au pilotage de l'évolution des diagnostics de consultation au niveau national ;
- ▶ Identifier plus précisément les facteurs de risque, par une meilleure connaissance de la symptomatologie des pathologies, dont, en priorité, la Covid-19.

Même si la nécessité d'un tel observatoire est admise de longue date, la crise sanitaire a accéléré ce projet. À l'heure où le recours à la médecine de ville doit trouver sa place à côté et en complémentarité des services d'urgences, la collecte à large échelle de données de ville offre un appui précieux à la gestion d'épidémies comme la Covid-19. Mais ces données, récoltées dans plus de 20 solutions différentes de dossiers médicaux électroniques utilisées par les médecins, ne sont aujourd'hui ni harmonisées, ni consolidées à l'échelle nationale. Le projet d'observatoire de la médecine de ville a pour ambition de constituer un entrepôt de données chaîné avec le SNDS, grâce à un circuit de remontées de données issues des principales solutions éditeurs utilisées par les médecins généralistes.

En incluant le HDH dans ce consortium, les porteurs du projet ont souhaité prévoir, dès son lancement, le partage des données à des fins d'intérêt général via le catalogue de données du HDH.

## ► Action 1.4 – Promouvoir les producteurs participant à la dynamique de partage des données

Pour promouvoir les producteurs participant à la dynamique de partage des données en 2021, le HDH a par exemple :

ÉLABORÉ UN PROTOTYPE DE CATALOGUE DE MÉTADONNÉES PERMETTANT DE METTRE EN AVANT LES BASES DE DONNÉES PARTAGÉES

PROCÉDÉ À UNE ANALYSE JURIDIQUE POUR ENCADRER LA MISE EN ŒUVRE DE LA TARIFICATION ET CONSTRUIRE UN RÉGIME DE REDEVANCE

RÉÉDITÉ SON COLLOQUE ANNUEL « DONNÉES DE SANTÉ ET INTELLIGENCE COLLECTIVE »

### Un catalogue de métadonnées

Pour faciliter l'accès aux données de santé et en intensifier l'usage dans les projets de recherche, une "simple" mise à disposition de ces données n'est pas suffisante. Il faut aussi recenser et documenter précisément le contenu des bases disponibles, afin que les utilisateurs convergent rapidement vers les données qui répondent à leur question de recherche. C'est dans cette optique que le Health Data Hub collecte et met en ligne les **métadonnées** de chacune des bases de son catalogue.

Concrètement, la consultation des métadonnées doit accélérer l'accès aux données, en optimisant les interactions entre les chercheurs, les producteurs et le HDH. Elle fournit également aux chercheurs les coordonnées des producteurs et favorise ainsi les échanges et collaborations entre scientifiques. Enfin, l'ouverture des métadonnées matérialise le souhait de transparence du HDH puisqu'elle offre, sans divulguer d'informations identifiantes, un maximum de détails sur les données collectées à des fins de recherche.

Techniquement, ce **catalogue de métadonnées** comporte trois niveaux d'information (Base, Tables, Variables) et dispose d'un moteur de recherche pour identifier les bases du HDH les plus proches de la problématique de l'utilisateur. Il permet ensuite d'en explorer le contenu, de la description générale d'une base, son contexte scientifique, les modalités de sa récolte et de sa mise à disposition, jusqu'à la description technique et statistique de chaque variable qu'elle renferme.

La solution déployée se base sur la connaissance métier de nombreux experts et son implémentation repose sur le système de gestion de données open-source CKAN, pour lequel le HDH a écrit une extension disponible dans un dépôt Gitlab public. Elle s'inscrit dans une démarche internationale de description standard des données de santé et garantira, à terme, les meilleures modalités d'accès et d'échange avec d'autres plateformes de données de santé telles que l'European Health Data Space (EHDS).

## Une redevance pour les acteurs privés accédant directement aux données du catalogue

Pour dynamiser le partage des données de santé, le HDH a instruit en 2021 la mise en place d'une redevance pour les acteurs privés souhaitant accéder directement aux données de son catalogue. Cette instruction s'est traduite par une phase d'analyse juridique pour identifier le cadre légal et réglementaire d'un éventuel régime de redevance. Plusieurs responsables de données ont été interrogés sur leurs pratiques de tarification. Ils estiment que ces pratiques ne leur permettent pas de garantir la pérennité des différentes bases et infrastructures de données publiques en santé. En ligne avec ce constat, le HDH a élaboré un argumentaire pour renforcer le financement public des infrastructures de données stratégiques des établissements de santé. La seconde phase, visant à construire et mettre en œuvre le futur régime de redevance de l'accès aux données du catalogue du HDH, s'ouvre dès 2022.

## Focus sur la 3<sup>ème</sup> édition du Colloque « Données de santé et intelligence collective »

Fort du succès des 2 précédentes éditions, le HDH a réédité, le 15 décembre 2021, son colloque annuel « Données de santé et intelligence collective ». Un événement co-organisé avec la Délégation Ministérielle au Numérique en Santé et le Grand Défi « Amélioration des diagnostics médicaux par l'IA », sur le site de Parisanté Campus. Plus de 350 participants y ont assisté, en ligne, et interagi avec une vingtaine d'intervenants. Au programme : les bénéfices des données de santé pour la recherche, la santé publique et le pilotage du système de santé, mais aussi les enjeux d'attractivité et de compétitivité que représente la réutilisation de ces données pour la France.

**350**  
**PERSONNES ONT PARTICIPÉ AU COLLOQUE DONNÉES DE SANTÉ ET INTELLIGENCE COLLECTIVE**



## ENJEU 2

# RENFORCER L'USAGE DES DONNÉES PAR UNE OFFRE DE SERVICE ATTRACTIVE

Pour pouvoir utiliser les données de santé à des fins de recherche, trois étapes doivent être franchies : l'étape réglementaire de la demande d'accès, l'accès effectif aux données et la réalisation du projet. Chacune de ces étapes présente des axes d'amélioration sur lesquels le HDH et ses partenaires s'investissent.

### ► Action 2.1 – Mettre en place des procédures simplifiées et un processus d'accès industrialisé

Le HDH accompagne les porteurs de projet tout au long de leur demande d'accès, avec l'appui indispensable des équipes de ses partenaires, notamment chez les acteurs qui administrent les bases de données.

### Comment le CESREES instruit les demandes d'accès aux données de santé ?

Créé en 2020, le CESREES, dont le HDH assure le secrétariat, rend son avis sur chaque demande d'accès aux données de santé. Il se prononce sur les points suivants :

- la finalité et la méthodologie de la recherche,
- la nécessité du recours à des données de santé à caractère personnel,
- la pertinence éthique,
- la qualité scientifique du projet,
- et, le cas échéant, l'intérêt public que sert le projet.

# 366

DOSSIERS NÉCESSITANT  
UNE AUTORISATION DE  
LA CNIL DÉPOSÉS AU HDH  
EN 2021



# 279

DOSSIERS EXPERTISÉS  
PAR LE CESREES  
EN 2021



## **Première contribution du HDH à procédure simplifiée de la CNIL**

Le 26 octobre 2021, la CNIL a publié un « référentiel relatif aux traitements de données à caractère personnel mis en œuvre à des fins de création d'entrepôts de données dans le domaine de la santé ».

Il s'agit d'une procédure simplifiée qui permet à tout responsable de traitement respectant les conditions énoncées dans ce référentiel de constituer un entrepôt de données de santé, après avoir adressé un engagement de conformité à la CNIL. Un entrepôt de données de santé est défini par la Commission comme un rassemblement de données massives, constitué pour une longue durée, à des fins de recherches, études et évaluations dans le domaine de la santé.

Préalablement à cette publication et conformément à l'article 66 II de la Loi Informatique et Libertés, le HDH a été saisi par la Commission afin de se prononcer sur ce référentiel. Pour y répondre, le HDH a travaillé en concertation avec les représentants de son conseil d'administration sur l'analyse juridique du projet de texte, afin de recueillir les observations de chacun. Cette contribution a été transmise à la Commission pour qu'elle puisse ajuster le projet de texte, en tenant compte des besoins des acteurs.

Tout au long de 2022, le HDH aura l'occasion de contribuer à de nouvelles procédures permettant de simplifier l'accès aux données de santé en France.

## **Une amélioration en continu de l'offre de service en 2021**

Pour donner à un projet toutes les chances d'obtenir un avis favorable du CESREES puis d'être autorisé par la CNIL, il faut bien maîtriser la procédure et ses différentes étapes, ainsi que le contenu du dossier à transmettre. Autant d'éléments qui peuvent sembler complexes, en particulier aux nouveaux entrants. C'est pourquoi le Health Data Hub, la Cnam et le CESREES ont conçu un "starter kit" destiné aux porteurs de projets. Ce kit pédagogique comprend non seulement des points de vigilance, des conseils et des exemples pour bien remplir son dossier mais aussi un protocole fictif complet et une « check list », pour que les porteurs de projets puissent identifier concrètement les éléments attendus.

Un 2<sup>ème</sup> guide pédagogique a également été édité avec la Cnam pour aider ceux qui souhaitent accéder aux données du SNDS, à remplir leur expression des besoins (EDB). Cette EDB est en effet essentielle pour préciser le périmètre de données à extraire.

S'y ajoute un 3<sup>ème</sup> outil, édité en 2021, visant à aider les porteurs de projets à mieux se repérer dans la diversité des procédures d'accès aux données de santé. Ce guide permet à ces derniers de caractériser leur projet et leur donne les éléments clefs pour sélectionner la procédure réglementaire adéquate à mettre en œuvre.

En complément de ces outils, le HDH a également mis en place en 2021 deux formations pour aider ceux qui souhaitent accéder aux données de santé.

► Une formation sur les données du SNDS et ses modalités d'accès. Animée en collaboration avec l'Assurance Maladie, cette session de 3h s'adresse à tous les porteurs de projet qui souhaitent utiliser ces données mais ne connaissent pas toujours leur contenu, la façon dont on y accède et les démarches réglementaires associées.

▶ Une présentation des procédures d'accès aux données personnelles de santé. Cette session d'1h en visioconférence s'adresse à tous les porteurs de projet qui veulent identifier les procédures les mieux adaptées à leurs projets. Elle détaille les éléments clés pour caractériser une recherche, savoir si elle implique la personne humaine ou non et décline les procédures réglementaires d'accès.

▶ **Action 2.2 – Renforcer et maintenir à l'état de l'art l'offre technologique du HDH**

**Un « espace d'analyse » sécurisé sur une plateforme technologique**

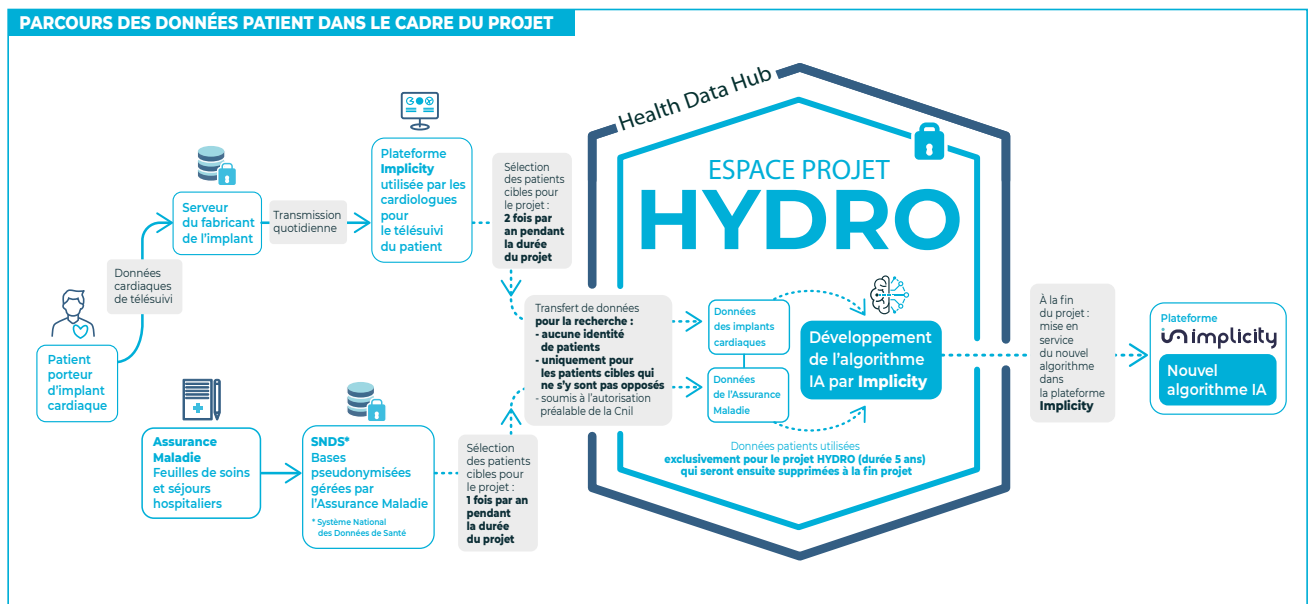
Le HDH accompagne les chercheurs, entrepreneurs et associations de patients qui souhaitent mener et accélérer leurs projets d'intérêt général, en mettant à leur disposition un « espace d'analyse » sécurisé sur une plateforme technologique. Cet espace permet de recevoir, stocker, traiter et visualiser les données de santé nécessaires aux projets sans pouvoir les exporter.

**6 FORMATIONS DISPONIBLES AU 7 OCTOBRE 2021**

**HYDRO** by **implicit** Un projet de recherche en intelligence artificielle (IA) porté par la société **Implicit**, en partenariat avec le **Health Data Hub**, avec l'appui de l'**Assurance Maladie**

1<sup>ER</sup> PROJET PILOTE DU **HEALTH DATA HUB**

Un algorithme d'IA pour prédire les crises d'insuffisance cardiaque et réduire les risques d'hospitalisation pour les patients porteurs d'implants cardiaques.



Il est isolé de l'extérieur, c'est-à-dire dédié uniquement au projet en question, sans transfert des données depuis et vers un poste local. Fondé sur une infrastructure dite « élastique », il offre notamment des ressources de calcul à capacité variable permettant de développer les usages les plus innovants.

Le contexte de sécurité de la plateforme technologique du Health Data Hub a été défini dans le cadre d'une démarche conduite en collaboration avec l'ANSSI (Agence nationale de la sécurité des systèmes d'information) et le Haut fonctionnaire de défense et de sécurité du ministère des Solidarités et de la Santé. La sécurité de la plateforme technologique est, en outre, testée et revue en permanence par le biais d'audits techniques menés par des prestataires indépendants qualifiés par l'ANSSI. Enfin, l'ANSSI et la CNIL participent à l'homologation de chaque nouvelle version de la plateforme technique.

En 2021, un audit PASSI a été réalisé sur la plateforme technologique préalablement à la réunion de revoyure de son homologation le 30 novembre 2021. Un audit conduit par l'ANSSI avait déjà été mené avec succès en début d'année. Dans le cadre de cette revoyure, le HDH a également mis à jour son analyse de risque via la méthodologie EBIOS RM. Cette révision a permis de prendre en compte les nouvelles menaces et d'identifier des risques émergents. Dans le cadre de l'amélioration continue, un plan de traitement des risques spécifique a été mis en place à l'issue de ces révisions.

Afin de répondre aux forts enjeux de sécurité, le HDH doit travailler de concert avec ses partenaires. Il a par exemple mis en place une procédure de sécurisation, validée par l'ANSSI, pour les transferts de données par les responsables de données et un guide de bonnes pratiques sur la pseudonymisation partagé avec la CNIL. Le HDH a également lancé des discussions avec l'Institut national de recherche en informatique et en automatique (INRIA) autour d'un partenariat sur l'analyse des requêtes suspectes.

### **Une offre technologique en évolution constante**

Traiter des données de santé à travers des approches analytiques innovantes nécessite de relever de nombreux défis technologiques. C'est pourquoi le HDH propose à tous les porteurs de projets une offre logicielle qui évolue en permanence, dans un environnement sécurisé et à l'état de l'art.

Pour réaliser leurs études, il met à leur disposition un espace projet dans lequel ils peuvent recevoir, stocker, traiter en mode collaboratif et visualiser les données de santé nécessaires à leur projet, en toute sécurité.

À l'écoute des utilisateurs, le HDH a mis en place en 2021 une feuille de route produit, pour prioriser leurs besoins en matière d'offre logicielle.



## Le HDH propose à ce jour dans son offre logicielle :

 <p>Les environnements de développement intégré (IDE) JupyterLab et RStudio, proposant de travailler respectivement avec les langages de programmation Python et R</p>	
 <p>Les modules PySpark et Sparklyr, permettant le traitement de grandes volumétries de données</p>	
 <p>L'outil de visualisation de données Superset</p>	
 <p>Les systèmes de gestion de bases de données PostgreSQL et MySQL</p>	<p>PostgreSQL</p> 
 <p>L'outil de visualisation et d'annotation d'images Cytomine</p>	
 <p>Les outils de collaboration, Git et Gitlab</p>	

## Zoom sur la réversibilité

### Poursuite du plan d'actions de réversibilité

Entièrement administrée par des agents du HDH, aux rôles et responsabilités bien définis, la plateforme technologique mobilise des composants industriels choisis pour leur performance, leur sécurité et leur élasticité.

L'offre logicielle à l'état de l'art permettra le développement des codes et algorithmes (R, python, spark.). Une « marketplace » est également prévue, à terme, pour favoriser le partage des outils pertinents avec la communauté.

Les enjeux de la plateforme technologique sont les suivants :

> La sécurité : assurer la protection des données de santé hautement sensibles qui y sont stockées.

> Les fonctionnalités et la performance : répondre aux besoins des porteurs de projets d'intérêt public, faisant appel aux approches issues de la science des données.

> Les coûts et les délais : mettre en œuvre rapidement une solution pour répondre aux enjeux de la recherche en santé et dans un contexte de compétition internationale.

Conçue dès l'origine pour être réversible en l'absence de solutions souveraines conciliant ces trois objectifs, la plateforme technologique s'appuie sur certaines solutions qui ne correspondent pas à la stratégie nationale pour un cloud de confiance. Pour autant, les travaux d'identification d'une telle solution se poursuivent avec une 3<sup>ème</sup> mise à jour de l'étude de marché visant à comparer

les solutions cloud susceptibles de répondre aux enjeux du HDH, tout en se conformant aux nouvelles exigences nationales. Cette étude, plus vaste que les précédentes, a été menée en grande partie en 2021. Elle est en cours de finalisation et sera enrichie des nouvelles solutions qui émergent, dont les solutions d'hybridation.

En parallèle, la documentation de la plateforme en Terraform se poursuit, préparant ainsi la migration vers la solution cible, ainsi que la réduction de l'adhérence aux fournisseurs actuels lorsque cela est possible.

Enfin, pour répondre à cet enjeu de souveraineté numérique de plus en plus présent, le HDH suit de près les discussions qui se tiennent dans différentes instances européennes.

### ▶ Action 2.3 – Développer une offre de service permettant un usage performant des données

#### Une animation ouverte de l'écosystème science

Si l'ambition du HDH est bien de permettre un accès plus rapide aux données de santé, il n'est pas toujours garanti que les données obtenues permettent d'aboutir à des résultats probants, fiables et valorisables. Afin de tirer le meilleur parti de ces données, un travail considérable est nécessaire pour comprendre leur organisation, la façon dont elles sont collectées, les biais et les limites qu'elles présentent. C'est dans ce but que le HDH propose à tous des informations, de la documentation et des formations,

en accès libre. Il s'inscrit, à ce titre, directement dans la démarche Science Ouverte portée par le ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation.

En pratique, le Health Data Hub et ses partenaires (DREES, Santé publique France, CNAM, ANSM, ARS, ATIH, ReDSiam, etc.) proposent depuis 2019 une documentation ouverte collaborative autour de la base principale du SNDS, dont le contenu sera complété en 2022 de nouvelles entrées et d'un guide d'initiation au SNDS. Cette documentation est accompagnée d'une série d'outils pour faciliter la prise en main des données. On y trouve des notebooks pédagogiques, un générateur de données de synthèse, un dictionnaire des tables et variables, ainsi que des programmes de manipulation des données. Ces outils, consultés

environ 500 fois par jour, dont les codes sources sont ouverts à tous dans des dépôts Gitlab, sont alimentés par la communauté et maintenus par le HDH.

S'y ajouteront, dès 2022, les nouveaux projets et outils initiés en 2021. Le projet BOAS (Bibliothèque Ouverte d'Algorithmes en Santé), mené en collaboration avec le réseau ReDSiam, en est un exemple. En soutenant et en accompagnant le développement et la validation d'algorithmes de ciblage dans le SNDS, il permettra à tous d'accéder à des implémentations et documentations indispensables pour réutiliser ces algorithmes de façon faible, ce qui facilitera la création de populations d'étude, ce projet se traduit concrètement par l'accompagnement de 4 premiers projets de validation d'algorithmes, sélectionnés lors d'un appel à manifestation d'intérêt qui se poursuit en 2022. Le HDH a également développé une cartographie des acteurs de la recherche sur le SNDS, compilant dans un tableau de bord les informations associées à plus de 900 publications. Enfin,

les travaux du HDH autour de la standardisation des données et de l'utilisation d'un modèle de données commun ouvrent de nouvelles perspectives dans le partage de programmes et de connaissances.

**11** **CONTRIBUTEURS  
À LA DÉMARCHE  
OPEN SOURCE**

**1587** **MEMBRES À LA  
COMMUNAUTÉ POUR  
30 MEETUPS ORGANISÉS**

**6049** **VISITEURS SUR LA  
DOCUMENTATION  
OUVERTE DU SNDS  
EN NOVEMBRE 2021**



#### 4 premiers lauréats sélectionnés le 8 novembre 2021

##### REPÈRES, SPF, CHRU BREST

Projet visant à développer et valider un algorithme pour identifier les sous-types d'AVC hospitalisés

##### BPE, REPÈRES, EDMUS, NEUROCENTRE MAGENDIE

Projet visant à évaluer et améliorer les performances diagnostiques d'un algorithme d'identification des poussées de Sclérose en Plaques

##### F-CRIN CRI-IMIDIATE, BNDMR, CÉPHÉPI, IPLESP

Projet visant à développer et valider des algorithmes autour des maladies inflammatoires systémiques rares



##### AHEAD (BPHC, INSERM), CHU BORDEAUX

Projet visant à développer et valider des algorithmes en orthopédie, notamment les chirurgies rachidiennes et de l'épaule

## ▶ **Action 2.4 – Diffuser l'offre de service auprès de l'ensemble des acteurs : organismes de recherche, industriels, partenariats publics privés, etc.**

L'écosystème dans lequel le Health Data Hub intervient est large. Il comprend notamment : les institutions françaises avec plus de 200 chercheurs affiliés de l'INSERM, du CNRS et de l'INRIA ; d'importants industriels, dont les grandes entreprises pharmaceutiques ; l'écosystème des start-ups (notamment via France Digitale) ; et les grands acteurs européens de la donnée de santé, dont les principales plateformes nationales, les infrastructures de recherche européennes et l'agence européenne du médicament, sans oublier les différents groupes de travail européens officiels qui travaillent sur la donnée de santé.

En 2021, 55 projets étaient accompagnés par le HDH. S'il est encore trop tôt pour mettre à l'actif de ces projets des avancées exploitables sur le plan médical, des jalons importants ont été franchis qui leur permettent de l'envisager. À titre d'exemple, le projet Hydro, porté par la start-up Implicit a requis un travail collectif de la part d'Implicit, de la Cnam et du Health Data Hub et le soutien d'une association de patients. Son objectif est de réaliser un croisement probabiliste des données de pacemaker avec les données d'hospitalisation de qualité suffisante pour que la suite de l'étude ne soit pas compromise. Le franchissement de ce jalon majeur a nécessité plusieurs mois d'engagement et d'intelligence collective.

D'une manière générale, les partenariats se sont multipliés avec des acteurs divers tels que le Datalab Normandie avec qui a été organisé un Appel à Manifestation d'Intérêt en début d'année 2021. Le Grand Défi "comment améliorer les diagnostics avec l'intelligence artificielle" est resté l'un de nos partenaires clé et a porté avec nous, en 2021, le programme Data Challenge qui a bénéficié à six sociétés savantes. Avec Unicancer, le Health Data Hub a lancé fin novembre un appel à projets pour préfigurer l'ambitieux projet de base de données majeure en oncologie.

Avec l'AP-HP, en plus des travaux stratégiques sur les conditions de succès du partage et de la valorisation des données, une dizaine de projets sont soutenus conjointement.

D'une manière générale, les initiatives s'accroissent avec des consortiums d'établissements de santé afin de favoriser un changement d'échelle. Des partenariats ont également été conclus avec des acteurs du monde de la Science Ouverte comme Redsiam, des associations de patients qui peuvent contribuer à la collecte ou à l'analyse de données ou encore des acteurs privés. Sans oublier les partenaires impliqués dans la constitution du catalogue de données, collection des bases de données qui seront partagées dans le HDH et dont l'accès sera de ce fait simplifié grâce à un cadre de partage clair et transparent, co-construit avec les acteurs.

## Bien avancés en 2021, ces projets sont actuellement en train de réaliser leur étude ou sur le point d'arriver sur la plateforme

PROJET	PORTEUR	OBJECTIF
DeepSarc	Centre Léon Bérard	Etudier l'impact des traitements du sarcome sur des données de vie réelle afin d'identifier les meilleurs schémas thérapeutiques
DEEP.PISTE	CRCDC Occitanie	Evaluer l'apport de l'intelligence artificielle dans le dépistage organisé du cancer du sein
DSI COVID	Université Toulouse III	Décrire le lien entre précarité sociale et risque d'infection par la COVID-19 et son évolution en France
CoviSAS	Université Grenoble - Alpes	Identifier la prévalence des formes sévères de COVID-19 chez les patients porteurs d'un syndrome d'apnées obstructives du sommeil (SAOS), et les combinaisons de comorbidités associées au SAOS corrélées à un taux supérieur de séjour en réanimation ou de décès.
CoData Cancer du Sein	ICANS	Anticiper et prioriser les prises en charge des patientes atteintes de pathologie mammaire non prises en charge pendant la crise sanitaire et analyser le changement des prises en charge de ces patientes, leurs éventuelles pertes de chance ainsi que l'effet de certains changements de pratiques.
HUGOSHARE	Réseau HUGO	Étudier les interactions médicamenteuses susceptibles de provoquer des effets indésirables, à partir des prescriptions médicamenteuses hospitalières
APRIORICS	CHU Toulouse	Développer un outil de description morphologique d'images microscopiques de cancers du sein pour mieux appréhender la variabilité de ces cancers.
SEDAAR	Hôpital Fondation Alphonse de Rothschild	Développer des algorithmes afin de créer un service d'aide à la décision et à l'interprétation pour les orthoptistes et ophtalmologistes.
RIAC	CHU Amiens-Picardie	Construire un ensemble d'indicateurs permettant de suivre la COVID-19 et de détecter les probables clusters et réaliser une étude épidémiologique rétrospective des données de régulation pour mieux décrire la population appelant le SAMU sans être toujours vue à l'hôpital en début de crise.
ARAC	Malakoff Humanis	Analyser les dépenses de santé en France et les restes à charge finaux des assurés
DAICAP	AP-HP	Développer un algorithme qui permettra d'obtenir un compte rendu d'IRM standardisé avec une prédiction de l'agressivité tumorale.

## **Un partenariat avec Datalab Normandie pour améliorer les soins grâce aux données de santé**

Pour faire émerger des projets améliorant la qualité des soins en Normandie, le HDH et Datalab Normandie ont conclu un partenariat et lancé un 1<sup>er</sup> appel à projets le 29 janvier 2021. 6 projets utilisant des données de santé ont été retenus. Ils recevront une aide minimum de 20 000€, chacun, grâce à la Région qui a mobilisé les fonds FEDER 2014-2020. Ils seront accompagnés sur le plan opérationnel. Datalab Normandie les mettra en relation avec les compétences régionales. Quant au HDH, il les aidera à mieux comprendre les démarches réglementaires et techniques.

Les 6 projets accompagnés par le HDH sont : Anthropometer3DNet, H3DMED, OVERCOME, Jumeau Numérique par le groupe VyV, PRIMEGE et EDSaNCoh

## **Jumeau Numérique – État de Santé d'un territoire**

Peu d'outils sont à disposition des collectivités pour simuler l'état de santé d'une population en lien avec des données environnementales. La modélisation d'un territoire via un «jumeau numérique», proposé par le groupe VyV et développé sur la base de données de santé et environnementales, permet de simuler des politiques publiques et leur impact avant leur mise en œuvre effective.

## **OVERCOME**

Le projet OVERCOME est porté par un consortium impliquant la société Hometrix Health, le laboratoire COMETE de l'Université de Caen, le CHU de Caen, et le réseau d'aides à domicile UNA-Pays d'Alençon. Il propose de recueillir les données de patients âgés suivis à domicile et présentant un risque de développer des troubles cognitifs évolutifs. Couplées à celles de l'Assurance Maladie, elles permettront de mieux catégoriser et prédire les évolutions du syndrome du risque cognitif moteur, qui peut se développer vers une maladie d'Alzheimer.

## **PRIMEGE**

Le projet PRIMEGE, porté par l'association de médecine générale ENSEREVA, vise à la création d'une base de données en médecine générale en Haute-Normandie et à la création d'une plateforme de connexion pour les professionnels de santé, afin de permettre une information du grand public et un retour aux professionnels sur leurs pratiques.

## **EDSaNCoh**

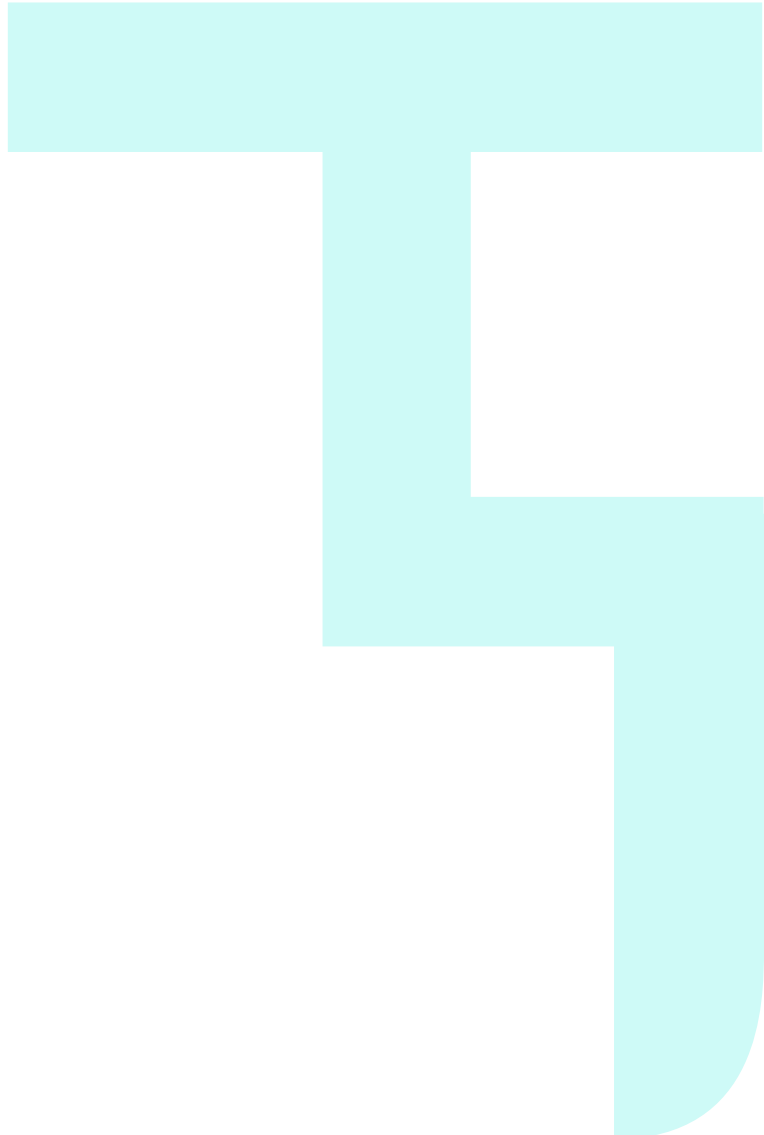
Une cohorte est un groupe de patients aux caractéristiques communes, suivis individuellement et de manière prospective, en recueillant des données de santé. Le CHU de Rouen a pour objectif de développer EDSaNCoh, une plateforme permettant de constituer des cohortes qui mutualisent les données de l'entrepôt de données de santé du CHU de Rouen, de l'Assurance Maladie et de questionnaires adressés aux patients.

## Anthropometer3DNet

Porté par le Centre de Lutte contre le Cancer Henri Becquerel de Rouen, le projet Anthropometer3DNet a pour objectif de créer une plateforme en ligne d'analyse automatique par intelligence artificielle de la composition corporelle de patients pris en charge pour un cancer, à partir de scanners réalisés en routine clinique. L'analyse des résultats peut donner un nouvel aperçu des facteurs de risque importants et améliorer le pronostic pour les patients.

## H3DMED

Le Centre de Lutte contre le Cancer François Baclesse de Caen s'est spécialisé depuis 20 ans dans la curiethérapie personnalisée, en particulier pour le traitement de cancers en gynécologie et de la peau. L'objectif du projet est de développer un algorithme d'IA approprié afin d'assister le médecin dans le processus de conception d'applicateurs personnalisés en curiethérapie.



# 6 PROJETS RETENUS DANS LE CADRE DE L'AAP DATALAB NORMANDIE

# 1 AAP AVEC DATALAB NORMANDIE "DONNÉES ET TECHNOLOGIES ASSOCIÉES, APPLIQUÉES À LA SANTÉ"

## Le partenariat avec AP-HP se développe autour de 8 projets pilotes

Amorcé en 2020, le partenariat entre l'AP-HP et le HDH s'est développé en 2021 avec des réalisations concrètes sur 8 projets pilotes et le lancement d'initiatives stratégiques conjointes.

Sur ces 8 projets, 7 nécessitaient une autorisation de la CNIL pour opérer. 6 ont d'ores et déjà obtenu un avis favorable du CESREES et le 8<sup>ème</sup> est en cours d'instruction.

PROJET	PORTEUR & PARTENAIRES	OBJECTIF
EMIR-Algo	AP-HP	Étude visant à développer et à valider des algorithmes de ciblage de personnes atteintes des différentes maladies auto-immunes et auto-inflammatoires systémiques (MAIS) rares à partir des données de la BNDMR chaînées aux données du SNDS.
BACTHUB	AP-HP, INSERM	Etablir le lien entre exposition aux antibiotiques et la survenue de bactériémies à bactéries résistantes.
DROMOS	AP-HP	Enrichir la Banque Nationale des Données de Maladies Rares avec les données de la base principale du SNDS pour mettre en place des indicateurs de santé publique.
AI Dream - STOIC	AP-HP, Therapanacea, GE	Développer des algorithmes d'apprentissage permettant le diagnostic de la COVID-19 et l'évaluation de sa sévérité à partir de scanners thoraciques.
NHANCE	AP-HP, INRIA	Développer et entraîner un algorithme d'IA détectant tous les organes de l'abdomen et leurs principales lésions.
DIGITAL HPV	AP-HP, INSERM	Déterminer le statut HPV sur les cancers épidermoïdes des voies aérodigestives supérieures et prédire la réponse aux thérapies.
EPIFRACTAL	AP-HP	Évaluer l'efficacité de la filière Fracture et proposer un modèle de sélection automatique des patients pour une cohorte à partir des comptes-rendus médicaux.
DAICAP	AP-HP, Inria, Incepto	Développer un algorithme qui permettra l'obtention d'un compte rendu d'IRM standardisé avec une prédiction de l'agressivité tumorale dans le cas des cancers de la prostate.



2021 a également été marquée par l'engagement de l'AP-HP à inscrire deux bases au catalogue du HDH : celle de la BNDMR qui sera chaînée au SNDS et la base STOIC. Le chaînage de la BNDMR avec le SNDS présente un enjeu majeur. Elle est en effet susceptible de multiplier les possibilités de recherches et donc, à terme, d'améliorer la prise en charge des patients atteints de maladies rares, dont l'errance diagnostique peut durer plusieurs années.

STOIC est une base COVID-19 réunissant un grand nombre d'images de cage thoracique de patients atteints de la COVID-19, à différents stades. Associée à des annotations médicales, cette base ouvre le champ des possibilités en termes d'études sur la COVID.

D'autres initiatives stratégiques sont à l'étude, parmi lesquelles une démarche coordonnée pour mettre les données de santé au format OMOP-CDM et ainsi pouvoir les partager à l'international. Par ailleurs, des réflexions sont en cours pour définir les conditions de partage et de valorisation des données qui seront inscrites au catalogue du HDH, à la fois sur le plan financier et scientifique.

## **Le HDH, membre fondateur de PariSanté Campus**

Annoncé le 4 décembre 2020 par le Président de la République et placé sous la responsabilité du ministère des Solidarités et de la Santé et du ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, le programme PariSanté Campus a pour ambition de créer un pôle d'excellence de recherche, d'innovation, de formation et d'entrepreneuriat unique au monde, sur le thème de la santé numérique. Le HDH en est membre fondateur, aux côtés de l'INSERM, l'Inria, l'Université PSL et l'Agence du numérique en santé. D'ici à 2028, ce campus de rang mondial va s'installer sur le site de l'ancien Hôpital du Val de Grâce à Paris. En attendant, un lieu préfigurateur a été inauguré en présence du Premier ministre le 14 décembre dernier. La force de PariSanté Campus sera de donner l'impulsion nécessaire pour encourager les rencontres entre opérateurs et acteurs, dans un lieu ouvert. En étant présent sur ce site, le HDH pourra décupler sa capacité à déployer des projets communs impactants, identifiables sur la scène internationale. Il a d'ores et déjà participé à la construction d'une offre spécifique à destination des startups sélectionnées par AMI qui viendront sur le site.

**5** MEMBRES FONDATEURS  
DE PARISANTÉ CAMPUS,  
DONT LE HDH



## **Le HDH membre fondateur de l'association Filière Intelligence Artificielle et Cancer (FIAC)**

En 2021, le HDH a rejoint les 10 autres membres fondateurs qui composent la FIAC. Une association qui fédère les acteurs publics et privés au sein d'une initiative nationale ambitionnant de profiter à l'ensemble de l'écosystème d'innovation en oncologie et de permettre le développement des solutions thérapeutiques et diagnostiques innovantes adressant des questions sans réponse.

**11**  
**MEMBRES COMPOSENT  
L'ASSOCIATION FIAC, DONT  
LE HEALTH DATA HUB**

## **Service de requête à la demande**

Le HDH a lancé, en 2021, une expérimentation visant à désigner un outil de "requête à la demande" destiné à des acteurs peu habitués aux données de santé. Ils pourront ainsi disposer d'éléments chiffrés issus du SNDS non disponibles en open data, sans aucune formalité réglementaire. Les éléments produits et la documentation associée seront ensuite accessibles à tous.

Pour préciser le périmètre de cette offre, des travaux ont été initiés en 2021 avec différentes associations de patients telles que France Vasculaires, l'association Journée Nationale de l'Audition ou l'Association des Porteurs de Dispositifs Electriques Cardiaques. Des chercheurs en sciences humaines et sociales ainsi que des représentants de journalistes comme le Quotidien du médecin ont aussi participé à cette réflexion. Ces travaux se poursuivront en 2022 pour aboutir à une ouverture du service au cours de l'année.



## ENJEU 3

# POSITIONNER LA FRANCE COMME UN LEADER DANS L'USAGE DES DONNÉES DE SANTÉ

Le HDH est un service à destination de l'écosystème, des acteurs à l'origine de la collecte de données, aussi bien que des utilisateurs de données et de la société civile. Pour atteindre son plein potentiel, le HDH doit être au rendez-vous des attentes d'un écosystème qui évolue désormais à une échelle européenne voire internationale.

### ► Action 3.1 – Alimenter la communauté scientifique internationale

En janvier 2021, un Conseil scientifique consultatif a été créé avec pour vocation d'accompagner la Direction du HDH à sa demande. Sa mission est de l'éclairer dans ses décisions pour tous les aspects scientifiques qui influent sur son développement. Ses membres se réunissent une fois par an en séance plénière, ainsi que de manière ponctuelle, selon les besoins du HDH. Pour aborder les sujets pouvant être très divers, y siègent 28 personnalités issues de nombreuses disciplines telles que médecine, pharmacie, génétique, épidémiologie, statistique, économie de la santé, juridique ou éthique.

Dès sa création, le conseil scientifique a donc été invité à rendre des avis sur des sujets d'intérêts au cœur de la stratégie de développement du Health Data Hub. Il a notamment instruit les décisions de la direction sur l'orientation du programme de financement de bourses postdoctorales, la construction d'un partenariat à fort impact avec Unicancer, l'intérêt scientifique des bases à ajouter au catalogue, ou encore les perspectives d'implication du HDH à l'échelle européenne.

### Le programme de bourses postdoctorales

Dans la ligne de son objectif d'alimenter la communauté scientifique sur le sujet des données de santé, le HDH a souhaité créer en 2021 un programme de financement de bourses postdoctorales. La direction a sollicité l'avis des membres du conseil sur le format du programme de financement ainsi que sur la pertinence et l'intérêt scientifique de sujets pré-identifiés. Ce programme de financement, ouvert à tous les laboratoires de recherche académique en France et à l'étranger, prévoit pour les bénéficiaires un soutien financier et humain en collaboration avec le laboratoire de recherche qui accueille le postdoctorant. Afin de contribuer à la promotion de l'usage des données à des fins d'intérêt public, les chercheurs bénéficiaires du programme de financement de bourse postdoctorale s'engagent à rendre disponible de manière ouverte à l'ensemble de la communauté scientifique le résultat des travaux, y compris les algorithmes développés, ainsi qu'à participer à une démarche de diffusion de leurs travaux dans le cadre d'événements scientifiques.

Trois sujets ont été choisis pour ces premières bourses postdoctorales :

- ▶ Étude comparative internationale de l'impact socio-économique, social et éthique de l'ouverture des données de santé,
- ▶ Visualisation des parcours de soins dans la base principale du SNDS.
- ▶ Implication des textes européens sur le HDH et propositions concrètes en matière de droits

### La mise au catalogue

L'ambition du catalogue est de rendre plus facile d'accès pour la recherche en santé des bases de données d'intérêt en les répliquant sur la plateforme technologique du HDH. Le catalogue dans sa première version comprendra une liste de bases qui seront chaînées à la base principale du SNDS quand cela est possible. En mars 2021, le conseil scientifique du HDH a été saisi pour avis sur l'intérêt des bases à inclure dans la première version du catalogue. L'objectif était de nourrir la réflexion du comité stratégique du SNDS en fournissant un avis scientifique sur les bases proposées en l'absence d'un processus bien établi, le comité stratégique ayant été créé en juillet dernier. Il s'est donc prononcé sur l'intérêt scientifique de chaque base en évaluant l'originalité au niveau national et international, la pertinence et la qualité de la base ainsi que les questions de recherche stratégiques qu'elle permettrait d'adresser, avec ou sans chaînage aux données de l'assurance maladie. Le conseil a évalué environ une trentaine de bases, 18 ont été retenues pour cette première version du catalogue.

### Le programme Unibase

Le programme UNIBASE est au cœur du partenariat avec Unicancer. En février 2021, le conseil a été appelé à fournir un avis sur l'intérêt scientifique des bases proposées par les Centres de lutte contre le cancer (CLCC) destinées à être partagées à terme dans le catalogue du HDH. Sur la base des éléments fournis par Unicancer quant aux types de données disponibles, le conseil a évalué les éléments suivants :

- ▶ l'impact potentiel des sujets proposés en termes de mobilisation des données pour la recherche académique et les soins des patients atteints de cancer,
- ▶ l'intérêt du chaînage des bases de données proposées à la base principale du SNDS,
- ▶ la pertinence et la réutilisabilité à des fins de recherche des bases proposées pour la mise au catalogue,
- ▶ la pertinence des thèmes proposés pour des projets aux finalités et aux méthodologies diverses,
- ▶ l'intérêt de la mise en qualité des données et de l'interopérabilité pour d'autres acteurs des données de santé.

Dans un deuxième temps, le conseil a rendu un avis sur les thèmes choisis pour le lancement d'un AMI à destination des CLCC permettant à trois projets pilotes d'être accompagnés par le HDH dès 2022.

## L'implication européenne du Health Data Hub

À la fin de l'automne 2021, un groupe de travail restreint du conseil a appuyé le HDH dans la construction d'une candidature à un appel à projets de la Commission européenne visant à construire et tester une première version d'Espace Européen des Données de Santé. Forts de leur expérience, les membres du groupe ont apporté un regard critique sur le choix des partenaires européens, les cas d'usage de recherche proposés, ainsi que l'architecture du projet.

## Un conseil scientifique ouvert à l'international

Le conseil scientifique du HDH a vocation à éclairer la direction en mobilisant également des sources d'inspiration à l'international. 7 membres internationaux ont donc rejoint en 2021 les 21 autres membres. D'une manière générale, l'ensemble de ces membres, qu'ils soient français ou étrangers, aident le HDH à identifier des partenariats hors de nos frontières ou de nouveaux outils et services. Ils contribuent à diffuser les savoirs de façon plus large au sein de la communauté scientifique mondiale et à intégrer le HDH dans le paysage européen des données de santé.

**28**  
**MEMBRES DANS LE  
CONSEIL SCIENTIFIQUE  
CONSULTATIF DONT  
7 INTERNATIONAUX**

## 6 nouveaux data challenges prévus en 2022

Le HDH organise régulièrement, avec ses partenaires, des événements fédérateurs ouverts à toute la communauté scientifique internationale. L'école d'hiver annuelle en est un exemple, tout comme les Data Challenges.

Le premier data challenge a été organisé en 2020 avec la Société Française de Pathologie, le Grand Défi "comment améliorer les diagnostics avec l'intelligence artificielle" et la BPI. Il a réuni près de 550 participants qui ont développé et testé des algorithmes d'IA pour détecter automatiquement des lésions pré-cancéreuses et cancéreuses pouvant survenir sur le col de l'utérus. Pour cela, ils se sont appuyés sur un jeu de données anonymes de près de 5 000 lames d'histopathologie (biopsies et conisations) en provenance d'une vingtaine de centres anatomopathologiques français volontaires, publics et privés. Le Health Data Hub a apporté une aide organisationnelle et logistique à la collecte de ces données.

A l'issue de l'événement, l'usage veut qu'on ouvre les données anonymisées et les algorithmes développés pour que la recherche puisse se poursuivre.

C'est pour encourager cette émulation que le HDH a renouvelé son partenariat avec le Grand Défi et la BPI et lancé en mars 2021 un appel à manifestation d'intérêt (AMI) destiné aux acteurs du monde médical. 6 data challenges ont été sélectionnés. Ils sont actuellement accompagnés par le HDH et soutenus financièrement par la BPI.

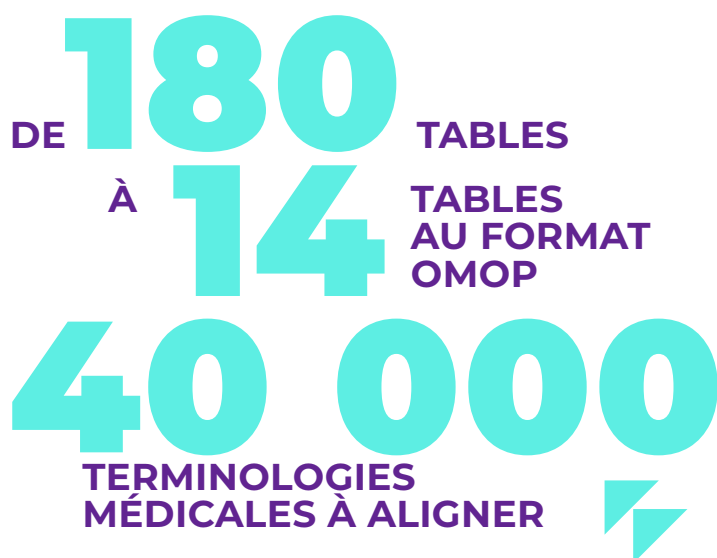
Ces initiatives portent sur des domaines médicaux divers et vont permettre d'apporter une véritable contribution sur ces questions :

- 1 Le Data Challenge DigiLUT, porté par l'Hôpital Foch, pour développer un algorithme capable de détecter un rejet de greffe à la suite d'une transplantation pulmonaire.
- 2 Le Data Challenge organisé par la Société Française de Chirurgie Thoracique et Cardiovasculaire (SFCTV) pour classifier les tumeurs pulmonaires à partir de données d'imagerie cellulaire.
- 3 Le D-IA-GNO-DENT, porté par les Hôpitaux Universitaires de Strasbourg, pour proposer des algorithmes de diagnostic automatique de maladies orales rares, à partir de photographies de dents.
- 4 Le Data Challenge organisé par la Société Française de Pathologie (SFP) et les dermatologues pour prédire le risque d'évolution métastatique de mélanomes.
- 5 L'Allergen Chip Challenge de la Société française d'allergologie (SFA) pour identifier les allergènes et associations d'allergènes en cause lors d'une manifestation allergique.
- 6 Le Data Challenge de la Société Française de Médecine Nucléaire (SFMN) pour développer un algorithme contribuant au diagnostic des syndromes parkinsoniens aux formes multiples.

▶ **Action 3.2 – Structurer et normaliser les données de santé pour faciliter leur usage dans le cadre de collaborations internationales**

**Vers une standardisation du SNDS pour favoriser sa réutilisation à des fins de recherche**

Le SNDS est l'une des bases de données médico-administratives les plus complètes au monde. Cependant, son utilisation nécessite d'être un expert de la base afin de comprendre le schéma de données, son potentiel, ses biais et ses limites. Mener des projets à l'échelle européenne avec le SNDS reste un véritable défi, par des schémas de données et des terminologies propres à chaque base de données / pays. Depuis 2020, le HDH se mobilise pour simplifier le schéma de données du SNDS et le faire évoluer vers un modèle de données commun OMOP-CDM afin qu'il soit plus facile à réutiliser dans des projets de recherche nationaux et européens. La transformation du SNDS au format OMOP-CDM permet de passer d'une base contenant plus de 180 tables à moins de 20 tables. Par ailleurs, le SNDS comporte plus de 70 000 terminologies médicales (CCAM, NABM, CIP, UCD, CSARR, LPP, etc.), des travaux d'alignement de ces terminologies au format standard sont en cours avec l'appui précieux d'internes en médecine.



En décembre 2021, le HDH a été lauréat pour la seconde fois d'un appel à projet EH DEN afin de continuer ces travaux de transformation, d'étendre le périmètre à tous les bénéficiaires et aux années antérieures à 2019.

▶ **Action 3.3 – Contribuer à la structuration du paysage européen autour de l'usage des données de santé**

Pour faciliter l'accès aux différents types de données disponibles dans ses Etats membres, la Commission européenne a fait du futur Espace Européen des Données de Santé (EHDS) l'une des priorités de sa politique de santé. Ce partage est en effet considéré comme fondamental pour développer la recherche et améliorer la santé des citoyens. Il permet, par exemple, de construire des algorithmes d'intelligence artificielle plus performants au service de la recherche ou de la performance industrielle, d'améliorer la gestion de pandémies telles que la COVID-19, ou encore d'étudier des maladies rares grâce à la mutualisation des données.

Plusieurs instruments de préfiguration du futur Espace Européen de Données de Santé ont déjà été mis en place par la Commission européenne depuis 2019. C'est le cas de TEHDaS (Towards a European Health Data Space), un programme de réflexion réunissant plus de 26 Etats membres pour lequel le Health Data Hub coordonne 5 entités affiliées : l'ANS, l'INSERM, les Hospices Civils de Lyon (HCL) et les universités Aix-Marseille et Toulouse III.

Plus récemment, le Health Data Hub a pris la tête d'un consortium qui a répondu à l'appel à candidature de la Commission européenne pour construire une version test du futur Espace Européen des Données de Santé. Ce consortium rassemble des plateformes nationales de données de santé, ainsi que des agences européennes

ou des infrastructures de recherche. Le projet vise à déployer un réseau de plateformes de données au sein de l'Union européenne.

Dans le cadre de l'utilisation des données en vie réelle, le Health Data Hub participe aussi au conseil scientifique de Darwin, le réseau d'analyse de données, d'expertise et de services, lancé par l'Agence européenne du médicament (EMA). Objectif : améliorer la prise de décision tout au long du cycle de vie des médicaments, grâce à l'exploitation de ces données.

Le Health Data Hub prend aussi part à un consortium mené par la startup Owkin en partenariat avec l'Inria, le Centre Léon Bérard, le CHU de Nantes, Apricity (France), TMF-EV, l'Université de médecine de Berlin, l'Université de Düsseldorf, BfArM (Allemagne), l'Université de technologie et d'économie de Budapest (Hongrie), l'Université de médecine de Graz (Autriche), l'Université de Gollnik (Slovénie) et la Fondation nationale italienne pour le cancer et l'Institut Jimenez Diaz (Espagne), en réponse à un appel à projets européen pour la mise en place d'un système d'apprentissage fédéré (federated learning) en Europe, en prenant en compte la confidentialité des données et dans lequel le Health Data Hub serait donc un nœud source de données. Le résultat sera annoncé en mars 2022.

### **De premiers résultats pour l'action conjointe TEHDaS**

Lancée en février 2021, l'action conjointe vers un "Espace Européen des Données de Santé" (TEHDaS) qui mobilise 25 pays, a pour objectif de soumettre, d'ici au 26 juillet 2022, des recommandations à la Commission Européenne pour le futur espace européen des données de santé. Pilotée par le HDH, la contribution française est particulièrement active sur l'engagement citoyen (groupe de travail 8) et la gouvernance (groupe de travail 5). Elle a posé ses premiers jalons en 2021, à travers des actions concrètes.

Une grande consultation en ligne a été lancée fin 2021 par le HDH, avec ses partenaires belge et anglais, pour mesurer la perception qu'ont les citoyens des données de santé et de leur utilisation. Disponible en français, anglais et néerlandais mais ouverte à tout citoyen européen, elle propose de l'information sur le partage et l'usage des données en Europe, des questionnaires individuels et un espace de discussion. Pour préparer cette e-consultation, une revue de littérature internationale de près de soixante-dix articles a été faite. Elle portait sur les mécanismes d'engagement citoyen quant à l'usage secondaire des données de santé en Europe et la réutilisation des données à des fins autres que le soin direct. Des associations d'usagers, des professionnels de santé, des acteurs de santé publics et privés, ainsi que des universitaires ont aussi été interrogés.

Pour travailler sur la future gouvernance du partage transfrontalier de données de santé, le Health Data Hub et ses partenaires finlandais, espagnol et néerlandais ont lancé un benchmark international permettant de comparer les procédures d'accès aux données de santé dans les quatre pays. Les premiers résultats ont mis en évidence les obstacles au partage transfrontalier des données de santé et les besoins juridiques spécifiques à la donnée de santé.

**9%**  
**DU BUDGET DE L'ACTION  
CONJOINTE EUROPÉENNE  
TEHDAS POUR LA FRANCE**

**60%**  
**FINANCÉS PAR  
LA COMMISSION  
EUROPÉENNE**





## Le HDH monte un consortium pour proposer une version test de l'EHDS à la Commission européenne

Le partage des données à l'échelle européenne est fondamental pour développer la recherche et améliorer la santé des citoyens. Il permet, par exemple, de construire des algorithmes d'intelligence artificielle plus performants, de mieux gérer des pandémies telles que la COVID-19, ou encore d'étudier des maladies rares.

C'est dans ce but que le HDH a monté un consortium pour répondre à l'appel à candidature de la Commission européenne, publié le 25 octobre 2021. Objectif : mettre en place une version test d'Espace Européen des Données de Santé (EHDS). Ce consortium regroupe une quinzaine d'acteurs européens majeurs, dont les plateformes nationales de données de santé de plusieurs Etats membres (Finlande, Norvège, Danemark, Allemagne, Belgique, Hongrie, Croatie), ainsi que des agences européennes comme l'Agence européenne du médicament (EMA) ou le Centre européen de prévention et de contrôle des maladies (ECDC) et des infrastructures de recherche (BBMRI, Elixir, eBrains, IACS).

Ensemble, ils construiront une infrastructure technologique en réseau qui connectera l'ensemble des partenaires fournisseurs de données (nœuds de données). Grâce à un portail unique, ce réseau permettra aux chercheurs souhaitant monter un projet d'interroger les catalogues de données de tous les nœuds, de demander l'accès à des données sélectionnées via un formulaire commun unique et, si nécessaire, de transférer des données en un seul endroit. L'idée est de rendre ce parcours aussi simple et fluide que possible pour les porteurs de projets.

Dans le dossier de candidature remis le 25 janvier dernier, sont également prévus une série de cas d'usage concrets permettant de tester le réseau

et de démontrer le bénéfice de cette réutilisation transnationale des données de santé pour la recherche, l'innovation, les politiques de santé, la réglementation et la médecine personnalisée.

Les résultats de l'appel à projet sont attendus d'ici à l'été 2022 et les travaux du consortium gagnant débiteront en septembre. Ils serviront de base au futur Espace Européen des Données de Santé.

**5** STRUCTURES DE RECHERCHE (BBMRI, ELIXIR), ASSOCIATIONS (EUPHA) OU AGENCES EUROPEENNES (EMA, ECDC) PARTENAIRES

**10** PAYS IMPLIQUÉS

### ► Action 3.4 – Promouvoir le Health Data Hub à l'international

L'une des missions du HDH est de contribuer au rayonnement international de la France quant à l'utilisation des données de santé, en valorisant notamment l'excellence des acteurs français. C'est dans ce but que le Health Data Hub a mis en place une newsletter internationale diffusée à plus de 800 contacts qualifiés dans le monde entier et participé à de nombreux événements de renommée internationale. À titre d'exemple, le HDH a organisé un séminaire commun avec le MIT et l'Académie Nationale de Médecine sur l'intelligence artificielle dans le monde médical et a été présent virtuellement sur le Pavillon français de l'Exposition universelle de Dubaï.

Le HDH a également signé un contrat de partenariat avec son homologue finlandais, Findata, et initié des échanges en vue de mettre en place des relations partenariales avec de nombreux acteurs en Europe comme à l'international (Canada, Israël, Royaume-Uni ...). L'objectif de ces partenariats est d'enclencher des réflexions communes sur l'usage des données à des fins de recherche ou d'innovation et de créer des projets impactants rassemblant des données et des chercheurs français comme étrangers.

L'année 2021 a par ailleurs été à nouveau marquée par l'organisation d'une École d'hiver dédiée à l'intelligence artificielle en santé, co-organisée avec 3 Instituts Interdisciplinaires en intelligence artificielle (Prairie, 3IA Côte d'Azur et MIAI Grenoble). 428 personnes, dont 56% d'étudiants y ont participé. Pendant cinq jours, des chercheurs de renommée internationale ont présenté leurs avancées et animé des sessions pratiques. Les conférences d'ouverture ont été réalisées par Yoshua Bengio (A.M. Turing Award, directeur scientifique de Mila et IVADO, professeur à l'Université de Montréal et codirecteur du programme d'apprentissage automatique et biologique au CIFAR au Canada), I. Glenn Cohen (professeur et doyen associé du Petrie-Flom Center for Health Law and Policy, Biotechnology and Bioethics à l'Université de Harvard) et Ran Balicer (professeur, directeur fondateur du Clalit Research Institute, directeur de l'innovation chez Clalit Health Services et membre du comité national d'experts Covid-19 pour le ministère de la santé en Israël).

**438**  
**PARTICIPANTS À**  
**LA WINTER SCHOOL**  
**"AI4HEALTH" 2021**

**50**

## **Le HDH s'associe à l'Académie nationale de médecine et au MIT pour un symposium sur l'IA**

Comment allier Intelligence Artificielle et pratique clinique ? Quelles réponses apporter aux questions éthiques et réglementaires que soulève l'IA ? Le rapprochement entre l'Intelligence Artificielle et la médecine, favorisé par les avancées technologiques, soulève encore beaucoup de questions. C'est pour tenter d'y répondre que le HDH, l'Académie nationale de médecine et le MIT, ont organisé un symposium, le 5 et 6 mai 2021. Plus de 700 participants y ont assisté en ligne. Ils ont pu échanger avec des cliniciens et des experts français et américains, à la croisée de l'IA et de la médecine, pour mieux comprendre les différentes approches et faciliter la collaboration transatlantique.

Dédiée aux avancées théoriques et techniques en intelligence artificielle appliquée à la médecine, la première journée du symposium a mis à l'honneur les apports de l'imagerie pour guider la pratique médicale dans des domaines comme le traitement de la Covid-19, les problèmes coronariens, la neurologie ou l'oncologie. L'occasion pour Nathalie Lassau, Professeur de radiologie à l'Institut Gustave Roussy (IGR), et ses confrères d'échanger sur l'importance des consortiums internationaux pour rassembler un grand nombre d'images annotées. Daniela Rus, professeure d'informatique au MIT CSAIL, a, quant à elle, présenté l'état de l'art en robotique pour la santé. La dernière session était consacrée à l'apport de l'intelligence artificielle pour l'analyse d'échantillons de cellules ou de tissus. Cécile Badoual, professeure d'anatomie et cytologie pathologiques à l'Université Paris Descartes, y a présenté des résultats probants sur la classification des cancers et engagé une discussion sur les données de microscopie et l'usage des modèles prédictifs, en hématologie notamment.

Ouverte par le Pr. Bernard Nordlinger, membre de l'Académie nationale de médecine, et Cédric Villani, député, la seconde journée a été consacrée aux questions éthiques et réglementaires autour de l'IA. L'accent a été mis sur la différence d'approche aux Etats-Unis et en Europe, compte-tenu de leur cadre réglementaire respectif. Dans une première session sur la régulation de la collecte et l'usage des données, Danny Weitzner, professeur d'informatique au MIT, l'avocate Jeanne Bossi-Malafosse et Nicholson Price, professeur de droit à l'Université de Michigan, ont montré pourquoi la donnée de santé se distingue des autres types de données et constitue un enjeu majeur. La façon dont la législation s'adapte à cet enjeu de chaque côté de l'Atlantique a également été évoquée, tout comme les insuffisances de la loi dans certains cas. Une loi trop ou pas assez protectrice selon les usages, source de risque pour la protection des données ou source de blocage pour la recherche. Lors de la deuxième session, Glenn Cohen, professeur de droit à l'université de Harvard et Claude Kirchner, directeur du Comité National Pilote d'Éthique du Numérique (CNPEN) ont abordé les enjeux éthiques de l'IA en médecine. Lors de ces échanges, Glenn Cohen a notamment appelé les décideurs à se pencher sur les questions éthiques autour des biais que peuvent comporter des solutions d'intelligence artificielle tel que l'apprentissage adaptatif.

Enfin, Emmanuel Bacry, directeur scientifique du Health Data Hub et directeur de recherche au CNRS ainsi que David Sontag, professeur d'informatique au MIT CSAIL, ont présenté les modalités d'accès aux données de santé en France comme aux Etats-Unis et les spécificités liées notamment aux grands programmes nationaux. Une discussion qui a mis en évidence des problématiques communes autour de la mise à niveau des données, leur format, l'interopérabilité des bases, l'ouverture des données et la relation au patient, notamment dans le cadre de la réutilisation des données de santé.

**704**  
**PARTICIPANTS À**  
**L'ÉVÈNEMENT "IA ET LA**  
**MÉDECINE : PROMESSES**  
**ET LIMITES"**



## ENJEU 4

# GARANTIR LA PARTICIPATION DE LA SOCIÉTÉ CIVILE ET PROMOUVOIR LES USAGES DES DONNÉES DE SANTÉ

Pour que le HDH atteigne son plein potentiel, il est essentiel que son objectif soit compris par l'ensemble des parties prenantes et en particulier la société civile. C'est dans ce but que le HDH a mis en place une direction citoyenne pour écouter la société civile, produire des contenus pertinents au regard de leurs attentes, et concevoir des partenariats.

### ► Action 4.1 – Écouter la société civile et comprendre ses attentes pour établir une relation de confiance

La direction citoyenne mène des actions pour comprendre les perceptions des citoyens, appréhender leur niveau de connaissance et recueillir leurs besoins et attentes quant aux données de santé. Pour ce faire, elle s'appuie sur les méthodes sociologiques, produit des revues de littérature, anime des groupes de travail et propose des consultations citoyennes. En 2021, plusieurs groupes de travail se sont tenus, notamment avec des étudiants en pharmacie et de la cohorte i-Share, pour les faire réagir à des propositions de contenus calibrés répondant à leurs attentes. Une revue de littérature, dans le cadre de l'action conjointe "Vers un Espace Européen des données de santé", a été publiée et un débat en ligne, au niveau européen a été lancé. Toutes ces démarches se poursuivront en 2022, afin d'alimenter l'observatoire des données de santé du Health Data Hub, et dans le cadre des engagements du gouvernement ouvert.

### Un grand débat en ligne pour sonder les citoyens européens sur les données de santé

Pour pouvoir utiliser les données de santé dans la recherche, la santé publique et la prévention, il est indispensable d'y associer les citoyens. C'est dans ce but que le HDH et ses partenaires belge et anglais ont lancé une grande consultation en ligne, de fin décembre 2021 à avril 2022. Une première soutenue par la Commission européenne, dans le cadre de l'action conjointe « Vers un espace européen des données de santé » (TEHDAS).

Son objectif premier est de sensibiliser les citoyens sur la façon dont leurs données sont utilisées, dans quel but et dans quelles conditions, ainsi que sur le rôle potentiel qu'ils peuvent jouer. Mais au-delà, c'est aussi l'occasion d'engager un échange avec eux, comme entre eux. Les enseignements tirés de ce débat alimenteront les propositions soumises à la Commission européenne en 2022 pour la mise en œuvre d'un espace commun européen des données de santé.

Pour ce faire, un site dédié a été créé, avec un espace pédagogique pour que les citoyens puissent s'informer, un test ludique pour susciter leur curiosité et recueillir leur perception et un espace de concertation sous forme de questions ouvertes auxquelles ils peuvent répondre avec des contributions, des commentaires, des « like » et des partages.

Proposée en plusieurs langues, cette plateforme est ouverte à tous les citoyens européens. Le HDH et ses partenaires s'engagent à y informer les citoyens en toute neutralité et transparence, à garantir la sécurité et la confidentialité des contributions, à communiquer sur la façon dont elles seront valorisées et, bien sûr, à tenir compte des avis émis pour rédiger les recommandations.

Pour préparer cette consultation, le HDH et ses partenaires ont mené une revue de littérature balayant plus de 70 publications européennes et rencontré une soixantaine d'acteurs concernés par l'usage secondaire des données de santé.

## ► **Action 4.2 – Expliquer et diffuser les engagements du Health Data Hub autour du respect des procédures, du contrôle des usages, de la maîtrise des données, de l'explicitation et l'accompagnement dans l'exercice des droits**

Après la publication du décret relatif au SNDS, qui précise les nouvelles modalités d'exercice des droits des personnes, le HDH a mis à jour son site internet afin d'expliquer ce que prévoit le nouveau texte. Le HDH contribue également à la réflexion sur la manière d'informer l'ensemble de la population française quant à l'existence du SNDS et des droits qui y sont associés.

Afin de mettre en œuvre les droits des personnes, à l'approche de la réception de la base principale du SNDS, le HDH a lancé la conception d'un outil de gestion des droits qui reposerait sur un formulaire accessible en ligne, permettant aux citoyens d'exprimer leur choix une fois pour toutes, en un lieu unique (voir aussi : action 1.2 – Installer une procédure nationale automatisée pour faciliter la sélection et la remontée des données au catalogue et leur appariement, p 23).

Par ailleurs, le HDH a produit des documents pour les responsables de données des futures bases du catalogue et les responsables de traitement des projets mobilisant des données de santé. Objectif : les aider à fournir une information exhaustive, compréhensible et conforme au RGPD, comme au décret SNDS. Ces documents pédagogiques et vulgarisés ont été conçus par des juristes et des citoyens pour que les personnes concernées comprennent leur contenu.

Enfin, en rejoignant le réseau de l'Association Française des Correspondants à la protection des Données à caractère Personnel (AFCDP), le HDH contribue à la communauté d'échanges sur le partage des données. Il y a travaillé sur des livrables, au sein des groupes de travail sur la réutilisation des données de santé et sur l'information Facile A Lire et à Comprendre (FALC).

### ▶ **Action 4.3 – Mettre en place une politique de communication et de diffusion de la culture de la donnée de santé**

Le Health Data Hub contribue à vulgariser et rendre accessible à tous, les enjeux relatifs à la réutilisation des données de santé. Pour ce faire, il conduit des actions de formations. L'une d'elles s'adresse aux citoyens pour les sensibiliser au Système National des Données de Santé (SNDS), avec le soutien de France Assos Santé et de l'Assurance Maladie.

Le Health Data Hub s'attache aussi à diffuser de l'information sur les projets qu'il accompagne, leurs finalités et leurs résultats. Sa direction citoyenne s'engage pour co-construire ces informations et mettre en œuvre la démarche Facile A Lire et à Comprendre (FALC). Des actions d'information avec des établissements de santé et des médiateurs numériques ont par ailleurs été initiées.

Enfin, la direction citoyenne met en œuvre des partenariats avec la société civile et les associations pour leur donner l'occasion de devenir acteurs des données de santé. C'est ainsi qu'elle a continué à accompagner, en 2021, France Assos Santé et ses associations membres (dont France Handicap, Unaf, FFAAIR, AFA, EFAPPE, Fibromyalgie France, France Rein, Unafam, France Alzheimer, Aides, FFD, FNAR, Ligue Nationale contre le cancer) pour réaliser l'étude VivreCovid19. Le HDH les accompagne dans la coordination de l'étude, la recherche d'autres partenariats avec des acteurs publics pour exploiter les données et le développement d'une doctrine de transparence et d'open source.

D'autres partenariats ont été signés, notamment avec la FFD, qui est la première association à avoir déposé un projet avec les données du SNDS, mais aussi avec APODEC, JNA, France Vasculaires, TousPartenairesCovid et l'ANEPF. De nouveaux partenariats sont en discussion.

## **16 FORMATIONS PROPOSÉES SUR LE SNDS**

### **Vivre Covid 19 : une étude associative ouverte à la recherche et la communauté scientifique**

Lancée en mai 2020, 6 semaines seulement après le début du confinement, VivreCovid19 est la première grande enquête de France Assos Santé (FAS), en partenariat avec le HDH. Plus de 5400 inscrits s'y expriment chaque mois sur leur niveau d'anxiété et leurs difficultés ou non d'accès aux soins durant la crise sanitaire. Cette étude, qui se termine en mai 2022, est ouverte à la recherche, ainsi qu'à toute la communauté scientifique.

Elle se démarque des études standards parce qu'elle a démarré dès le premier confinement et parce qu'elle interroge aussi bien des patients chroniques ou en situation de handicap, qu'un échantillon témoin sans pathologie. Avec plus de 3,5 millions de données collectées d'ici à la fin de l'étude, FAS va pouvoir tirer des enseignements solides. Il a d'ailleurs fait appel à des partenaires publics de recherche afin d'exploiter les données le plus largement et finement possible.

Son protocole a été mis en ligne sur le site de l'étude, pour que tous les partenaires associatifs et la communauté scientifique s'approprient bien son contexte, ses objectifs, ses méthodes, ses moyens, ainsi que ses mesures de sécurité, d'information et ses analyses. Cette publication en open access a aussi pour objectif de démontrer la robustesse scientifique et méthodologique de l'étude, la pertinence des prestataires choisis, ainsi que la capacité du monde associatif à adopter les meilleures pratiques et standards dans l'intérêt des patients.

Une partie des données sera également accessible en accès libre sous forme de données anonymisées, pour augmenter l'opportunité de nouvelles découvertes. Chaque structure ayant recours à un jeu de données anonymisées devra cependant s'engager notamment à les utiliser à des fins de santé publique et à ne pas les commercialiser.

Le HDH accompagne FAS tout au long du projet. Il a par exemple contribué à animer des ateliers et à concevoir puis mettre en œuvre la doctrine éthique.

**+5000**  
PARTICIPANTS À  
L'ÉTUDE VIVRE  
COVID-19

#### ▶ **Action 4.4 – Développer la mise à disposition de données strictement anonymes en libre accès**

La réutilisation des données de santé s'est aujourd'hui démocratisée grâce à un accès totalement ouvert à certaines données. Pour cela et compte tenu de la sensibilité de ces données, il est nécessaire de produire des données fictives, ou des données réelles complètement anonymisées. En reflétant certaines propriétés des données réelles, ces données fictives facilitent l'appropriation et stimulent les usages des données de santé. C'est pour créer ce type de données à grande échelle que le HDH a sensiblement amélioré le générateur de données synthétiques, initié en 2020 et proposé en accès libre sur Gitlab. Un partenariat a en outre été noué avec l'INRIA et avec l'équipe REPERES, pour générer des données synthétiques représentatives de la base principale du SNDS. Ces travaux permettront à terme au HDH de rendre disponibles, avec son catalogue de métadonnées, des données de synthèse pour accompagner chacune des bases de données par ailleurs disponibles de manière pseudonymisée dans sa plateforme hautement sécurisée.

Le HDH promeut également l'ouverture des données réelles lorsqu'elles peuvent être parfaitement anonymisées. Ainsi, grâce au déploiement de l'outil open source Dataverse, les utilisateurs pourront en 2022 accéder aux données ouvertes dans le cadre des data challenges accompagnés.

**5000**  
LAMES NUMÉRISÉES DE  
BIOPSIE D'UTÉRUS EN COURS  
DE MISE À DISPOSITION  
EN OPEN DATA



3

# BILAN DE L'ACCÈS AUX DONNÉES DE SANTÉ





**Pour améliorer la qualité des soins, les données de santé doivent être accessibles au plus grand nombre. Mais cela doit se faire dans le respect des droits des personnes concernées.**

**En effet, ces données, même pseudonymisées, restent des données personnelles à moins qu'elles ne soient parfaitement anonymisées. Dans ce dernier cas, leur utilisation est libre et aucune démarche n'est à entreprendre. À défaut, l'accès à ces données est réglementé.**

## **ACCÈS AUX DONNÉES ET PERSPECTIVES**

### **Un accès réglementé**

Dans la plupart des cas, il est nécessaire d'obtenir une autorisation auprès de la CNIL pour traiter des données personnelles de santé. En amont de cette autorisation, la procédure prévoit l'intervention de deux institutions :

- ▶ Le HDH, point d'entrée unique des projets mobilisant des données de santé. Près de **380** projets y ont été déposés en 2021.
- ▶ Le CESREES rend un avis sur la finalité, la pertinence et la qualité scientifique du projet

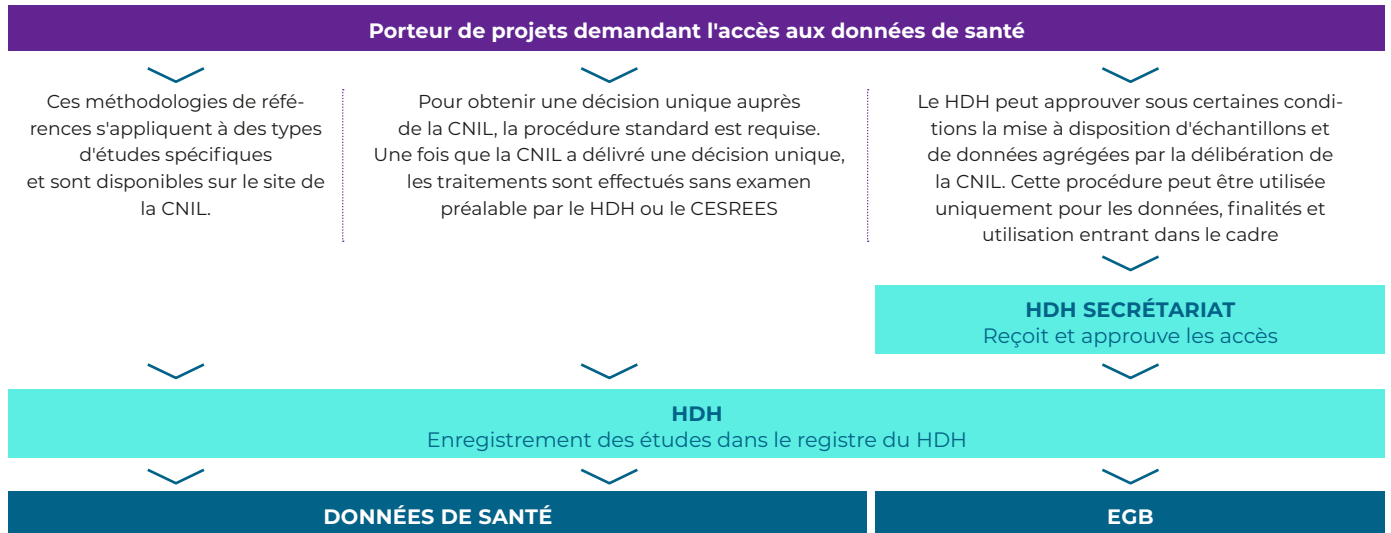
Afin de faciliter l'accès aux données et permettre aux porteurs de projet de réaliser leur traitement dans les meilleurs délais, le temps imparti à chacune de ces institutions est encadré. Le HDH dispose ainsi de 7 jours pour transmettre le projet complet au CESREES. Ce dernier a ensuite un mois pour rendre son avis éclairé. Si l'avis est favorable, la CNIL dispose de 2 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.

### 3 PROCÉDURES SIMPLIFIÉES CORRESPONDANT À DES USAGES ET DES DONNÉES

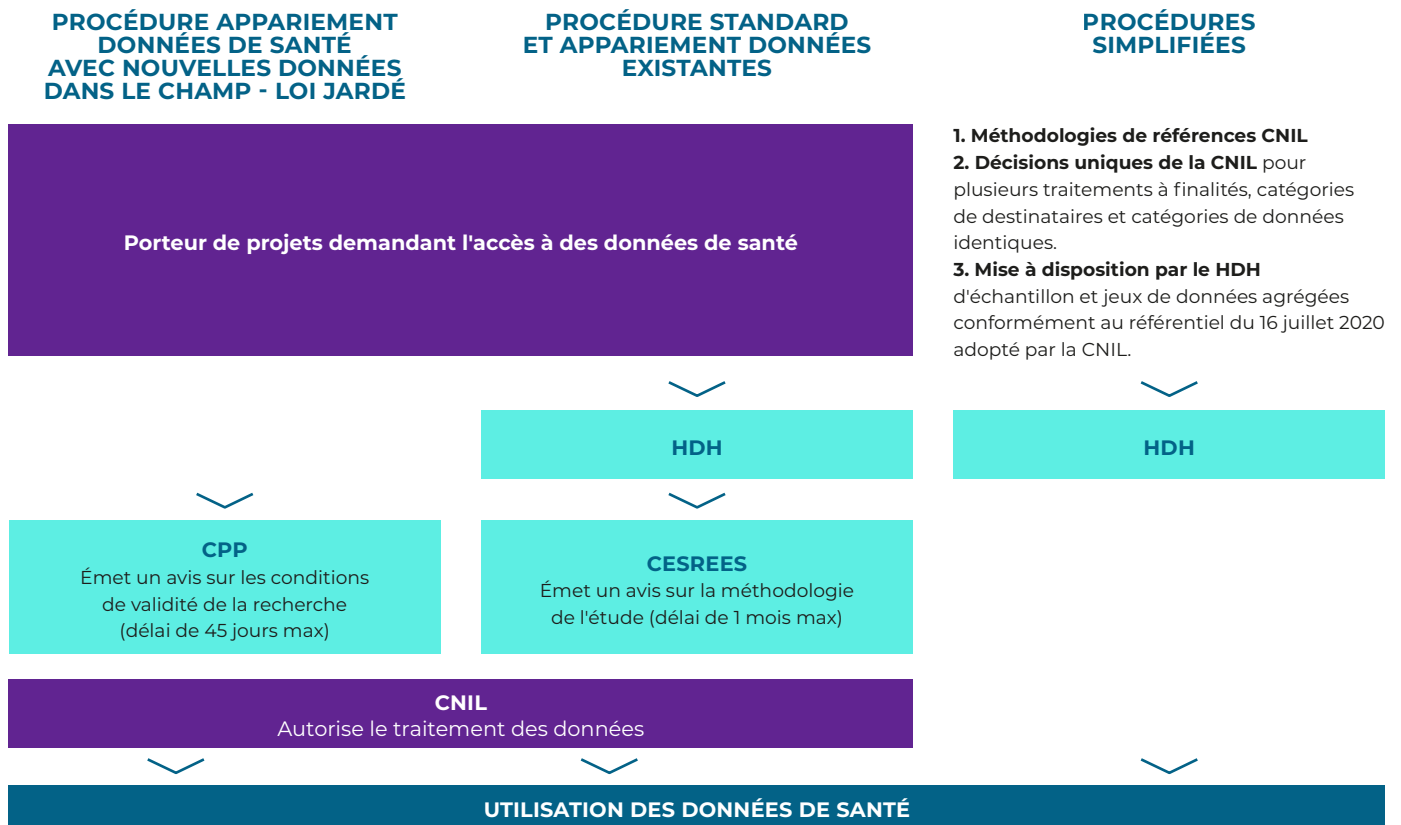
**1 PROCÉDURE DES MÉTHODOLOGIES DE RÉFÉRENCES MENTIONNÉE À L'ARTICLE 73 DE LA LIL**

**2 PROCÉDURE DES DÉCISIONS UNIQUES MENTIONNÉE À L'ARTICLE 66 DE LA LIL**

**3 PROCÉDURE D'ÉCHANTILLON ET DONNÉES AGRÉGÉS MENTIONNÉE AU II DE L'ARTICLE 66 DE LA LIL**



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



Il existe des procédures d'accès dites simplifiées, dans le cadre desquelles l'autorisation de la CNIL n'est pas requise. C'est notamment le cas si le traitement envisagé est conforme à une méthodologie de référence (dite MR). Afin de promouvoir l'accès aux données et favoriser leur usage, le HDH contribue au déploiement de ce type de procédures en étroite collaboration avec la CNIL. En 2021, il s'est ainsi prononcé sur le référentiel relatif aux entrepôts de données de santé. Grâce à cette procédure simplifiée, un responsable de traitement voulant mettre en œuvre un entrepôt de données dans un établissement de santé n'a plus à solliciter une autorisation de la Commission dès lors que son projet est conforme au référentiel.

Enfin, le HDH peut promouvoir aussi un accès plus simple à des jeux de données agrégées ou des échantillons, dans des conditions préalablement autorisées par la CNIL.

### L'accès en quelques chiffres (chiffres 2021)

<b>Nombre total de projets déposés au guichet en 2021</b>	<b>380</b>
<b>Nombre de projets déposés au guichet en vue d'une autorisation CNIL</b>	<b>366</b>
dont transmis au CESREES	<b>279</b>
dont réorientés MR ou CPP	<b>56</b>
dont non complets	<b>31</b>
<b>Nombre de demande d'accès à l'EGB simplifié</b>	<b>14</b>
<b>Nombre de projets relevant d'une MR (méthodologie de référence)</b>	<b>1411</b>
MR004	<b>1279</b>
MR005	<b>54</b>
MR006	<b>78</b>



## Une offre de services destinée aux porteurs de projet

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, avant même la naissance du HDH, la DREES a élaboré une collection d'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. En 2021, le HDH y a apporté les évolutions suivantes :

- ▶ Le kit pédagogique, permettant aux porteurs de projet de bien comprendre quelles pièces sont attendues dans le cadre d'une demande d'autorisation, a été actualisé. Ce kit comprend plusieurs éléments dont un guide pour bien remplir le protocole scientifique, un guide pour bien remplir l'expression des besoins lorsque des données de la base principale du SNDS sont requises et un guide relatif à l'information des personnes exigée par le RGPD. Ces éléments ont été conçus en lien avec le CESREES et la CNAM.
- ▶ Un guide présentant la réglementation française en matière d'accès aux données de santé a été créé. Il permet aux acteurs de s'y retrouver par le biais de critères bien identifiés. Ce guide a été traduit en anglais et a fait l'objet de formations régulières. Un guide spécifique a en outre été conçu pour aider ceux qui veulent accéder à l'Échantillon Généraliste de Bénéficiaires.
- ▶ L'offre de formation a été élargie, avec notamment une formation citoyenne aux données de la base principale du SNDS. Destinée à tous ceux qui se posent des questions sur les données de santé et comment elles sont utilisées dans les projets de recherche, cette présentation fait le point sur ce qu'est le SNDS et comment les chercheurs l'utilisent. Y sont notamment abordés l'accès aux données de santé, leurs parcours, la notion d'open data et le rôle du HDH. Chaque module dure 20 minutes. Cette formation est en accès libre sur le site du HDH. La formation sur les données du SNDS et ses modalités d'accès animées en collaboration avec l'Assurance Maladie a été revue. Cette formation s'adresse à tous les porteurs de projet qui ont l'habitude d'utiliser des données mais ne connaissent pas toujours les détails du SNDS, ses principales bases, la façon dont on y accède et les démarches réglementaires associées. Proposée en visioconférence ou en présentiel, cette formation dure 3h.
- ▶ La documentation collaborative permettant aux porteurs de projet de disposer d'information sur les caractéristiques et le contenu de la base principale du SNDS s'est enrichie.
- ▶ 307 questions ont été posées sur notre forum d'entraide en ligne en 2021.
- ▶ 7 meetups organisés sur des sujets en lien avec le SNDS ont été proposés auprès de nos 1639 membres.

## Usage des données de santé et transparence

Le répertoire public des projets, prévu par la loi, s'est enrichi en 2021 de 1800 projets supplémentaires. Au total, près de 6 000 projets mobilisant des données personnelles de santé y étaient répertoriés fin 2021. Pour chacun d'entre eux, une fiche dédiée présente le bénéfice attendu, les données mobilisées, la méthode, les coordonnées du porteur de projet et de son DPO ainsi qu'un calendrier de mise en œuvre. Un recueil des résultats issus de ces projets a en outre été initié, conformément au code de la santé publique. Pour faciliter la compréhension de cette obligation par les acteurs concernés, un guide pédagogique a été conçu et sera publié prochainement sur notre site internet. Les résultats des projets pourront par ailleurs être téléchargés depuis le répertoire public dans les prochains mois.

En 2021, une réflexion a par ailleurs été lancée afin de faire évoluer la plateforme de soumission des demandes d'accès aux données. Elle se poursuivra en 2022, afin de mieux répondre aux besoins des acteurs intervenant dans cette procédure.



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](mailto:hdh@health-data-hub.fr)**

**Édition : Health Data Hub**

**Conception et réalisation : Parimage**

**Mars 2022**

