



RAPPORT ANNUEL 2022



HEALTH
DATA HUB

SOMMAIRE

Préface de Gérard Raymond	3
Édito de Stéphanie Combes	4

1

LE HEALTH DATA HUB, UNE INFRASTRUCTURE CLÉ POUR LA RECHERCHE EN SANTÉ	5
QUATRE MISSIONS AU SERVICE DE L'INTÉRÊT GÉNÉRAL	6
L'ORGANISATION DU HEALTH DATA HUB	9

2

RETOUR SUR LA FEUILLE DE ROUTE PLURIANNUELLE 2019-2022 DU HEALTH DATA HUB	16
ENJEU 1 : DÉCLOISONNER LE PATRIMOINE DE DONNÉES DE SANTÉ ET AMÉLIORER SA QUALITÉ	18
ENJEU 2 : RENFORCER L'USAGE DES DONNÉES PAR UNE OFFRE DE SERVICE ATTRACTIVE	27
ENJEU 3 : POSITIONNER LA FRANCE COMME UN LEADER DANS L'USAGE DES DONNÉES DE SANTÉ	39
ENJEU 4 : GARANTIR LA PARTICIPATION DE LA SOCIÉTÉ CIVILE ET PROMOUVOIR LES USAGES DES DONNÉES DE SANTÉ	50

3

BILAN DE L'ACCÈS AUX DONNÉES	54
---	-----------



Gérard Raymond,
Vice-Président
du Health Data Hub

PRÉFACE

C'est ce que confirme le bilan de notre première feuille de route. En 2019, notre défi était non seulement d'offrir une plateforme de services très étendue pour accompagner les porteurs de projets, mais aussi de susciter et participer à la naissance d'une culture partagée de la donnée de santé. Ce défi, nous l'avons relevé avec succès, avec la mise en service de notre plateforme technologique, à la pointe en termes de sécurité, de stockage et d'analyse de données. Si nous l'avons relevé, c'est aussi grâce à nos collaborations avec une centaine de partenaires. Nous accompagnons actuellement 80 projets. Et, en lien avec le ministère de la Santé et de la Prévention, nous accompagnons les entrepôts de données hospitaliers de santé qui se mettent en place dans les établissements de santé.

Mais il reste encore beaucoup de chemin à parcourir. Les acteurs de la recherche ont des attentes à la fois sur l'accès aux données et la nature de ces données. C'est le cas par exemple de l'enrichissement des données de l'Assurance Maladie par d'autres sources. Et puis, il est surtout indispensable d'accélérer les délais d'accès. Le futur texte européen sur l'espace européen des données de santé a d'ailleurs placé cette question en tête de ses priorités. C'est dans ce but que nous nous mobilisons pour que le partage des données de santé devienne la norme. L'étude Vivre-Covid 19, que nous avons lancée en 2020 avec France Assos Santé en est un exemple. Il n'a fallu que six semaines pour la mettre en ligne, alors que nous étions en plein confinement. Nous venons

de signer un premier partenariat avec Santé publique France pour partager ces données à des fins de recherche. Cette ouverture est une première pour un collectif d'associations comme France Assos Santé ! Si nous adoptons tous les bons réflexes, ce type de projet ne sera plus l'exception mais la règle.

Pour autant, l'utilisation des données de santé génère encore des craintes, même si les citoyens sont globalement ouverts au partage de leurs données, moyennant un usage éthique et responsable. C'est ce qu'a confirmé la e-consultation citoyenne menée en 2022 dans le cadre du projet européen TEHDAS. C'est pour préserver cette confiance que nous formons les citoyens comme les usagers du système de santé à la seconde vie des données, et que nous les impliquons dans la gouvernance.

Car pour que les patients bénéficient des innovations, la recherche doit se faire, et doit se faire le plus vite possible.

Trois ans après sa création, le Health Data Hub a rempli ses objectifs, avec une ambition renouvelée au service de la recherche en santé.



Stéphanie Combes,
Directrice
du Health Data Hub

ÉDITO

Parce qu'il existe en France l'une des bases de données médico-administratives les plus importantes au monde. Notre pays a été précurseur dans l'exploitation des données de santé. Cette base, qui comporte toutes les données des remboursements des actes de soin sur l'ensemble du territoire, vise à mieux connaître les dépenses de l'Assurance Maladie. Mais, en exploitant ses données, elle a aussi le potentiel pour faire progresser collectivement le système de santé. C'est dans ce cadre, et pour prolonger des initiatives déjà mises en œuvre, que le Health Data Hub (HDH) a fait son entrée dans le paysage, fin 2019. Objectif : accélérer, simplifier, unifier et sécuriser l'accès aux données de santé, en toute transparence, notamment grâce à l'intelligence artificielle (IA).

En trois ans, les usages de ces données de santé se sont multipliés, avec de premiers résultats pour les projets accompagnés ; le périmètre de ces données s'est élargi et s'est ouvert à de nombreux gisements de données en France. Ce projet d'ouverture des données nécessitait à la fois de relever des défis techniques et, pour beaucoup, de s'acculturer à l'usage et au partage des données de santé. C'est pourquoi, l'ambition du HDH était double : non seulement offrir une plateforme de services très étendue permettant d'accompagner les porteurs de projets de l'idée jusqu'aux résultats ; mais aussi, susciter et co-construire une culture partagée de la donnée de santé. Objectif atteint, puisqu'il dispose désormais d'une plateforme à l'état de l'art et d'un dispositif pérenne d'écoute et de participation citoyenne. Mais en trois ans, le Health Data Hub s'est aussi inscrit dans le paysage européen, notamment en tant qu'autorité compétente dans l'action conjointe TEHDAS « Vers un espace européen des données de santé ».

2022 a été une année décisive, permettant au HDH de passer à l'échelle. Il n'est plus un prototype, mais s'inscrit désormais dans le paysage de la recherche en santé grâce à un ensemble de projets qui ont été particulièrement structurants. Il a notamment participé à d'importants travaux issus de la gouvernance stratégique pour formuler la vision à moyen-long terme du patrimoine national des données de santé et les règles du jeu du partage. Sur un plan plus opérationnel, il a aussi mis à la disposition de ses publics une offre de formation et des guides, variés et complémentaires ; il a structuré une offre pour accompagner les entrepôts de données hospitaliers, en aidant le ministère de la Santé et de la Prévention à organiser un appel à projets (AAP) de 50 millions d'euros. Enfin, le HDH a été retenu par la Commission européenne comme chef de file pour construire le prototype de l'espace européen des données de santé, via un consortium.

Le Health Data Hub a ainsi trouvé sa place dans l'écosystème des données de santé, au service de ceux qui les collectent comme de ceux qui les utilisent. Il entend aujourd'hui poursuivre ses efforts pour démocratiser l'usage de ces données, et permettre à tous de mieux connaître les bases de données qui demandent une forte expertise pour être utilisées.

1

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 celles 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.

Le HDH accompagne aujourd'hui 80 projets, dont les premiers ont franchi d'importants jalons permettant d'envisager des avancées exploitables sur le plan médical. Ces projets concernent aussi bien l'amélioration drastique des diagnostics, l'efficacité des parcours de soins, l'analyse des effets secondaires des médicaments, la prédiction d'événements comme les crises d'insuffisance cardiaque, la lecture automatique des comptes-rendus médicaux ou encore la résistance aux antibiotiques. En voici quelques-uns, à titre d'exemples :

● Étudier le lien entre la prise de glucocorticoïdes pour traiter la Covid-19 et le risque d'événements indésirables.

Projet GLUCO porté par IQVIA et PeLyon.

Ce projet demandé par l'Agence européenne du médicament (EMA) est terminé et ses résultats seront publiés dans un article scientifique d'ici mi-2023.

Le HDH y a contribué en transformant au format OMOP-CDM l'extraction de la Caisse nationale d'assurance maladie (CNAM) portant sur 300 000 patients atteints de la Covid-19 hospitalisés, et en testant la qualité des données ainsi transformées. Il s'agit du premier projet conduit sur les données du Système national des données de santé (SNDS) transformées vers un format international et interopérable. Il a été présenté dans différents congrès tels que le Medical Informatics Europe 2022 et l'OHDSI Europe 2022. Par ailleurs, le HDH a ouvert les scripts permettant de passer du SNDS natif vers le SNDS OMOP en open source.

● Comprendre le lien entre la prise d'antibiotiques et le développement ultérieur d'une infection à bactérie résistante.

Projet [BACTHUB](#) porté par l'AP-HP et l'Institut national de la santé et de la recherche médicale (Inserm).

La propagation de bactéries résistantes aux antibiotiques est un problème majeur de santé publique en France. Elle s'accompagne d'infections difficiles à traiter et d'une mortalité accrue. Or, les français consomment beaucoup d'antibiotiques ce qui contribue au développement de la résistance bactérienne. Le risque à l'échelle d'un individu, qui prendrait par exemple des antibiotiques pour soigner une bronchite ou une angine, reste mal connu.

C'est pour étudier ce lien que le projet BACTHUB a été lancé. Pour ce faire, le HDH a aidé les porteurs du projet à consolider les données de plus de 30 000 patients hospitalisés dans 14 hôpitaux de l'AP-HP sur une période de quatre ans.

Deux de ses data ingénieurs y ont travaillé pendant plus d'un an. Ils ont créé une base de données unique sur les infections bactériennes dans le sang, base qui couple des données cliniques et microbiologiques de qualité. Un article universitaire présentant cette base a été publié dans *Clinical Epidemiology*¹, et ces travaux ont fait l'objet d'une communication orale et d'une communication affichée au congrès international de l'ES-CMID. Cette base de données sera prochainement enrichie avec des données de l'Assurance Maladie. Cela permettra d'évaluer la délivrance d'antibiotiques aux patients avec une bactériémie, avant leur hospitalisation ; d'étudier le lien entre la prise d'antibiotiques et le développement d'une infection à bactérie résistante.

● **Développer un outil permettant de quantifier la proportion de patients touchés par les effets indésirables d'un médicament.**

Projet [ORDEI](#) porté par l'Agence nationale de sécurité du médicament et des produits de santé (ANSM).

Nausée, maux de tête, démangeaison, somnolence, diarrhée, vision floue, fatigue, éruption cutanée... Les effets indésirables, ou effets secondaires, sont des réactions nocives et non voulues, susceptibles de se produire après la prise d'un médicament. Les professionnels de santé, les patients ou associations de patients, et les laboratoires pharmaceutiques peuvent les signaler en quelques clics aux autorités sanitaires sur le site signalement-sante.gouv.fr.

Le projet ORDEI a pour but de développer un outil capable de calculer automatiquement le taux de déclaration des effets indésirables. Mis en production fin 2022, il sera accessible à tous dès avril 2023.

● **Analyser les comptes-rendus des urgences dans les dossiers médicaux des patients victimes d'un accident pour identifier le lien avec une éventuelle prise de médicaments.**

Projet [TARPON](#) porté par l'équipe IETO-BPH de l'Université de Bordeaux.

Pouvoir croiser les informations sur les chutes, accidents de la route, du travail ou de la vie courante contenues dans les dossiers des urgences, avec les données de ces mêmes patients relatives à la prise éventuelle de médicaments avant leur accident, telle est l'ambition du projet TARPON.

Une première phase du projet consistant à construire la méthodologie a permis de développer un algorithme d'identification des patients pris en charge pour traumatisme aux urgences. Plusieurs articles ont été publiés ou soumis à des revues scientifiques^{1,2,3 et 4}.

● **Accélérer et améliorer la prise en charge des patients en ophtalmologie.**

Projet [SEDAAR](#), porté par l'Hôpital Fondation A. de Rothschild.

1,7 million de personnes vivant en France seraient atteintes de handicaps visuels. Ces troubles sont essentiellement liés à des pathologies de la rétine, dont le diagnostic est en grande majorité établi par les ophtalmologues grâce à divers examens d'imagerie. L'enjeu est d'automatiser en partie ces diagnostics car plus vite ils sont posés, plus les traitements sont efficaces. Aussi, les spécialistes se raréfient. La mise à disposition des données sur l'espace d'analyse du HDH est en cours en janvier 2023 et des résultats sont attendus d'ici mars 2023. Par la suite, la base de référence qualifiée, multimodale et multi-pathologies unique en France créée pour le projet, pourra être annotée par des spécialistes puis partagée aux professionnels de santé et aux industriels pour faciliter d'autres projets de recherche clinique.

1. Salam Abbara, Didier Guillemot, Salma El Oualydy, Maeva Kos, Cécile Poret, Stéphane Breant, Christian Brun-Buisson & Laurence Watier (2022) Antimicrobial Resistance and Mortality in Hospitalized Patients with Bacteremia in the Greater Paris Area from 2016 to 2019, *Clinical Epidemiology*, 1547-1560, DOI: 10.2147/CLEP.S385555

2. Chenais G., Gil-Jardiné C., Touchais H., Combes X., Avalos-Fernandez M., Contrand B., Tellier E., Bourdois L., Revel P., Lagarde E. Development and Validation of Deep Learning Transformer Models for Building a Comprehensive and Real Time Trauma Observatory. *JMIR AI*. Preprint doi: 10.2196/40843

3. Chenais G., Lagarde E., Gil-Jardiné C. Artificial Intelligence in Emergency Medicine: a Viewpoint of Current Applications, Foreseeable Opportunities and Challenges. *J Med Internet Res*. 2022. Preprint doi: 10.2196/40031

4. Les liens vers les articles mentionnés en pre-print : <https://hal.inria.fr/hal-03276129> ; <https://hal.inria.fr/hal-03276124>

L ORGANISATION DU HEALTH DATA HUB

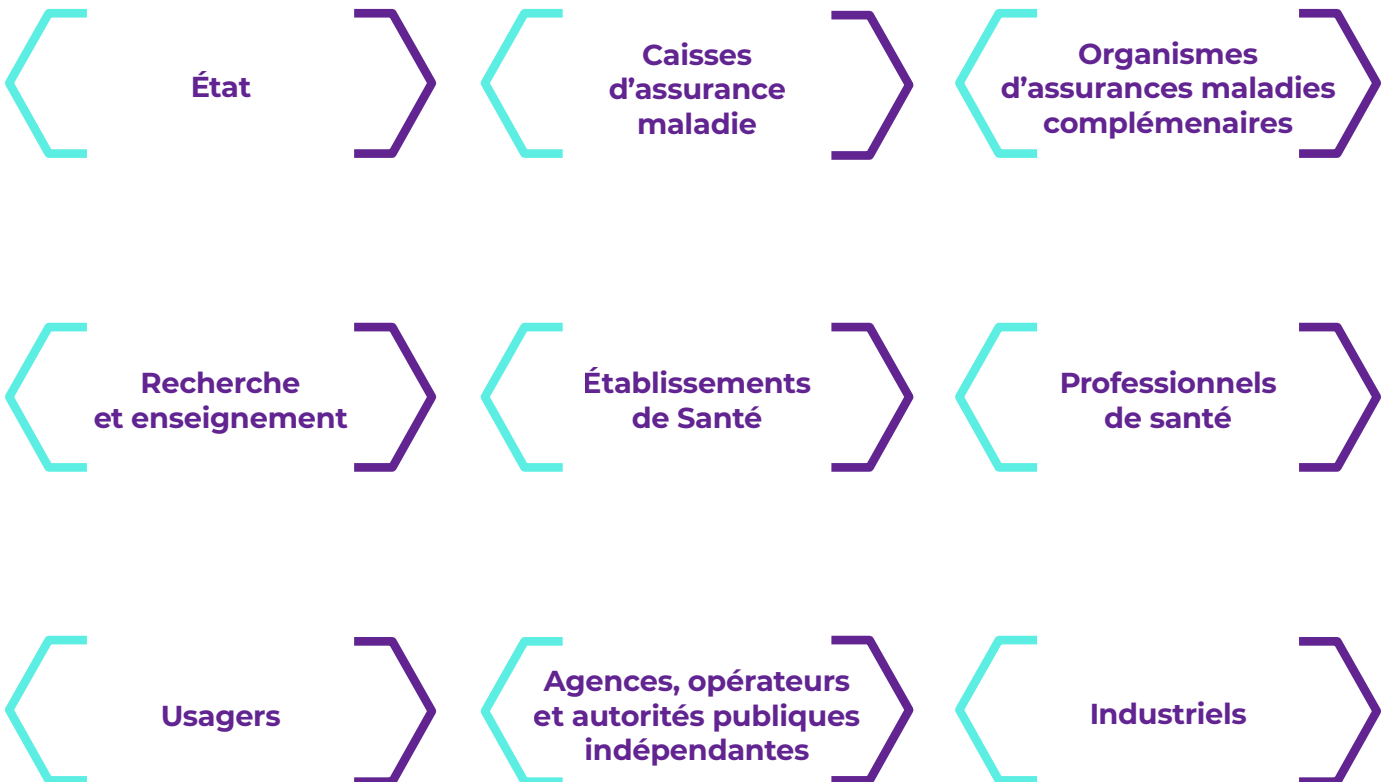
L'équipe du Health Data Hub

L'équipe du HDH est composée de 89 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 Health Data Hub est un Groupement d'intérêt public, créé par la loi du 24 juillet 2019 relative à l'organisation et la transformation du système de santé. Ses statuts sont publiés le 30 novembre 2019.

Son assemblée générale réunit 56 membres, répartis en neuf collèges : l'État, les caisses d'assurance maladie, les organismes d'assurances maladies complémentaires, la recherche et l'enseignement, les établissements de santé, les professionnels de santé, les agences, opérateurs et autorité publique indépendante, les usagers du système de santé et les industriels. Elle fixe les grandes orientations du programme de travail annuel du HDH.



Un représentant de chaque collège siège au conseil d'administration, sauf pour l'État qui en a deux : la DREES, pour le ministère de la Santé et de la Prévention ainsi que la DGRI, pour le ministère de l'Enseignement supérieur et de la Recherche. La CNAM, l'UNOCAM, l'Inserm, le CHU de Toulouse, l'UNPS, France Assos Santé, Santé publique France (SPF) et l'AFROS en sont également membres.

L'assemblée générale et le conseil d'administration ont été présidés par Bruno Maquart, jusqu'en juillet 2022. Gérard Raymond, également président de France Assos Santé, lui a succédé, par intérim. Le président de France Assos Santé est également vice-président du HDH, témoignant de la place importante qu'occupent les associations dans la gouvernance de la structure.

En janvier 2021, le HDH s'est doté d'un conseil scientifique, composé de 28 experts (dont sept experts internationaux) aux compétences complémentaires : médecins, pharmaciens, spécialistes de l'intelligence artificielle, juristes et spécialistes de l'éthique, généticiens, épidémiologistes, économistes de la santé, etc. Depuis septembre 2022, le conseil est présidé par la généticienne et épidémiologiste Ségolène Aymé, ex vice-présidente. Elle a succédé à Alain Livartowski qui en reste membre. Le Pr Bertrand Lukacs, chirurgien urologue, spécialiste des données en urologie et cancérologie, a pour sa part été nommé vice-président en remplacement de Ségolène Aymé.

Le conseil scientifique consultatif émet des avis pour 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 financement du HDH, de 21 millions d'euros par an en moyenne, est principalement public. 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 2022, le HDH comptait 89 collaborateurs, dont 54 femmes (soit 60.7 %) travaillant dans des équipes métier qui accompagnent les projets de bout en bout (le guichet, la direction des partenariats, la direction des données, la direction technique et le produit, la direction projets et services utilisateurs), dans des directions stratégiques (la direction scientifique, la direction médicale et la direction citoyenne) ou des fonctions support (la direction de la sécurité des systèmes d'information, le secrétariat général, la direction de la communication et la direction juridique).

7 GRANDES ORIENTATIONS POUR 2022

En 2022, le HDH a poursuivi ses actions pour déployer les projets qu'il accompagne, étendre ses relations avec les associations et les citoyens, développer ses partenariats avec des organismes publics et privés, ou encore augmenter son empreinte européenne et internationale.

Ces actions ont été organisées autour de sept grandes orientations :

1 Intégrer et enrichir la base principale et le catalogue.

Suite à la publication du décret du 29 juin 2021 relatif au traitement de données à caractère personnel dénommé « Système national des données de santé » et à la publication de l'arrêté du 12 mai 2022 relative à la première version du catalogue, le HDH s'est donc préparé à accueillir les données dans sa plateforme. Une offre de services à destination des acteurs à l'origine de la collecte et de la mise en qualité des données a été créée pour mettre en valeur les données auprès de l'écosystème.

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

L'attractivité des données de santé pour la France et l'international dépend des délais d'accès. Le HDH a donc poursuivi avec ses partenaires tels que la CNAM, l'industrialisation de ses processus clés : ingestion, pseudonymisation, préparation et mise à disposition des données de santé. Ces travaux devront se poursuivre en 2023.

3 Consolider une plateforme technologique respectant un principe de réversibilité et la stratégie nationale pour un cloud de confiance.

Développée depuis 2019, la plateforme technologique du HDH a été mise en service dès le printemps 2020. Cette plateforme constitue un défi technologique car elle doit répondre à d'importantes exigences de sécurité mais également proposer une offre logicielle permettant de répondre à des projets innovants. Le HDH a également poursuivi ses actions de réduction de l'adhérence aux services proposés par son hébergeur actuel et a réalisé un benchmark auprès des fournisseurs de cloud de confiance.

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

Quatrième pilier de la feuille de route stratégique du HDH, cette orientation, française et européenne, se renforce à mesure que le HDH se déploie. L'écoute, l'information, la formation et les partenariats avec les associations d'utilisateurs du système de santé constituent les quatre priorités de sa direction citoyenne.

5 Contribuer à l'élaboration d'une stratégie de financement de bases de données pour promouvoir leur partage et leur développement.

Pour que le HDH atteigne son plein potentiel, il est nécessaire que les conditions du partage des données et la juste rétribution des efforts des acteurs soient assurées. Ces travaux ont vocation à être menés au sein du comité stratégique des données de santé et représentent une priorité pour le Health Data Hub qui en est membre.

6 Contribuer à préfigurer l'espace européen des données de santé.

Le 3 mai 2022, la Commission européenne a publié un projet de règlement sur l'espace européen des données de santé - European Health Data Space (EHDS) - dont l'objet est de faciliter l'accès aux différents types de données disponibles au sein des États membres pour développer la recherche et améliorer la santé des citoyens. Afin de tester une première version de cet espace, la Commission a par ailleurs lancé un appel à projets afin de construire une première version test de

cet EHDS. En juillet 2022, le HDH a été désigné par la Commission européenne avec 16 autres partenaires pour un budget de 5 millions d'euros, pour mener un projet visant à tester un parcours utilisateur pour la création, le déploiement et la gestion de projets de recherche sur les données de santé au niveau européen, autour de cinq cas d'usage.

7 Poursuivre la structuration interne.

Comptant aujourd'hui 89 collaborateurs, le HDH doit poursuivre son passage à l'échelle et le recrutement des expertises clés nécessaires à la bonne réalisation de ses missions.

UNE NOUVELLE FEUILLE DE ROUTE POUR 2023-2025

La feuille de route stratégique pluriannuelle 2019-2022 du HDH arrivant à échéance le 31 décembre, une nouvelle feuille de route a été proposée, en lien avec les instances, pour la période 2023-2025. Pour la construire, le HDH a consulté 26 acteurs de l'écosystème et animé quatre groupes de travail, pour 46 organismes rencontrés. Présentée en conseil d'administration et en assemblée générale, elle a été approuvée à l'unanimité le 9 juin dernier. Grâce à ce travail en profondeur, qui aura duré six mois, le HDH a pu reformuler ses axes stratégiques afin d'aligner ses priorités sur ses nouveaux enjeux.

Le HDH a mené une phase de consultation auprès des acteurs de l'écosystème pour construire sa feuille de route. Parmi les principaux constats qui ont émergé, il y a le fait que la donnée de santé

permet d'accélérer l'arrivée d'innovations sur le marché. Néanmoins, les délais d'accès à ces données restent trop importants, ce qui retarde également l'accès aux innovations pour les patients. La réussite dépendra d'abord de la valeur d'usage constatée par toutes les parties prenantes, dont les citoyens. Parmi les autres constats : les difficultés rencontrées dans la mise en œuvre pratique de certaines exigences réglementaires, des échanges plus étroits entre les différentes instances sont à prévoir pour permettre une plus grande fluidité. Enfin, de manière unanime, les parties prenantes ont demandé à ce que l'enrichissement de la base principale du SNDS par des données cliniques soit inscrit comme une priorité pour les prochaines années de vie du dispositif.

Ainsi, durant les trois prochaines années, le HDH va renforcer ses actions préexistantes autour de ses quatre axes stratégiques originels, tout en resserrant ses liens avec l'Europe ainsi qu'avec les autres « hubs » de santé à l'échelle régionale et locale. La nouvelle feuille de route pluriannuelle du HDH se compose de quatre grands axes :



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

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

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

Renforcer ses connexions avec les acteurs de l'écosystème



LES TEMPS FORTS ET CHIFFRES CLÉS DU HDH

80 PROJETS INNOVANTS...



... QUI PROGRESSEDENT



UN RAYONNEMENT INTERNATIONAL

17

PARTENAIRES AU SEIN DU PROJET « HEALTHDATA@EUPILOT » VISANT À CRÉER ET TESTER UNE VERSION BETA DE L'ESPACE EUROPÉEN DES DONNÉES DE SANTÉ AUTOUR NOTAMMENT DE CINQ CAS D'USAGE

1 PARTICIPATION À L'APPEL À PROJETS EUROPÉEN TEF-HEALTH

8 ÉVÈNEMENTS DE PORTÉE INTERNATIONALE SUR LES 15 ORGANISÉS PAR LE HDH EN 2022 :

...dont **5** évènements labellisés dans le cadre de la présidence française du Conseil de l'Union européenne (PFUE)

470 participants à l'évènement « IA et la médecine : promesses et limites » co-organisé avec le MIT, l'Académie nationale de médecine et l'Ambassade de France aux USA en octobre 2022

405 participants à la 4^e édition du colloque « données de santé et intelligence collective » organisé par le HDH et la DREES en décembre 2022

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

UN PROJET RÉSOLUMENT CITOYEN

10 PARTENARIATS AVEC DES ASSOCIATIONS


5699 CONTRIBUTIONS DÉPOSÉES DANS LE CADRE DU PREMIER DÉBAT CITOYEN EUROPÉEN SUR LES DONNÉES DE SANTÉ

2

RETOUR SUR LA FEUILLE DE ROUTE PLURIANNUELLE 2019-2022 DU HEALTH DATA HUB



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.

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. 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, d'entrepôts de données hospitaliers, etc. La grande base de données médico-administratives, rassemblant historiquement le Système national d'information inter-régimes de l'Assurance Maladie (SNIIRAM), le Programme de médicalisation des systèmes d'information (PMSI) et le Centre d'épidémiologie sur les causes médicales de décès (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 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 de la Santé et de la Prévention 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 technologiques dédiés par projet et hautement sécurisés, mais aussi la maintenance des environnements.

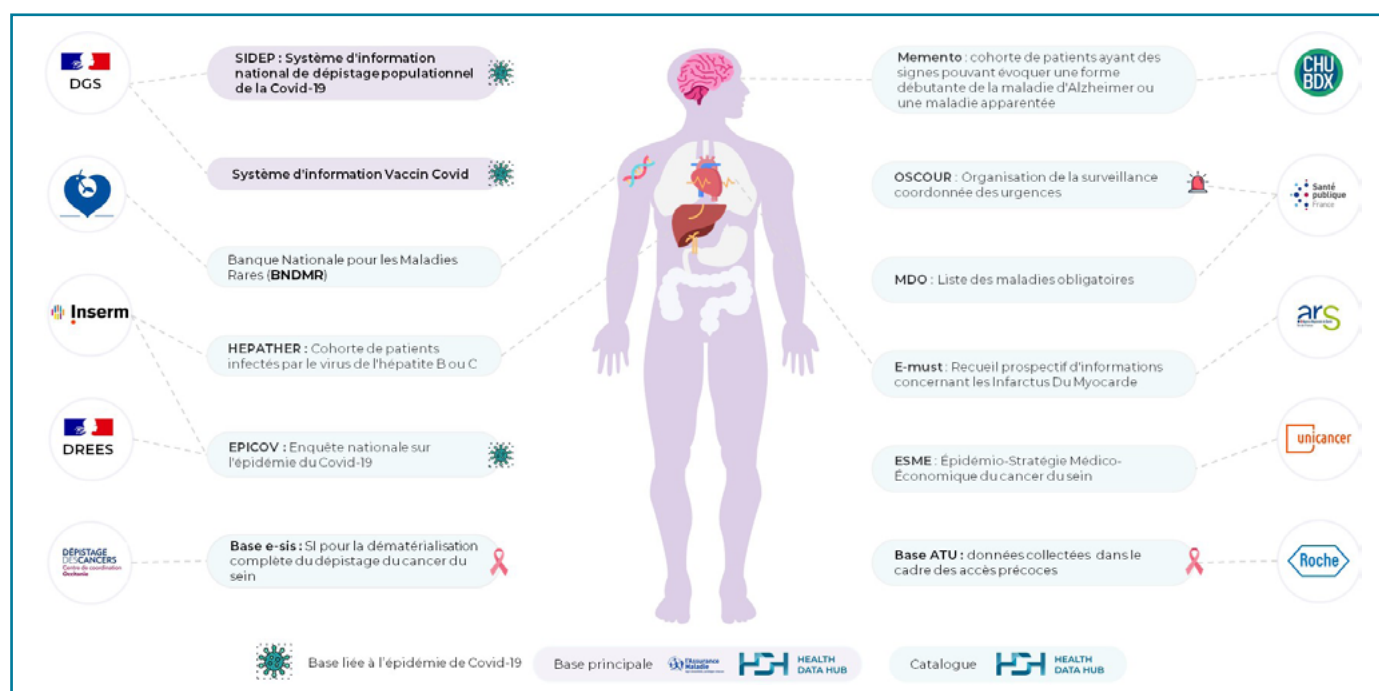
Toutes les bases de données ne peuvent cependant pas être concernées par ce dispositif. Elles sont donc sélectionnées, et la collection retenue compose ce que l'on appelle le « catalogue » du HDH. Ce dernier se construit donc progressivement, en lien avec les acteurs à l'origine des bases de données concernées ou ceux qui les administrent, sous le décret dit SNDS d'application de la loi Organisation et Transformation du Système de Santé du 29 juin 2021.

Une première liste des bases de la base principale et du catalogue du SNDS

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

Le 12 mai 2022 a été publié le premier arrêté listant les bases qui composent la base principale et le catalogue du SNDS.

Le comité stratégique des données de santé existant alors depuis peu, la sélection des bases s'est faite en plusieurs étapes. Un groupe de travail du ministère de la Santé et de la Prévention, réunissant plusieurs acteurs clés de l'écosystème, a proposé une première liste de bases de données de natures diverses (administratives, nationales, cohortes ou registres et entrepôts thématiques en réponse à des priorités de santé publique) ainsi qu'une base d'origine privée. Le conseil scientifique consultatif du HDH a ensuite émis un avis sur cette première liste, alors que dans le même temps la conformité juridique des bases proposées était vérifiée. Ces étapes ont permis, en lien avec le ministère de la Santé et de la Prévention de présenter une deuxième liste, plus courte, au comité stratégique des données de santé. Après avis de ce dernier, un projet d'arrêté listant les bases de la base principale et la liste des bases du catalogue a été soumis pour instruction à la CNIL.



Un dispositif en six étapes pour inscrire les bases de données au catalogue

Créé par l'arrêté du 29 juin 2021, le comité stratégique des données de santé, présidé par le ministre de la Santé et de la Prévention, réunit plusieurs acteurs clés de l'écosystème. Son rôle est principalement de proposer des orientations sur le développement du Système national des données de santé. A ce titre, il identifie notamment les bases de données éligibles à l'arrêté détaillant la liste des bases composant la base principale et le catalogue du SNDS, recommande leur inscription et, le cas échéant, émet un avis sur l'ordre de priorité de cette inscription.

Un groupe de travail a été créé pour prendre en charge cette mission. Il regroupe les représentants des institutions du comité stratégique des données de santé et s'est réuni à deux reprises entre décembre 2021 et février 2022. Objectifs : définir un dispositif d'inscription des bases à l'arrêté catalogue et créer un kit documentaire pour mettre en œuvre ce dispositif, avec une grille permettant d'analyser la pertinence d'une base selon des critères bien définis.

Ce dispositif prévoit six étapes :

Identifier les bases de données pressenties.

Ces bases ou thématiques peuvent être directement proposées par le comité stratégique des données de santé, en fonction des priorités de santé publique ou de recherche du moment. Les responsables de données peuvent aussi eux-mêmes solliciter le HDH pour que leur base de données soit inscrite au catalogue du SNDS.

Décrire les bases de données pressenties.

Pour ce faire, le responsable de chaque base doit compléter une fiche descriptive, rédiger une lettre d'intention et fournir les autorisations réglementaires relatives à sa base. Pour enrichir sa description, il dispose d'une fiche d'auto-évaluation de la qualité et d'une fiche d'expertise qu'il peut compléter et joindre à son dossier.

Organiser une expertise externe. Après avoir rassemblé ces éléments pour les bases de données pressenties, le HDH organise leur expertise externe à partir de la fiche d'expertise. Sont évalués les critères suivants : pertinence, potentiel de valorisation, originalité, caractéristiques et qualités. L'expert peut s'entretenir avec le responsable de données, s'il le juge nécessaire.

Vérifier la complétude des dossiers. Avant de transmettre tous les éléments au comité stratégique des données de santé, une check-list, faisant partie du kit documentaire, permet au HDH de préciser que les éléments clés ont bien été discutés en amont avec le responsable de données.

Prioriser les bases. Le comité stratégique des données de santé se réunit et priorise les bases de données dont les dossiers complets lui ont été transmis par le HDH.

Publier l'arrêté détaillant la liste des bases composant la base principale et le catalogue, après avis de la CNIL.

Ce dispositif, tout comme le kit documentaire, a été approuvé par le comité stratégique des données de santé le 6 juillet 2022.

Des engagements vis-à-vis des responsables de données

Pour favoriser le partage de ces données, le HDH a pris plusieurs engagements vis-à-vis de ceux qui les collectent et les mettent en qualité, 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 pour la mise en conformité avec le règlement général sur la protection des données (RGPD).
- Il valorise leur travail et leur expertise. Il renforce la visibilité des bases de recherche, ce qui est susceptible d'augmenter 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.

En 2022, le HDH a notamment lancé des ateliers pour aider à la mise en conformité des bases de données souhaitant intégrer le catalogue. Pour cela, il s'est appuyé notamment sur le référentiel que la CNIL a publié le 26 octobre 2021 pour encadrer la constitution d'entrepôts de données de santé (EDS) et sur la « check-list » qu'elle a partagé sur son site en septembre 2022. Ce document reprend toutes les dispositions du référentiel et permet aux acteurs de s'assurer de la conformité de leur projet.

Action 1.2 **Installer une procédure nationale automatisée pour faciliter la transmission des données à des fins d'appariements directs via le numéro d'inscription au répertoire (NIR)**

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, lorsqu'il est

disponible. Cet identifiant est 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 certaines actions doivent se faire 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 national qui permettra d'automatiser certaines étapes et de rendre la procédure plus accessible à ses nombreux bénéficiaires. Un outil qui va intégrer l'automatisation des croisements impliquant une reconstitution du NIR par la Caisse nationale d'assurance vieillesse (CNAV). Ce système d'information pourra ainsi réceptionner les NIR (ou des informations d'état civil pour la reconstitution), et assurer leur pseudonymisation et leur transmission à la CNAM pour faciliter les appariements et la mise en œuvre des projets.

Au-delà de simplifier les appariements, cet outil permettra aussi de manipuler le NIR en toute sécurité pour que les personnes concernées puissent faire valoir leurs droits et que celles qui le souhaitent puissent, par exemple, s'opposer à la réutilisation de certaines de leurs données à des fins de recherche.

Elles pourront le faire via une interface accessible à tous les citoyens. Son développement a démarré en 2022 et un cahier des charges pour la procédure de facilitation des appariements a été conçu. La mise en pratique de cet outil est prévue pour 2023.

En attendant, des échanges ont eu lieu entre la CNAV, l'Inserm, la CNAM et le HDH pour accélérer la mise en œuvre de l'appariement avec reconstitution du NIR pour certaines cohortes.

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

Trois projets sélectionnés dans le cadre du partenariat avec Unicancer autour du programme UNIBASE

Avec plus de 380 000 nouveaux cas par an en France, les cancers représentent la première cause de décès chez les hommes et la deuxième chez les femmes. 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 dix CLCC disposant d'entrepôts de données ConSoRe, le moteur de recherche pour le big data en cancérologie.

Trois projets pilotes, en phase avec le Plan cancer 2021-2030, ont été sélectionnés en 2022. Avec, à la clé, un accompagnement et un financement du HDH.

ISIS

Le projet ISIS a pour objectif principal d'évaluer le risque d'infertilité chez les femmes jeunes ayant souffert d'un cancer du sein. À plus long terme, les résultats pourraient ouvrir la voie à de nouvelles recommandations pour préserver la fertilité chez ces patientes. Pour informer les patientes et leur proposer des solutions préservant leur fertilité, il est indispensable de mieux connaître les risques d'infertilité liés aux thérapies anticancéreuses, y compris les thérapeutiques nouvelles et à venir. Ce projet qui implique l'Institut de cancérologie de l'Ouest, le centre Léon Bérard, l'Institut Curie et le centre Oscar Lambret, est d'ores et déjà accompagné par Unicancer et le HDH qui viennent de signer une convention. En cours d'élaboration, le protocole scientifique devrait pouvoir être soumis au comité scientifique et éthique national (CESREES) d'ici début 2023 et faire l'objet d'une convention de partenariat avec le HDH.

REALIGIST

Le projet REALIGIST a pour objectif de constituer une cohorte inédite de patients atteints de tumeurs stromales gastro-intestinales (GIST). Avec 800 nouveaux cas par an en France, les GIST sont des cancers rares, de la famille des sarcomes, qui débutent dans des cellules de la paroi du tube digestif. Malgré la découverte d'anomalies génétiques au sein de ces tumeurs et des avancées thérapeutiques majeures, un grand nombre de questions cliniques et thérapeutiques restent non

résolues. En associant pour la première fois les données cliniques et biologiques issues du soin « en vie réelle » et les données issues de l'Assurance Maladie (base principale du SNDS), ce projet permettra d'identifier de potentiels facteurs de risque, d'analyser les parcours de soin des patients et d'identifier de nouveaux facteurs associés à la réussite ou à la résistance aux traitements anticancéreux. Soutenu par sept CLCC, ce projet a obtenu l'avis favorable du CESREES et s'apprête à demander une autorisation à la CNIL. Il fait l'objet d'une convention de partenariat avec le HDH.

DASTO

Le projet DASTO s'intéresse à la prédiction du risque de maladie thromboembolique veineuse (MTEV) chez les patients atteints de cancer, avec pour objectif de comprendre les facteurs de risque au sein d'une cohorte de 100 000 patients. Les données de la base principale du SNDS seront corrélées avec celles des dossiers médicaux analysés par l'outil ConSoRe. Impliquant quatre CLCC, ce projet a récemment obtenu l'avis favorable du CESREES et s'apprête à demander une autorisation à la CNIL. Il fait l'objet d'une convention de partenariat avec le HDH.

Forts de cette expérience, Unicancer et le HDH ont souhaité ouvrir UNIBASE aux CLCC n'utilisant pas ConSoRe, ainsi qu'à d'autres structures hospitalières. Pour lancer la seconde vague de l'AMI, ils se sont associés à la Ligue nationale contre le cancer.

Quatre projets pilotes, en phase avec le Plan cancer 2021-2030, seront sélectionnés en 2023. Tous seront accompagnés par Unicancer et le HDH, trois bénéficieront d'un financement du HDH et un d'un financement de la Ligue nationale contre le cancer.

Des travaux seront menés pour que les données soient interopérables et modélisées conformément à des standards internationaux.

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. En pratique, il s'agit de réaliser des projets de recherche exploitant des bases de données de santé mutualisées entre les différents centres hospitaliers.

Parmi les différentes initiatives en cours, les deux projets les plus avancés sont les suivants :

AVC ischémiques est issu d'une collaboration entre l'Hôpital Saint-Joseph, le CHRU de Nancy et le CHU de Caen. Ce projet vise à construire un score prédictif de gravité en revisitant le National Institutes of Health Stroke Scale, outil utilisé actuellement pour le suivi évolutif d'un accident vasculaire cérébral (AVC), pour y intégrer notamment des dimensions biologiques et radiologiques. Ce score permettrait de mieux prédire les séquelles et le handicap fonctionnel du patient à la suite d'un AVC. Le parcours de soins (rééducation, réadaptation, réhospitalisation) pourrait être envisagé dès la sortie de l'hôpital pour les six mois à venir.

Dyspnée Aiguë est issu d'une collaboration entre le CHRU de Nancy et les Hospices civils de Lyon (HCL). Ce projet a pour objectif de mieux identifier les caractéristiques des patients pris en charge aux urgences pour une dyspnée aiguë. Les résultats du projet permettront de proposer des pistes pour améliorer la filière de soins et construire un score de risque de morbi-mortalité.

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

Fin 2021, l'Agence européenne du médicament a retenu la candidature du HDH et de ses partenaires dans le cadre d'un appel d'offres sur des données en vie réelle. Objectif : pouvoir accéder à des bases européennes de données individuelles hospitalières, et au format OMOP-CDM.

Pour y répondre, le HDH 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'Agence du numérique en santé (ANS) et Capgemini.

Ensemble, ils ont défini un socle commun de données hospitalières pour alimenter leur futur entrepôt de données de santé multicentrique. Une base de données clinique sera ensuite enrichie par des données médico-administratives de la base principale du SNDS, et transformée dans un modèle commun international permettant de les rendre interopérables. L'entrepôt de données sera hébergé sur la plateforme technologique du HDH, à laquelle pourront accéder les partenaires hospitaliers, l'European Medicines Agency (EMA), ainsi que les porteurs de projet autorisés.

Le Health Data Hub accompagne le déploiement des entrepôts de données de santé hospitaliers

Les données disponibles dans les établissements de santé sont nombreuses et présentent un important potentiel pour la recherche, l'innovation et l'amélioration des systèmes de soin, notamment via le pilotage. Pour autant, on ne peut pas toujours les analyser car elles ne sont pas toujours organisées en entrepôts de données de santé. Dans un contexte où l'utilisation des données de santé revêt un caractère de plus en plus stratégique au niveau international et avec la dynamique enclenchée par la Commission européenne autour de l'espace européen des données de santé, la constitution

d'un réseau d'EDS hospitaliers en France est devenue prioritaire, ce que souligne la stratégie d'accélération Santé Numérique (SASN).

C'est pour accélérer l'émergence de ces EDS à l'échelle nationale, que le ministère de la Santé et de la Prévention a lancé en 2022 un appel à projets en deux vagues, doté d'une enveloppe de 50 millions d'euros avec, pour les lauréats, un financement à hauteur de la moitié de leurs dépenses. Les candidatures peuvent regrouper jusqu'à six partenaires par projets éligibles (établissements de santé, publics ou privés et groupements de coopération sanitaires) sous réserve d'une activité combinée suffisante.

En pratique, le ministère s'appuie sur le HDH qui a participé à la rédaction du cahier des charges ainsi qu'au pilotage opérationnel de l'appel à projets auprès de la Banque publique d'investissement (BPI). Le HDH est également chargé de l'accompagnement des candidats et lauréats pendant toute la durée de l'appel à projets. Il pourra ainsi identifier au plus près des acteurs du terrain les sujets d'accompagnement prioritaires et élaborer une doc-trine nationale en matière d'EDS, avec les établissements retenus. En outre, le HDH accompagnera les lauréats pour inscrire dans son catalogue un périmètre de données issues de leur EDS, ce qui était d'ailleurs l'un des critères de sélection de l'appel à projets.

En 2022, le HDH a lancé deux premières actions pour aider les candidats à constituer leur dossier : un événement consacré aux EDS à Parisanté Campus le 12 septembre et un catalogue de services. Plus de 600 personnes y ont participé, dont 300 en présentiel, l'occasion d'échanger sur des sujets liés aux entrepôts de données. Un temps d'échange entre une vingtaine d'éditeurs de solutions d'EDS et des établissements de santé a également été organisé afin de créer des interactions entre ces communautés et stimuler les partenariats.

Le HDH a également accompagné la direction interministérielle du numérique (DINUM) dans la mise à jour de son catalogue [GouvTech](#) qui recense les solutions numériques proposées aux services publics pour y ajouter une catégorie dédiée aux entrepôts de données de santé. Cette initiative, qui a déjà permis en quelques mois de centraliser les informations d'une quinzaine d'éditeurs, offre une meilleure connaissance du marché aux établissements de santé.

En 2023, le HDH poursuivra son travail auprès des lauréats, avec des groupes de travail sur des sujets prioritaires et des sessions collectives pour mutualiser les bonnes pratiques, les outils et les connaissances entre les lauréats, puis au-delà. Un premier groupe de travail a d'ores et déjà été lancé sur le financement des bases de données de santé publique (cf. infra).

Le Health Data Hub contribue au lancement de l'entrepôt national de médecine de ville

En tant que membre du consortium porté par le Collège national des généralistes enseignants (CNGE), le HDH a contribué en 2022 au lancement de l'entrepôt national des données de médecine de ville. Ce projet d'entrepôt, baptisé Platform for Data in Primary care (P4DP), est lauréat de l'AMI Santé numérique. Son budget de 14,38 millions d'euros sera financé par Bpifrance à hauteur de 9,73 millions d'euros. Objectif : créer le premier entrepôt de données de santé de médecine de ville à l'échelle nationale chaîné avec la base principale du Système national des données de santé. Une partie de cet entrepôt, partagée avec le HDH, permettra de développer des projets de recherche à large échelle au-delà de son rôle premier consistant à permettre aux généralistes chercheurs d'accéder à des données fiables, des tableaux de bord de suivi de leur activité et des outils de modélisation prédictive.

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

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 ne suffit pas. Il faut aussi recenser et documenter précisément le contenu des bases disponibles, pour que les porteurs de projets puissent converger rapidement vers les données qui les intéressent. C'est dans cette optique que le HDH collecte et met en ligne les métadonnées de chacune des bases de son catalogue.

Concrètement, en consultant ces métadonnées qui préservent la confidentialité, les porteurs de projets peuvent identifier rapidement les bases d'intérêt grâce à des informations standardisées et exhaustives. Cette démarche favorise les échanges et les collaborations entre les scientifiques et le HDH.

Une première version de ce catalogue de métadonnées a été déployée par le HDH fin 2022. De nouvelles fonctionnalités seront accessibles au cours du premier semestre 2023.

Cette première version comporte trois niveaux d'information (bases, tables, variables) pour les métadonnées des bases relationnelles. Grâce à un moteur de recherche, l'utilisateur peut identifier les bases qui répondent le mieux à sa problématique et en explorer le contenu : description générale, contexte scientifique, modalités de récolte et de mise à disposition, description technique et statistique de chaque variable.

Ce catalogue se base sur la connaissance métier de nombreux experts et son implémentation a été entièrement faite par les équipes du HDH. Elle s'inscrit dans une démarche internationale de description standard des données de santé et

5. Consortium regroupant le HDH, la start-up Loamics, ainsi que le CHU de Rouen, l'Université Nice Côte-d'Azur et l'Université de Rouen Normandie, avec le soutien de la CNAM et du Ministère de la Santé et de la Prévention.

garantira, à terme, les meilleures modalités d'accès et d'échange avec d'autres plateformes de données de santé telles que l'espace européen des données de santé.

Une réflexion approfondie sur le financement pérenne des bases de données en santé

L'exploitation des données de santé rapporterait chaque année près de 7,3 milliards à l'économie française, selon l'Organisation de coopération et de développement économiques (OCDE). C'est ce que révèle l'étude du cabinet Veltys remise au HDH en juillet 2022. La France dispose de tous les atouts pour s'imposer pleinement comme leader naturel sur le sujet : nouveaux médicaments, recherche, outillage des professionnels de santé... Les innovations à venir ne manquent pas et l'écosystème est en attente. Des efforts restent à faire, notamment pour construire une véritable politique publique de la donnée de santé et pour que les grands programmes d'investissement accordent une place plus importante au financement de la collecte, de la mise en qualité et de la réutilisation innovante des données.

En 2021, le HDH avait commencé à instruire la mise en place d'une redevance pour les acteurs privés souhaitant accéder directement aux données de santé. Cette instruction avait montré que le contexte réglementaire laissait peu de place pour des modèles de valorisation économique permettant de couvrir tous les coûts de collecte et de mise en qualité de ces données. Ces travaux sont importants mais nécessaires pour obtenir des données de santé fiables et protégées.

En ligne avec cette première conclusion, le comité stratégique des données de santé a lancé à l'été 2022 un groupe de travail rassemblant institutionnels et responsables de bases de données de santé. Objectif : préciser les contours d'un éventuel modèle de financement public pérenne permettant d'alimenter et d'actualiser les bases de

données d'intérêt dans la durée mais aussi de favoriser leur partage. Portés par l'actualité et le lancement de l'AAP du ministère de la Santé et de la Prévention sur les d'entrepôts de données de santé hospitaliers, les travaux de ce groupe ont été fructueux. Un premier livrable pour un modèle de financement public des EDS a été constitué au second semestre 2022.

Ces travaux se poursuivront en 2023 pour définir un régime commun et uniformisé de redevances d'accès aux données de santé.

Focus sur la 4^{ème} édition du colloque du Health Data Hub « Données de santé et intelligence collective »

Organisée en partenariat avec la Direction des recherches, des études, de l'évaluation et des statistiques (DREES), la 4^{ème} édition du colloque du HDH « Données de santé et intelligence collective », s'est tenue le 7 décembre dans les locaux de Parisanté Campus en présence de 400 participants.

L'occasion de revenir sur les réalisations concrètes du HDH au terme de ses trois premières années d'activité, en compagnie du ministre de la Santé et de la Prévention, François Braun ; mais aussi d'échanger sur l'usage des données de santé, notamment en matière de prévention, et les freins à lever.

 400 participants pour la 4^{ème} édition du 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.

195
dossiers expertisés
par le CESREES
en 2022

259
dossiers déposés
au HDH en 2022
pour obtenir une
autorisation de la
CNIL

Plusieurs contributions du Health Data Hub pour définir et mettre à jour des procédures simplifiées

Les référentiels relatifs aux accès précoces et compassionnels

Le 22 septembre 2022, la CNIL a publié deux référentiels destinés aux laboratoires pharmaceutiques qui exploitent un médicament autorisé à titre précoce ou compassionnel dès lors qu'un suivi des patients est requis par l'ANSM.

Grâce à ces procédures simplifiées, les professionnels qui respectent les conditions énoncées dans ces référentiels peuvent traiter ponctuellement des données personnelles après avoir adressé un engagement de conformité à la CNIL.

Comme le veut la loi, avant de publier ces référentiels, la CNIL a saisi le HDH le 15 février 2022 pour qu'il se prononce. Le HDH a alors analysé les projets de textes sur le plan juridique, en lien avec son conseil d'administration. Cette contribution a été transmise à la CNIL pour qu'elle puisse ajuster les textes, en tenant compte des besoins des acteurs.

Les méthodologies de référence (MR) 005 et 006

La CNIL a lancé le 10 janvier 2022 une consultation visant à faire évoluer les MR-005 et MR-006 bien connues des acteurs de l'écosystème car elles facilitent l'accès aux données hospitalières (PMSI) sans autorisation de la CNIL. Compte tenu de la réticence des acteurs quant aux évolutions proposées, le HDH a proposé, comme d'autres, que les MR-005 et MR-006 soient conservées en l'état et qu'une discussion soit ouverte pour élaborer une MR propre aux données de la base principale du SNDS.

Tenant compte de ce retour, la CNIL a donc proposé une simple mise à jour des MR-005 et MR-006 intégrant les évolutions juridiques qui s'appliquent au traitement des données du SNDS

et de sa doctrine. Un groupe de travail institutionnel a été créé pour s'accorder sur ces ajustements. Ses travaux se poursuivront début 2023 pour que les nouveaux textes soient publiés avant la fin du premier semestre.

En parallèle, des projets de MR reprenant certaines propositions de cette concertation seront élaborés pour la base principale du SNDS. Des groupes de travail se tiendront avec les acteurs de l'écosystème, la DREES, la CNAM, et l'Agence technique de l'information sur l'hospitalisation (ATIH) avant que les futures MR-007 et MR-008 ne soient publiées. Ce projet permettra, comme le souhaitent ces acteurs, de co-construire les procédures simplifiées avec la CNIL pour qu'elles répondent à leurs besoins.

Le référentiel relatif à l'échantillon du Système national des données de santé (ESND) et aux données agrégées du SNIIRAM

L'échantillon du Système national des données de santé (ESND) a été créé par la CNAM en juin 2022 pour remplacer l'actuel échantillon généraliste de bénéficiaires (EGB). Il rassemble les parcours de soins en ville (base du SNIIRAM) et de l'hôpital (base du PMSI) de 2% de la population présente dans la base principale du SNDS. Il repose sur la même architecture pour que les programmes informatiques d'analyse de données puissent être utilisés dans les deux cas.

Le HDH a saisi la CNIL en mai 2022 pour adapter le référentiel d'accès simplifié à l'EGB aux caractéristiques de ce nouvel échantillon, sur la base d'un dossier élaboré avec la CNAM. Publié au JORF le 4 février 2023, le référentiel ajusté comporte quelques ajustements mineurs. Pour accéder à l'ESND, chaque porteur de projet doit désormais préciser et justifier la pertinence de la profondeur historique demandée, parmi les deux proposées, ainsi que l'utilisation de l'un des identifiants potentiels. Il doit par ailleurs prendre des mesures

appropriées pour rendre publique l'information relative à l'étude menée.

Pour que les porteurs de projets puissent mobiliser cette procédure dans les meilleurs délais, le HDH a mis à jour sa documentation pédagogique.

Une offre de service qui n'a cessé de s'améliorer en 2022

Deux nouveaux guides sur l'anonymisation

Pour dissiper toute confusion entre donnée anonyme et donnée pseudonymisée, le HDH a publié deux nouveaux guides sur l'anonymisation. Une donnée simplement déidentifiée - c'est-à-dire dont les données directement nominatives ont été retirées - n'est en effet pas toujours une donnée parfaitement anonyme. Un guide a donc été mis en ligne en septembre 2022 pour aider les acteurs à comprendre les notions d'anonymisation et de pseudonymisation. Une fois cette distinction faite, il appartient à chacun de s'assurer du caractère anonyme des données de santé qu'il souhaite traiter et d'en documenter l'analyse. C'est pourquoi un deuxième guide a été publié. Il permettra d'évaluer le risque de réidentification d'un jeu de données et d'attester que le risque est bien nul.

Un starter-kit MR-004

Le HDH a également engagé des travaux pour mettre à la disposition des porteurs de projets un starter-kit dédié aux études réalisées dans le cadre de la méthodologie de référence MR-004. Il s'agit d'une procédure d'accès simplifiée aux données de santé à caractère personnel permettant d'éviter une demande d'autorisation. Toutefois, il est toujours indispensable que l'étude, l'évaluation et la recherche présentent un intérêt public. En contrepartie de cet allègement de formalités, le responsable de traitement s'engage à respecter toutes les

obligations du référentiel encadrant cette procédure. Cette procédure largement utilisée par l'écosystème a servi pour 1096 études enregistrées sur le répertoire public des projets du HDH, en 2022. Pour autant, les porteurs de projets n'identifient pas toujours bien les obligations associées à cette procédure ni les éléments à documenter. Le HDH mettra en ligne en 2023 un starter-kit regroupant différents supports, dont un guide mode d'emploi, le template de protocole et la grille de conformité à cette MR.

Une vingtaine de formations

Enfin, en 2022 le HDH a assuré une vingtaine de formations sur les démarches réglementaires d'accès aux données, sur la base principale du SNDS et ses modalités d'accès (en partenariat avec la CNAM), et sur la plateforme technologique du HDH. Ces sessions, ouvertes à tous grâce au catalogue de formation en ligne, ont également été dispensées auprès des startups de Parisanté Campus dont le HDH est membre fondateur.

Pour enrichir cette offre, le HDH a commencé, au second semestre, à concevoir des Massive Open Online Course (MOOC). Cette initiative se poursuivra en 2023 avec un Data Challenge mis en ligne au premier trimestre.

Action 2.2 Renforcer et maintenir à l'état de l'art l'offre technologique du Health Data Hub

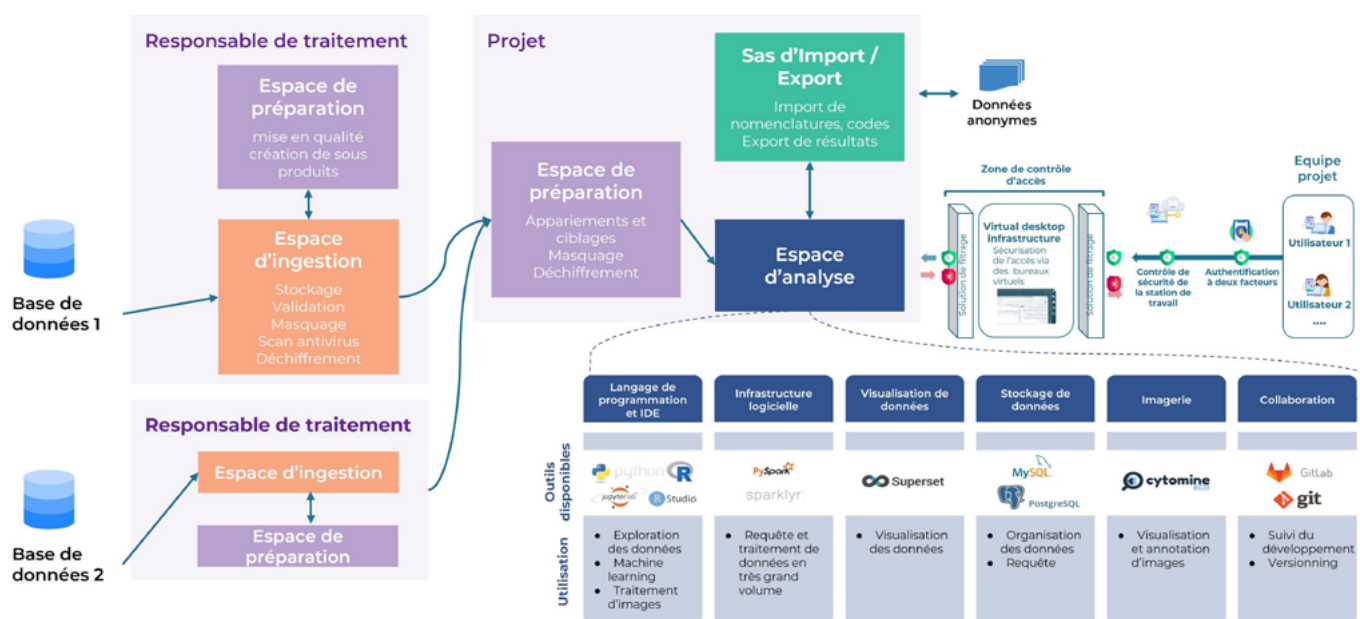
Un « espace d'analyse » sécurisé sur une plateforme technologique

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.

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 HDH a été défini, en respectant les principes de défense en profondeur, dans le cadre d'une démarche conduite en collaboration avec l'Agence nationale de la sécurité des systèmes d'information (ANSSI) et le fonctionnaire sécurité des systèmes d'information (FSSI) du ministère de la Santé et de la Prévention. La sécurité de la plateforme technologique est, en outre, testée et revue périodiquement par le biais d'audits techniques menés par des prestataires spécialisés en sécurité des systèmes d'information (PASSI) indépendants qualifiés par l'ANSSI. Deux ont déjà été réalisés avec succès depuis sa mise en œuvre. Enfin, l'ANSSI et la CNIL participent à l'homologation de chaque nouvelle version de la plateforme technique.

Focus : comment se compose la plateforme technologique ?



Zoom sur l'espace d'analyse sécurisé (SOC)

Afin de garantir le plus haut niveau de sécurité des données de santé, le HDH met en place son centre opérationnel de sécurité (SOC), qui vise à détecter au plus tôt les comportements anormaux afin de mettre en œuvre une réponse adaptée aux incidents de sécurité, aussi bien au sein de la plateforme technologique du HDH que vis-à-vis de son patrimoine informationnel.

Le centre opérationnel de sécurité permet de poursuivre six missions permettant d'appréhender les menaces cyber dans leur globalité.

- Anticiper : analyse de la menace afin d'anticiper les scénarios d'attaque, pour être en mesure de les détecter et d'y répondre au plus vite. Un service de veille pour la sécurité assure la surveillance sur le réseau internet et le réseau internet sombre afin de détecter au plus tôt les éventuelles préparations d'attaques dirigées contre les actifs du Health Data Hub.
- Détecter : gestion et corrélation des traces afin de détecter au plus tôt toutes tentatives d'attaques et ou comportements suspects vis-à-vis des actifs du Health Data Hub.
- Remédier : gestion des incidents de sécurité, mise en place d'une posture adaptée à la menace détectée.
- Améliorer : mise en place en continu de nouvelles règles de détection en cohérence avec l'évolution des fonctionnalités de la plateforme et de la menace.
- Investiguer : possibilité de faire appel à une cellule d'expertise supplémentaire afin de mener des investigations post-incident.
- Sensibiliser : organisation de campagnes de sensibilisation en présentiel ou en e-learning pour tous les acteurs qu'ils soient employés du HDH, utilisateurs de la plateforme, chercheurs.

En 2022, le chantier s'est principalement concentré sur l'élaboration du cahier des charges et le lancement du marché. Le dernier étant notifié, les travaux ont pu démarrer pour mettre en production le service dès 2023.

Participation aux travaux de refonte du référentiel de sécurité

Institué le 22 mars 2017, le référentiel de sécurité du SNDS constitue un socle d'exigences solide pour la sécurité des SI et la protection des données. Il permet de répondre de façon spécifique et avec une logique « sur-mesure » à la sensibilité des données du SNDS et au risque de réidentification des citoyens.

Le HDH met en œuvre les principes de ce référentiel de sécurité du SNDS qui spécifie notamment que les données doivent toujours être pseudonymisées.

Le dernier texte applicatif à revoir, depuis la loi relative à l'organisation et à la transformation du système de santé (OTSS), est un document technique, porté par arrêté pris après avis de la CNIL. Le HDH a contribué à sa mise à jour. Le ministère de la Santé et de la Prévention y a en effet travaillé en s'appuyant sur la CNAM et le HDH. Le projet d'arrêté a été soumis à la CNIL début 2023.

Une offre technologique en évolution constante

Traiter des données de santé à travers des approches analytiques innovantes nécessite de relever de nombreux défis technologiques. 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 qu'ils puissent réaliser leurs études, le HDH met à leur disposition un espace d'analyse dans lequel ils peuvent recevoir, stocker, traiter en mode collaboratif et visualiser les données de santé nécessaires à leur projet, en toute sécurité.

L'offre logicielle du HDH comprend à ce jour :

- Les environnements de développement intégré (IDE) JupyterLab et RStudio, proposant de travailler respectivement avec les langages de programmation Python et R.
- Les outils de collaboration Git et Gitlab.
- L'outil de calcul distribué Spark pour Jupyterlab, permettant d'effectuer des analyses complexes à grande échelle (Big Data).
- L'outil de visualisation de données Superset.
- Les systèmes de gestion de bases de données PostgreSQL et MySQL.
- Les outils de visualisation et d'annotation d'images Cytomine (JPEG, PNG, TIFF, etc.) et OHIF Viewer (DICOM).
- L'outil d'analyse de données Atlas, permettant de créer des cohortes et d'effectuer des études scientifiques poussées sur des données converties au format OMOP.

Dans le cadre du Plan d'action pour un gouvernement ouvert, la direction technique du HDH met également à disposition le code source de transformation des données liées à leur ingestion au sein de la plateforme technologique.

Zoom sur le plan d'action réversibilité

La plateforme technologique du HDH est entièrement administrée par des agents du HDH, dont les rôles et les responsabilités sont bien définis. Ses composants industriels ont été choisis pour leur performance, leur sécurité et leur élasticité.

Elle a été conçue dès sa création pour répondre à trois enjeux :

- 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 comme l'intelligence artificielle et nécessitant de grandes capacités de calcul et de stockage.

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

En l'absence de solution souveraine permettant de répondre totalement à ces trois objectifs, la plateforme technologique s'appuie sur des solutions complémentaires. Un plan de migration vers une solution souveraine a été défini car la plateforme a été conçue pour être réversible by design et réduire au maximum les adhérences avec le fournisseur de services cloud actuel. Cela lui permettra d'être redéployée sur une autre solution du même niveau de maturité le moment venu.

C'est dans cette optique que le HDH étudie les pistes de réversibilité depuis 2019 et a établi avec la DINUM une liste des services indispensables pour faire fonctionner sa plateforme en toute sécurité. La recherche d'une solution, répondant aux critères actuels, a été lancée au plus tôt. En 2022, une troisième étude de marché a permis de comparer les solutions cloud susceptibles de répondre aux enjeux du Health Data Hub. Cette étude a été transmise à la DINUM, au ministre chargé de la Transition numérique et des Télécommunications, et à la Direction générale des entreprises (DGE). Elle confirme toujours que pour l'instant aucune offre de marché souveraine ne couvre les besoins fonctionnels et de sécurité permettant à la plateforme de répondre aux exigences légales requises. Les rencontres avec les acteurs français et européens témoignent néanmoins d'importantes avancées au fil des années.

Confiant dans les capacités industrielles françaises et européennes, le HDH s'implique, avec la DINUM, pour faire évoluer les solutions existantes. Il se met à la disposition des entreprises pour partager avec elles ses besoins essentiels. Par ailleurs, il fait aussi appel à des solutions françaises pour les projets qu'il porte actuellement, qu'il s'agisse de cybersécurité ou de projets d'outils permettant de faciliter les demandes d'exercice des droits ou d'appariements.

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

Une animation ouverte de l'écosystème science

Pour faciliter la recherche, réduire les délais des projets de recherche et mutualiser les efforts de data management, le HDH a mis en place, dès 2019, une démarche science ouverte. Une démarche à laquelle contribuent de nombreux partenaires tels que la DREES, la CNAM, l'ATIH, Santé publique France, le ReDSiam ou les agences régionales de santé (ARS). Les chantiers menés sont cohérents avec les recommandations du député Eric Bothorel dans son rapport de mission parlementaire. Ils s'inscrivent dans la feuille de route de l'administrateur ministériel des données, algorithmes et codes, source du ministère de la Santé et de la Prévention.

C'est ainsi que le 14 décembre 2022, le HDH organisait un événement dans le cadre du programme « bibliothèque ouverte d'algorithmes en santé » pour soutenir le développement d'algorithmes facilitant l'exploitation des données de la base principale du SNDS et leur mise en open source. Cette rencontre était l'occasion de dévoiler les cinq lauréats de l'AMI BOAS et de revenir plus largement sur la démarche science ouverte du HDH et de ses partenaires. Des nouveautés ont également été présentées comme le guide d'initiation à la base principale du SNDS, les scripts de transformation de la base principale du SNDS au format OMOP-CDM, ou le script de pseudonymisation utilisé par la plateforme technique du HDH.

La démarche science ouverte comporte quatre axes de travail :

1. L'axe « connaissances » a pour but de mieux comprendre les données et l'écosystème des utilisateurs du SNDS. Divers outils, utilisés plus de 500 fois par jour, ont été développés et ouverts dans le cadre de ce chantier tels que la documentation

collaborative du SNDS. Quatre nouvelles ressources ont été ouvertes en 2022 :

- Un catalogue de métadonnées permettant de mieux cerner le contenu des bases de données.
- Un guide d'initiation de la base principale du SNDS à destination des nouveaux utilisateurs de cette base.
- Une cartographie de l'écosystème SNDS à laquelle se sont ajoutées plusieurs sources de données (publications scientifiques en français et anglais, thèses, etc). Cette dernière a fait l'objet d'une présentation au congrès Emois 2022.
- La documentation permettant de standardiser la base principale du SNDS au modèle de données commun OMOP.

2. L'axe « données » a pour but de favoriser et promouvoir le partage de données, avec en 2022 les avancées suivantes :

- L'ouverture des scripts permettant de transformer les données au format OMOP-CDM pour la base principale du SNDS.
- La préparation de nombreux data challenges.
- La plateforme de données pour le partage en open data dont la mise en production est prévue en 2023.

3. L'axe « programmes » a pour objectif de mutualiser les efforts de conception de scripts et d'algorithmes sur la base principale du SNDS. En 2022, un groupe de travail regroupant le HDH, le ReDSiam et la CNAM a travaillé sur l'ouverture du script de ciblage des diabétiques dans la base principale du SNDS, prévue pour 2023. L'appel à manifestation BOAS a permis de sélectionner 15 lauréats au total, dont la dernière vague de sélection a eu lieu en décembre 2022. Des réflexions sur le futur de l'AMI BOAS ainsi que sur la structuration de la bibliothèque BOAS sont en cours.

4. L'axe « animation » a pour but de partager les résultats et scripts ouverts à l'écosystème. Quatre meetups ont eu lieu en 2022, avec à chaque fois une centaine d'inscrits. Plus de 2020 membres se sont inscrits en 2022 pour en suivre un de ces événements.

Un groupe de travail pour favoriser l'usage du traitement automatique du langage en santé

Le traitement automatique du langage naturel (TAL) est un des champs de recherche les plus actifs de l'intelligence artificielle. Il a pour but d'apprendre à une machine à comprendre le langage humain. Dans le domaine de la santé, beaucoup d'informations se présentent sous forme de textes libres, notamment les comptes-rendus médicaux des patients, ce qui constitue une matière très riche mais coûteuse à exploiter et à interpréter.

Avec le TAL, il serait possible de les récupérer de manière automatique pour faciliter l'interprétation des médecins et les aider dans leur diagnostic. Il serait possible par exemple d'imaginer des algorithmes permettant d'identifier dans ces comptes-rendus, quelles patientes sont enceintes, lesquelles ont déjà fait une échographie et si cette échographie est normale. Cela représenterait non seulement un gain de temps mais aussi une précision accrue pour le système de santé et ses professionnels.

Cependant, de tels algorithmes, à la fois généraux et performants en open source n'existent pas. C'est pour favoriser leur développement grâce au TAL, que le coordinateur de la stratégie nationale pour l'intelligence artificielle (SNIA) a mandaté le Health Data Hub en avril 2022. Sa mission : monter un groupe de travail qui va identifier les actions stratégiques pouvant bénéficier d'un financement national.

Le HDH a ainsi réuni 10 chercheurs spécialisés dans le traitement automatique du langage naturel. Le premier constat qu'ils ont fait est que les algorithmes de TAL s'appuient sur les comptes-rendus médicaux, qui contiennent des données sensibles et potentiellement identifiantes. Ainsi, la pseudonymisation des comptes-rendus, qui vise à y enlever les informations directement identifiantes (nom, prénom, etc.), est un préalable indispensable à tout projet de TAL. Ils ont donc suggéré de développer des outils permettant de pseudonymiser automatiquement les textes afin de pouvoir les réutiliser pour entraîner les algorithmes. En effet, de tels outils de pseudonymisation transposables et utilisables par tous n'existent pas. La plupart des centres hospitaliers n'ont tout simplement pas les ressources pour en développer.



4

ressources ouvertes en 2022

2020

participants aux meetups

15

membres meetup



Pour répondre à cette problématique et plus largement concourir à la réutilisation des données médicales grâce au TAL, le groupe de travail a proposé deux actions qui pourraient être déployées dès 2023 :

- Un guide méthodologique de référence pour assister les chercheurs souhaitant développer un algorithme de pseudonymisation des comptes rendus. Ce guide serait complété par un outil en accès libre permettant d'évaluer la performance de ces algorithmes.
- Une base de comptes-rendus de patients fictifs permettant d'entraîner ces nouveaux algorithmes et ainsi améliorer leur performance sans risque de réidentification.

Le HDH présentera ces actions au coordinateur de la SNIA, début 2023.

Action 2.4 Diffuser l'offre de service au sein de l'écosystème : organismes de recherche, industriels, partenariats publics et privés, etc

Mettre en valeur les activités de soutien à l'innovation

Chargé de soutenir l'innovation, le HDH se penche en permanence sur les dernières avancées en science des données et ouvre régulièrement des appels à projets destinés à l'écosystème de la santé. Ces multiples appels à projets visent à appréhender la diversité des données et des usages (IA, chaînage, ciblage, données de vie réelle, etc) ainsi que tous les enjeux de santé publique (cancérologie, essais cliniques, prévention, etc).

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

- **BOAS (vagues 2 et 3)** : constitution d'une bibliothèque ouverte d'algorithmes de cibrages visant à soutenir le développement et la validation d'algorithmes de ciblage dans le SNDS (deux fois cinq lauréats).
- **France-Québec - IA en santé** : appel à projets transatlantique franco-québécois, en partenariat avec le Fonds de recherche Québec-santé (deux lauréats).
- **Accompagnement et soutien à la constitution d'entrepôts de données de santé hospitaliers** : en partenariat avec la DGOS et Bpifrance, ce dispositif de la stratégie d'accélération « santé numérique » vise à constituer un réseau d'EDS hospitaliers.
- **Data Challenges en santé (vague 3)** : compétitions en data science visant à répondre à des enjeux médicaux en un temps donné. Cet AAP a été lancé en partenariat avec la Délégation ministérielle au Numérique en Santé et Bpifrance.
- **AAP CSF-ITS** : appel à projets dont le HDH est partenaire avec le Comité stratégique de filière des industries et technologies de santé (CSF-ITS) « data, IA et parcours de santé ». L'objectif est d'analyser des parcours de santé pour en cibler les ruptures et proposer des innovations (cinq lauréats).
- **AAP GD4H** : Green Data For Health (GD4H) et le HDH ont organisé un appel à projets commun afin d'améliorer la mobilisation des bases de données environnementales et de santé au service des projets de recherche et d'innovation (quatre lauréats retenus).

■ **AMI Unicancer (vague 2)** : en partenariat avec Unicancer et la Ligue contre le cancer, lancement d'un appel à manifestation d'intérêt visant à accélérer la recherche à partir de données de vie réelle en cancérologie en construisant des bases de données de référence en la matière (quatre lauréats).

Données de vie réelle : lancement d'une réflexion sur leur mobilisation

Les « données de vie réelle » désignent des données qui reflètent la pratique courante, par opposition à des données collectées dans un cadre expérimental. Elles peuvent provenir des dossiers des patients ou de collectes spécifiques (registres, cohortes, études de pharmacovigilance, etc). La richesse de ces données est une opportunité à ne pas manquer pour l'écosystème de la santé. Afin d'encourager les projets innovants, le HDH a engagé une réflexion sur la mobilisation de ces données dans le processus de développement des médicaments, pour développer par exemple une médecine ciblée ou pour optimiser les essais cliniques. Après avoir rencontré de nombreux acteurs de la filière (industriels, startups et partenaires technologiques), les équipes construisent actuellement un premier cas d'usage test, en prévision d'un futur appel à projets.

Prévention en santé : explorer le potentiel des données de santé pour les politiques publiques

Définie par l'Organisation mondiale de la Santé comme « l'ensemble des mesures visant à éviter ou réduire le nombre et la gravité des maladies, des accidents et des handicaps », la prévention est l'une des principales pistes pour améliorer la santé des citoyens dans la durée. Selon l'Organisation de coopération et de développement économique (OCDE), en 2021, plus de 3 millions de décès auraient pu être évités par des mesures de prévention. L'exploitation des données de santé peut contribuer à mieux comprendre les déterminants

(socio-environnementaux, comportementaux, génétiques, biologiques, etc) de la santé et ainsi éclairer les décideurs pour optimiser l'efficacité des politiques de prévention à mettre en œuvre. C'est pourquoi le HDH explore les moyens par lesquels mobiliser ces données.

Certains projets déjà accompagnés par le HDH répondent partiellement à cet enjeu, et peuvent constituer des premiers cas d'usages de prévention primaire et secondaire. C'est le cas du projet « parcours IDM en Île-de-France », qui évalue les parcours de soin après un infarctus du myocarde, en termes de qualité, de sécurité et de pertinence. C'est aussi le cas du projet Deep-piste, qui vise à développer un algorithme pour analyser des mammographies et définir les facteurs de risque.

Il permettrait de personnaliser le dépistage, dans une démarche de prévention d'évolution péjorative des cancers.

Afin d'aller plus loin dans cette réflexion et favoriser le développement de cas d'usage innovants, le HDH envisage de lancer des appels à projets dédiés à la prévention.

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

En 2022, le Health Data Hub s'est associé au Green Data for Health, dispositif du Commissariat général au développement durable (CGDD) visant à favoriser le recueil de données environnementales. Ensemble, ils ont lancé un appel à projets conjoint « La donnée pour la recherche et l'innovation en santé-environnement ». Centré sur l'impact du bruit, des bassins industriels et des perturbateurs endocriniens sur la santé, il va faciliter les croisements de données au service de la prévention. En enrichissant le SNDS avec des données de divers domaines de la vie et en les croisant avec des données de contexte, on pourra en effet étoffer nos connaissances et mieux caractériser les facteurs de risque.

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

Créé en 2021, PariSanté Campus (PSC) a pour ambition de devenir un pôle d'excellence unique au monde, pour la recherche, l'innovation, la formation et l'entrepreneuriat en santé numérique. Confondé par l'Inserm, l'Inria, l'Université PSL, l'Agence du numérique en santé et le HDH, ce campus de rang mondial va s'installer en 2028 sur le site de l'ancien hôpital du Val-de-Grâce à Paris.

Sa force sera de donner l'impulsion nécessaire pour que les opérateurs et les acteurs puissent se rencontrer dans un lieu ouvert. Ces synergies devraient décupler la capacité collective à déployer des projets communs à fort impact, identifiables sur la scène internationale.

En un an, le partenariat avec PariSanté Campus a offert de nombreuses opportunités de collaboration, qui ont permis de mettre à profit l'écho qu'offre PariSanté Campus pour organiser plusieurs événements à destination de l'écosystème national de la e-santé. En effet, tout au long de l'année, le HDH a organisé 15 événements avec la coopération de PariSanté Campus, dont quatre labellisés par la présidence française du Conseil de l'Union européenne (PFUE), contribuant ainsi à la diffusion d'une culture de la donnée de santé. En 2022 également, le HDH a développé une offre spécifique à destination des startups sélectionnées par AMI et qui sont incubées au sein de PSC.

Le Health Data Hub, membre fondateur de l'association Filière intelligence artificielle et cancer (FIAC)

En 2022, deux membres industriels supplémentaires ont rejoint le projet. Pour rappel, la Filière IA et cancers est une association qui fédère les acteurs publics et privés au sein d'une initiative nationale. Le projet souhaite profiter à l'ensemble de l'écosystème en oncologie, et permettre le développement de solutions thérapeutiques et diagnostiques innovantes. L'association réunit à ce jour, l'Institut national du cancer, le HDH, l'Alliance pour la Recherche et l'Innovation des Industries de Santé (ARIIS), France Biotech et neuf industriels de santé : Amgen, AstraZeneca, Bristol-Myers Squibb, Janssen-Cilag, MSD, Novartis, Pfizer, Pierre Fabre, et Roche Diagnostics. Cinq projets reposant sur la réutilisation de données de données en cancérologie (PRC) ont été approuvés par les membres fondateurs de l'association. Ils concernent notamment les cancers du poumon, broncho-pulmonaire et bronchique.



5

membres fondateurs
de PariSanté Campus,
dont le HDH

Un service de requête à la demande

Grâce à l'outil de « requête à la demande » que la HDH expérimente depuis 2021, les acteurs peu habitués aux données de santé pourront disposer, sans aucune formalité réglementaire, d'éléments chiffrés issus du SNDS mais non disponibles en open data. Les éléments produits et la documentation associée seront ensuite accessibles à tous puisque le HDH a la volonté d'élargir la connaissance ouverte relative à la base principale du SNDS.

Pour préciser le périmètre de cette offre, des travaux ont eu lieu avec différentes associations de patients telles que France Vascularites, l'association Journée Nationale de l'Audition ou l'Association des Porteurs de Dispositifs Electriques Cardiaques, avec la participation de chercheurs en sciences humaines et sociales, ainsi que des représentants de journalistes comme le Quotidien du médecin.

En 2022, cette initiative s'est élargie à de nouveaux partenaires tels que l'afa Crohn RCH France, l'Association Porteurs de défibrillateurs cardiaques et l'Association francophone pour vaincre les douleurs (AFVD), tandis que le travail autour des requêtes a suivi son cours, pour que ce service puisse être lancé en 2023.

13
membres composent
l'association FIAC, dont
le HDH

ENJEU 3

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

Le Health Data Hub est un service qui s'adresse tant aux acteurs qui collectent et/ou utilisent les données, qu'à la société civile. Pour atteindre son plein potentiel, le HDH doit s'adapter aux attentes d'un écosystème qui évolue désormais à l'échelle européenne voire internationale.

Action 3.1 Alimenter la communauté scientifique internationale

Un conseil scientifique consultatif ouvert à l'international

En janvier 2021, un conseil scientifique consultatif a été créé avec pour vocation d'accompagner la direction du Health Data Hub dans toutes ses missions, de rendre des avis, faire des propositions et l'éclairer dans ses décisions. Y siègent 28 personnalités issues de nombreuses disciplines telles que la médecine, la pharmacie, la génétique, l'épidémiologie, la statistique, l'économie de la santé, juridique, les sciences des données et intelligence artificielle ou éthique. Dès la fin de l'année 2021, le conseil scientifique s'était élargi pour accueillir sept membres internationaux et la rentrée 2022 a donc été l'occasion de tenir, dans les locaux de PariSanté Campus, la première réunion en présentiel du conseil depuis sa création. Ainsi, le 3 février 2022, les membres du conseil ont pu prendre connaissance du bilan 2021 du Health Data Hub et apporter leur regard sur les grandes activités en cours, telles que les partenariats académiques et internationaux, les nouvelles activités européennes, les appels à projets en élaboration, l'accès aux données, etc.

Durant l'année, les membres du conseil consultatif ont aussi appuyé les équipes du Health Data Hub dans le déroulement de ses activités. Par exemple, le HDH a organisé deux

délégations en 2022, la première en Israël, la seconde au Japon, dans l'optique de développer des partenariats stratégiques entre acteurs français et internationaux. Pour ces rendez-vous, le HDH a souhaité s'entourer d'experts seniors et s'est donc naturellement tourné vers les membres du conseil scientifique. Martine Gilard et Alain Livartowski se sont joints à la délégation envoyée à Tokyo en octobre 2022, qui a ainsi pu bénéficier de leurs expertises respectives en radiologie et oncologie.

Enfin, 2022 a aussi été l'occasion pour le conseil scientifique de lancer ses propres travaux sur des sujets de prospective majeurs identifiés avec la direction du HDH. Les membres du conseil scientifique ont en effet la possibilité de former des groupes de travail pour se prononcer sur des questions requérant une expertise spécifique. Sept thématiques d'avenir, chacune portée par un membre du conseil ou un expert désigné, ont été retenues et présentées lors d'une réunion en juillet 2022 : l'éthique et la santé ; les registres ; les données sociales ; l'anatomopathologie ; la psychiatrie et la santé mentale ; les données génomiques ; le cancer.

La thématique des données génomiques, portée par Thomas Bourgeron et Ségolène Aymé, est ainsi apparue comme clé en raison de l'impact indéniable de la génétique sur la vulnérabilité des patients face à certaines maladies. De même, le sujet de la psychiatrie et la santé mentale, porté par Hossein Khonsari, directeur médical du Health Data Hub, apparaît comme

un enjeu prioritaire de santé publique alors que l'organisation des soins ainsi que la nature des affections psychiatriques, combinant des composantes médicales, sociales et médico-sociales, ont pour conséquence de produire des données hétérogènes et dispersées. Afin d'identifier les potentielles actions à porter par le HDH sur ces axes stratégiques, des travaux d'état des lieux des bases de données et de leurs utilisations en France et dans le monde seront lancés fin janvier 2023, les rapports finaux sont attendus pour le premier semestre. Des groupes de travail seront ainsi lancés sur la base de ces travaux au second semestre 2023.

Le sujet « éthique et santé » fait l'objet d'un groupe de travail porté par Ségolène Aymé et a pour ambition d'explicitier les grands enjeux sociétaux auxquels répondent les services du HDH afin d'améliorer la lisibilité des missions du HDH pour les membres de la communauté des utilisateurs potentiels de ses services. Le rapport final, explicitant les choix du HDH en termes de valeurs à servir et ayant l'ambition de servir de support de communication et de guide pour l'action, sera finalisé au premier semestre 2023.

D'autres thématiques identifiées en juillet 2022 nécessitent, quant à elles, d'être instruites davantage afin d'identifier les axes les plus pertinents à aborder lors des futurs groupes de travail. Ainsi, le sujet des données sociales, porté par Michelle Kelly-Irving, a fait l'objet de temps d'échange avec des acteurs clés de l'écosystème. Il apparaît évident que ces données sont indispensables pour comprendre comment les inégalités sociales se transforment en inégalités de santé. De nouveaux échanges préparatoires puis le lancement d'un groupe de travail sont prévus pour le premier trimestre 2023.

Le sujet des registres, porté par Martine Gilard,

a fait l'objet de premiers travaux du HDH en janvier 2023 dans le but d'initier la réflexion suite aux rapports du Haut Conseil en santé publique de septembre 2021 et de l'Académie nationale de médecine de janvier 2022. Des rendez-vous avec des acteurs clés, dont la direction générale de la santé (DGS) et Santé publique France, suivis du lancement d'un groupe de travail, sont prévus pour le premier trimestre 2023.

Le sujet du cancer, porté par Alain Livartowski, apparaît également comme stratégique du point de vue de la quantité et de la diversité des données qu'il génère. Or, dans l'ambition d'améliorer la médecine de précision, il est nécessaire de parvenir à réunir des données récentes, multicentriques et de natures variées qui soient utilisables et interopérables. Pour mieux comprendre quel éclairage le futur groupe de travail pourrait apporter sur le sujet, le conseil scientifique envisage des travaux préparatoires durant l'année 2023.

Par ailleurs, une réflexion sur la thématique de l'anatomie et cytologie pathologiques, porté par le P^r Philippe Bertheau de l'AP-HP, membre extérieur du conseil scientifique, est aussi en cours d'élaboration et des travaux sont prévus pour 2023.

Le conseil scientifique a ainsi clôturé son année par une réunion de travail tenue à Paris, le 29 novembre 2022. Les directrices et directeurs du Health Data Hub ont eu l'occasion de présenter au conseil le programme de travail pour 2023, puis des ateliers au cours desquels les membres ont été invités à partager leur expérience et réflexion autour de thématiques clés pour le HDH telles que la prévention, l'intelligence artificielle en santé, les données de vie réelle et le partage des données.

L'initiative Data challenge se poursuit en 2022

Depuis 2020, le Health Data Hub établit des partenariats avec des acteurs du monde médical pour organiser des Data Challenges en santé. En 2020, un premier Data Challenge portant sur la détection des lésions du col de l'utérus a été organisé avec la Société Française de Pathologie (SFP). Face au succès de ce premier challenge, le HDH a renouvelé et multiplié l'opération en lançant au printemps 2021 un appel à manifestation d'intérêt destiné aux acteurs du monde médical souhaitant organiser un Data Challenge sur une thématique médicale de leur choix. Six projets ont été sélectionnés à l'occasion de cet AMI et se sont développés au cours de l'année 2022. Lancés courant 2023, ils apporteront une véritable contribution sur des domaines médicaux divers :

- 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.
- 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.
- Le Data Challenge 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.
- Le Data Challenge organisé par la Société Française de Pathologie et les dermatologues pour prédire le risque d'évolution métastatique de mélanomes.
- 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.

- Le Data Challenge de la Société Française de Médecine Nucléaire (SFMN) pour développer un algorithme contribuant au diagnostic des syndromes parkinsoniens aux formes multiples.

L'intérêt suscité par ces Data Challenges va croissant, aussi bien auprès de ceux qui veulent en organiser, qu'auprès des potentiels compétiteurs prêts à relever les défis posés par les experts médicaux. Ils contribuent à l'innovation ouverte grâce au partage de connaissances qui découle de ces compétitions. Grâce au plan « France 2030 », la Délégation ministérielle au Numérique en Santé, Bpifrance et le HDH ont poursuivi ainsi la démarche en lançant, à l'été 2022, un nouvel AAP afin d'établir des partenariats avec des acteurs médicaux désireux d'organiser une compétition Data Challenge sur de nouvelles thématiques médicales d'intérêt.

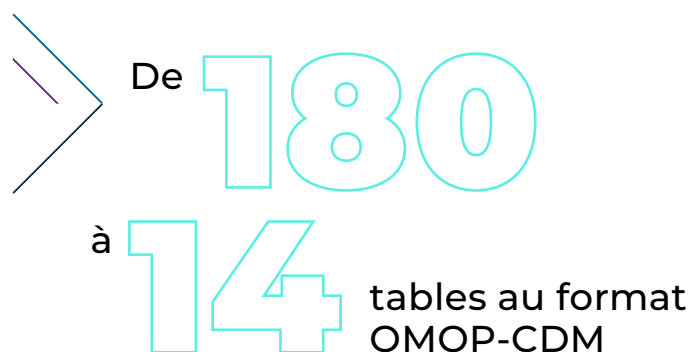
Ce nouvel AAP destiné aux sociétés savantes, établissements de santé, groupements de coopération sanitaire et instituts hospitalo-universitaire avait pour vocation de sélectionner jusqu'à 10 nouveaux projets répondant à des critères scientifiques, techniques et réglementaires rigoureux décrits dans un nouveau cahier des charges. L'objectif étant de sélectionner des projets de haute qualité et d'intérêt clinique documentés qui soient faisables d'un point de vue scientifique et logistique, et dans le respect du RGPD.

Dans ce contexte, le HDH a souhaité mettre à disposition des candidats présélectionnés, un appui juridique et technique afin de consolider leur projet sur les plans scientifique et réglementaire avant sélection. Les lauréats pourront bénéficier d'un soutien financier et de l'accompagnement du HDH pendant toute la durée de leur organisation.

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

Vers une standardisation de la base principale du SNDS pour favoriser sa réutilisation à des fins de recherche

La base principale du SNDS reste encore sous exploitée alors qu'elle constitue l'une des bases de données médico-administratives les plus larges au monde. Il faut en général plusieurs mois, voire années, de montée en compétence, pour en comprendre le schéma, les biais et les limites. Mener des projets à l'échelle européenne avec la base principale du SNDS reste un véritable défi à cause de terminologies propres à la France et d'un schéma de données non standard. Depuis 2020, le HDH standardise ces données et les transforme vers le modèle de données commun OMOP-CDM. Cette standardisation, qui permet de passer de plus de 180 tables à moins de 20, ouvre donc son utilisation à des projets européens. Des internes en médecine ont également travaillé sur les alignements de plus des 70 000 terminologies (CCAM, NABM, LPP, CSARR, etc). Les scripts (ETL) permettant de transformer les données au format OMOP-CDM ont été ouverts sur Gitlab, et une documentation détaillant les règles de transformation a été publiée sur la documentation collaborative du SNDS. Des travaux d'extension du périmètre à tous les bénéficiaires et aux années antérieures à 2019 sont en cours.



Ouverture de l'ETL de transformation

Ouverture des scripts de transformation du SNDS au format OMOP-CDM



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 les Etats membres, la Commission européenne a fait du futur espace européen des données de santé 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. Une initiative qui a pris un nouvel élan avec la proposition d'un règlement pour l'EHDS, dévoilé par la Commission européenne le 3 mai 2022.

Plusieurs instruments de préfiguration du futur espace européen des données de santé avaient 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 dans lequel le Health Data Hub a coordonné cinq partenaires français. Pilotée par le HDH, la contribution française a été particulièrement active sur l'engagement citoyen - groupe de travail 8 - et la gouvernance - groupe de travail 5 - (cf. infra).

En juillet 2022, le HDH a pris la tête du consortium lauréat d'un appel à candidatures 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, et vise à déployer un réseau de plateformes de données dans l'Union européenne. Les travaux ont été lancés en octobre 2022, lors d'un événement organisé par le HDH à Paris dans les locaux de PariSanté Campus. Le HDH a d'ailleurs également été nommé autorité compétente

française pour répondre au financement européen Setting up services by Health Data Access Bodies dans le cadre du programme « l'UE pour la santé » (EU4Health).

Au-delà de ces actions préfiguratrices de l'EHDS, le HDH est partie prenante de plusieurs autres projets européens. Il a ainsi intégré par exemple un très large consortium mené par la Charité Berlin (Allemagne) dans le cadre du projet Testing and Experimentation Facility, Health (TEF Health). Les autres partenaires français impliqués dans le projet sont le Laboratoire national de métrologie et d'essais (LNE), le Commissariat à l'énergie atomique et aux énergies alternatives (CEA), le Centre national de la recherche scientifique (CNRS), le CHU de Grenoble, l'Université de Grenoble Alpes, les Hospices Civils de Lyon, le CHU de Rennes, l'IHU de Strasbourg et l'EIT Hub France. Le but de ce projet est de proposer des services de test à des PME et startups, pour qu'elles puissent accélérer la mise sur le marché de leurs solutions innovantes d'algorithmes et de robotique. Par ailleurs, en tant que lauréat du programme European Health Data & Evidence Network (EHDEN), le HDH travaille à l'alignement des données du SNDS sur le modèle de données OMOP-CDM et collabore avec l'agence européenne du médicament ainsi que quatre établissements de santé pour constituer une base de données multicentrique.

Le HDH participe aussi au conseil scientifique du Data Analysis and Real World Interrogation Network (DARWIN EU), le réseau d'analyse de données, d'expertise et de services, lancé par l'Agence européenne du médicament. Objectif : améliorer la prise de décision tout au long du cycle de vie des médicaments, grâce à l'exploitation de ces données. Enfin, le HDH a organisé et participé à de multiples événements à l'échelle européenne pour présenter ses activités et réalisations, y compris son engagement européen. Ainsi, durant la PFUE, le HDH a organisé cinq événements labellisés.

La contribution du Health Data hub au règlement européen EHDS

Le 3 mai 2022, la Commission européenne dévoilait son projet de règlement pour l'espace européen des données de santé. Un projet qui est actuellement débattu au sein du Conseil de l'Union européenne et du Parlement européen. Le HDH, aux côtés de multiples acteurs publics, contribue à élaborer la position française.

Ce projet de règlement, qui couvre l'utilisation primaire et secondaire des données de santé, constituerait une avancée majeure pour en faciliter l'usage. Il s'inspire en bonne partie du modèle français et semble aborder un certain nombre de points dont la clarification permettra d'accélérer la réutilisation des données de santé en France comme en Europe. Il devrait notamment se traduire par une réduction des délais d'accès en France et une harmonisation des règles. Enfin, il prévoit que tous les États membres mettent en place un organisme responsable de l'accès aux données. Le HDH fait déjà figure de modèle.

Ce projet de règlement propose aussi un cadre pour l'utilisation transfrontière des données de santé, grâce à une nouvelle infrastructure européenne, baptisée «HealthData@EU» et pilotée par le consortium que mène le HDH (cf. infra). Sa vocation est d'aider les acteurs à identifier et à réutiliser des données au niveau européen.

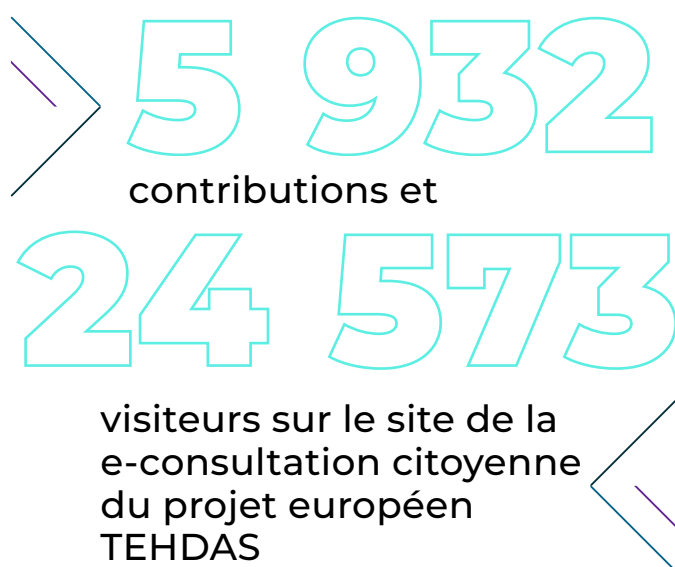
De nouveaux résultats pour l'action conjointe TEHDAS

Lancée en février 2021, l'action conjointe TEHDAS qui mobilise 26 pays, avait pour objectif de soumettre des recommandations à la Commission européenne pour le futur espace européen des données de santé. Pilotée par le HDH, la contribution française a été particulièrement active sur l'engagement citoyen (groupe de travail 8) et la gouvernance (groupe de travail 5).

L'action conjointe a continué ses travaux en 2022 et se terminera en août 2023.

Lancée fin 2021 par le HDH, avec ses partenaires belge et anglais, la consultation en ligne, pour mesurer la perception qu'ont les citoyens des données de santé et de leur utilisation, s'est clôturée en avril 2022. Disponible en français, anglais et néerlandais mais ouverte à tout citoyen européen, cette consultation proposait de l'information sur le partage et l'usage des données en Europe, des questionnaires individuels et un espace de discussion. [Elle a recueilli plus de 6 000 contributions et le site a été consulté par plus de 24 500 visiteurs.](#) Le résultat des contributions a été publié en 2022 et la compilation des recommandations pour le futur EHDS le sera début 2023, après une série de workshops.

De plus, le Health Data Hub et ses partenaires finlandais, espagnol et néerlandais ont réalisé une comparaison internationale des procédures d'accès aux données de santé dans leurs quatre pays. Enfin, basé sur l'expérience du partenariat mis en place entre le HDH et son homologue finlandais (Findata), un guide pour la mise en place de tels partenariats a été publié.



Le projet HealthData@EU pilote, pierre fondatrice du futur EHDS

En 2021, le HDH a formé un consortium pour répondre à l'appel à candidatures de la Commission européenne et déposé sa candidature en janvier 2022. Objectif : mettre en place une version test de l'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 ou le Centre européen de prévention et de contrôle des maladies (ECDC) ainsi que des infrastructures de recherche (BBMRI, Elixir, eBrains).

Officiellement retenu en juillet 2022, ce projet a été officiellement lancé les 11 et 12 octobre 2022 à Parisanté Campus. Un événement auquel le HDH a convié les partenaires du consortium, ainsi que la Commission européenne. Pendant deux jours, les participants ont discuté du plan de travail et tissé des liens qui faciliteront le déroulement du projet.

Les travaux ont pour but de construire une première version de l'infrastructure transfrontière HealthData@EU, prévue dans le projet de règlement pour l'EHDS. Ce réseau connectera tous les partenaires fournisseurs de données (nœuds de données), ainsi que les services centraux mis en place au niveau de la Commission européenne. Grâce à un portail unique fourni par la Commission européenne, ce réseau permettra aux chercheurs souhaitant monter un projet d'interroger les catalogues de données de tous les nœuds du réseau et d'en demander l'accès via un formulaire unique. L'idée est de construire un parcours utilisateur aussi simple et fluide que possible pour les porteurs de projets.

Cinq cas d'usage proposés lors de la candidature ont été retenus par la Commission européenne en septembre 2022 pour tester le réseau et 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 :

- **Le Centre européen de prévention et de contrôle (ECDC)** cherchera à démontrer la possibilité d'utiliser l'EHDS pour assurer la surveillance de la résistance antimicrobienne.

Données : Finlande, Belgique, Croatie, Norvège

- **L'Agence européenne des médicaments** devra identifier les risques de troubles de la coagulation chez les patients atteints de Covid-19.

Données : France, Finlande, Belgique, Danemark, Croatie, réseau Darwin EU

- **L'institut belge Sciensano** comparera l'utilisation de tests, les hospitalisations et l'adhésion à la vaccination chez les sous-populations vulnérables.

Données : France, Finlande, Belgique, Danemark, Croatie, Hongrie

- **Le HDH** mènera un projet destiné à anticiper les parcours de soin dans les maladies cardio-métaboliques grâce à de l'IA.

Données : France, Finlande, Danemark, Hongrie, Norvège

- **L'infrastructure de recherche européenne Elixir** tentera d'identifier des signatures génomiques caractéristiques des différents types de cancers colorectaux.

Données : Belgique, Danemark, Hongrie, Norvège.

Le Direct grant, une opportunité pour consolider les services du Health Data Hub et construire un réseau national avec des partenaires stratégiques)

Le Direct Grant Setting up services by Health Data Access Bodies est un financement européen du programme EU4Health. Il s'inscrit dans la continuité du projet HealthData@EU pilote et de la proposition de règlement EHDS. Son but est d'accélérer la mise en place d'organismes responsables de l'accès aux données de santé dans les États membres et/ou d'en élargir les capacités et missions.

En France, le HDH, qui a été désigné autorité compétente française par le ministère de la Santé et de la Prévention pour répondre à cet appel à candidatures, a déjà mis en place une partie des services et infrastructures demandés. Sa candidature va donc se concentrer sur la consolidation de ces services mais va aussi mettre l'accent sur la construction d'un réseau de partenariats stratégiques qui pourront ainsi participer et gagner en visibilité au niveau européen. La préparation de la candidature française est en cours pour pouvoir être déposée début 2023.

Cinq événements labellisés présidence française du Conseil de l'Union européenne

Au cours du premier semestre 2022, une vingtaine d'événements consacrés à l'e-santé ont été organisés en France et au sein de l'UE, dans le cadre de la présidence française du Conseil de l'Union européenne. Cinq d'entre eux, portés par le HDH, ont mobilisé 24 partenaires et plus de 100 intervenants, à Paris Santé Campus. Conçus pour partager de la connaissance et de l'expertise, mais aussi pour former (en particulier dans le cadre de l'AI4Health Winter School, du 10 au 14 janvier 2022, réunissant plus de 420 participants), ils ont

donné lieu à beaucoup d'échanges. Des experts et des acteurs de l'écosystème ont débattu avec les participants sur les grands défis et les enjeux liés aux données de santé, dont la création de plateformes de données de santé à l'échelle nationale et européenne, l'interopérabilité des données, et l'importance de faire collaborer professionnels de santé et data scientists dans le cadre de Data Challenges.



5

événements labellisés
par la PFUE (Présidence
française du Conseil
de l'Union européenne)



Action 3.4 Promouvoir le Health Data Hub à l'International

Pour promouvoir les initiatives françaises au-delà des sphères françaises et européennes mais aussi dans une démarche d'ouverture et de partage des connaissances, le HDH a renforcé en 2022 ses relations avec les acteurs internationaux et a redoublé d'effort pour proposer à la communauté internationale des événements favorisant de véritables échanges.

Des événements pour rassembler et partager avec la communauté scientifique internationale

Une conférence pour revenir sur les défis et les enjeux liés à la création de plateformes de données de santé en France et en Finlande

Organisée le 11 mai 2022, la conférence « Quelles opportunités pour les plateformes nationales de données de santé ? » est un des événements organisés dans le cadre de la présidence française du Conseil de l'Union européenne. Quelques mois après la signature d'un premier accord de collaboration visant à promouvoir le partage de bonnes pratiques favorisant le partage transfrontalier de données de santé, les acteurs français et finlandais se sont donnés rendez-vous. Objectif : partager les temps forts de la création des deux plateformes et en dresser le bilan deux ans après leur lancement. En plus des représentants de ces deux plateformes, les séquences étaient animées par des représentants des ministères concernés - ministère de la Santé en Finlande et ministère de la Santé et de la Prévention en France - mais également la Caisse nationale d'assurance maladie pour la France, un intervenant de la Commission européenne et un intervenant de Sitra, le fond d'innovation finlandais. À cette occasion, les acteurs à l'initiative de ces plateformes ont pu revenir sur les enjeux

qui ont permis d'ouvrir les données de santé en France et Finlande. 260 participants en ligne et à Paris Santé Campus ont assisté à cette conférence.

Un symposium international dédié à l'Intelligence Artificielle en médecine

Capitalisant sur le succès de sa première édition, le MIT Computer Science and Artificial Intelligence Lab (CSAIL), le MIT Institute for Medical Engineering and Science (IMES), l'Académie nationale de médecine et le HDH se sont associés une nouvelle fois pour organiser la seconde édition du symposium « Intelligence Artificielle et Médecine : promesses et limites ». Lors de cette édition, introduite par Daniela Rus (directrice du Computer Science and Artificial Intelligence Laboratory et doyenne adjointe de la recherche au Schwarzman College of Computing du MIT), les experts européens et américains se sont rassemblés lors de trois conférences. Objectif : éclairer les approches en IA et santé de part et d'autre de l'Atlantique et échanger sur les possibilités et réalités pratiques. Ces trois conférences ont été animées notamment par Leo Anthony Celi, chercheur scientifique principal à l'Institute for Medical Engineering and Sciences (IMES) du MIT ; Farhad Rikhtegar Nezami, chercheur au Brigham and Women's Hospital à la Harvard Medical School et Nicolas Do Huu, cofondateur et Chief AI Officer chez Iktos AI.

Grâce à leurs interventions, les invités ont pu se confronter à des problématiques d'actualité dont l'accès aux données de santé à l'échelle nationale et internationale, les défis de l'IA dans le domaine médical, les questions de validation d'algorithmes ainsi que les potentiels biais de l'IA mais également les besoins des cliniciens ou encore l'interaction homme-machine. Accessible en ligne, l'événement organisé l'après-midi du 20 octobre à l'Académie nationale de médecine de Paris a rassemblé près de 500 participants du monde entier.

AI4health

Coorganisée avec trois instituts interdisciplinaires en intelligence artificielle (Prairie, 3IA Côte d'Azur et MIAI Grenoble), l'édition 2022 de l'école d'hiver dédiée à l'intelligence artificielle en santé, a rassemblé 428 personnes, dont 56% d'étudiants. Pendant cinq jours, des chercheurs de renommée internationale ont présenté leurs avancées et animé des sessions pratiques. Les conférences d'ouverture ont été assurées 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). Pour 2023, le HDH se lance à nouveau dans l'organisation d'une telle école, dont l'apport scientifique et pédagogique a été démontré à plusieurs reprises, avec pour objectif de la tenir en présentiel à Paris.



428
participants à la Winter
School «AI4Health» 2022

Des collaborations binationales à fort impact pour soutenir la recherche de pointe en IA et santé mobilisant des données de santé

En 2022, le HDH a signé son premier accord de collaboration internationale avec un partenaire non européen : le Fonds de recherche du Québec - Santé (FRQS) au Canada. L'objectif est de fédérer et favoriser les collaborations entre les acteurs de l'écosystème de la donnée autour de projets à fort impact sur les systèmes de santé français et québécois. Pour marquer le début de cette collaboration, le FRQS et le HDH se sont associés pour soutenir deux projets mobilisant des données de santé françaises et/ou québécoises dans le secteur de l'intelligence artificielle. Les deux projets lauréats de cet appel ont été dévoilés lors d'un webinaire le 15 décembre 2022. Ils seront financés durant 18 mois, à travers un appel à projets commun.

Zoom sur les projets sélectionnés

Améliorer le traitement et la prise en charge des patients souffrant de sténose valvulaire aortique grâce à des modèles prédictifs d'apprentissage

Ce projet est porté par des équipes de l'Université Côte d'Azur en France et de l'Université Laval au Québec. La sténose aortique est l'une des maladies cardiovasculaires les plus courantes dans le monde, notamment dans les pays développés où l'espérance de vie est en hausse constante. C'est une pathologie dont on ne sait pas contrôler la progression et qui ne peut être traitée à ce jour que par une intervention chirurgicale. Il est donc important de déterminer le meilleur moment pour opérer. C'est précisément le but de ce projet qui souhaite que les patients atteints des formes les plus graves puissent bénéficier d'une prise en charge personnalisée. Pour ce faire, les équipes analyseront des données cliniques et d'imagerie (e.g. échocardiographie, scanner) issues

de la cohorte québécoise PROGRESSA - regroupant des données de patients sur 15 ans - grâce à des algorithmes d'intelligence artificielle.

Prévenir les risques de malformations congénitales en limitant l'exposition des femmes enceintes aux médicaments tératogènes grâce à l'apprentissage automatique

Ce projet est porté par des équipes de l'Université Toulouse 3 - Paul Sabatier en France et du Centre hospitalier universitaire Sainte-Justine au Québec. Identifier les médicaments tératogènes - susceptibles de provoquer des malformations congénitales - est un processus long et complexe, entraînant des risques inutiles pour certaines femmes enceintes et pour leur fœtus. Le projet IN-uTERO a pour objectif d'identifier de nouveaux médicaments tératogènes, en analysant les données de deux grandes cohortes de femmes enceintes/enfants en France (cohorte CQG) et au Québec (cohorte EFEMERIS) grâce à des algorithmes d'intelligence artificielle. Ce projet a également pour objectif de caractériser les profils de femmes enceintes à risque de faire une malformation congénitale, pour soutenir les médecins quand ils leur prescrivent un médicament.

Aller à la rencontre d'écosystèmes des données de santé étrangers : apprendre et partager avec des grands acteurs internationaux

Échanges, apprentissage et partages d'expériences au contact d'autres acteurs de l'écosystème des données de santé, y compris à l'étranger, sont au cœur de la stratégie du HDH. Accompagnés d'un panel d'experts, les équipes du HDH sont donc allées, pour la première fois en 2022, à la rencontre d'acteurs non européens. Ces délégations ont été des occasions précieuses d'en apprendre davantage sur la gestion des données de

santé dans d'autres pays et d'identifier des pistes de collaborations internationales avec des pays confrontés aux mêmes problématiques d'accès et de partage des données de santé.

La première délégation est allée à la rencontre des acteurs clés israéliens du 26 au 29 juin (grâce au soutien du ministère de la Santé israélien et sur invitation de l'ARC Innovation Summit). Elle a réuni des experts du HDH, de l'AP-HP et de la direction générale de l'offre de soins (DGOS) qui ont échangé avec des représentants du ministère de la santé et de la Prévention israélien ainsi qu'avec les principaux acteurs de la recherche en santé et producteurs de données de santé en Israël (les caisses d'assurance maladie et les grands hôpitaux du pays). La délégation a notamment été accueillie par le Sheba Medical Center, Rambam Healthcare Campus, Clalit research institute, Ichilov et Leumit.

Pour la seconde délégation, les experts réunis par le HDH se sont retrouvés à Tokyo du 24 au 28 octobre, grâce au soutien de l'Ambassade de France au Japon. Ils ont pu échanger avec les équipes du ministère de la Santé japonais, les responsables de la gestion des deux plateformes de données de santé nationales (l'une équivalente à la base principale du SNDS, la Japanese National Database of Health Insurance Claims and Specific Health Checkups), et l'autre dédiée à la surveillance des médicaments, permettant le chaînage de données cliniques à des données médico-administratives (MID-NET ou Medical Information Database NETWORK). Étaient également présents lors de ces rencontres, les principaux acteurs de la recherche en intelligence artificielle et santé du pays, le National Cancer Center Research Institute et sa plateforme d'oncogénomique le C-CAT (Center for Cancer Genomics and Advanced Therapeutics), le RIKEN Research institute, Keio University, Sony Computer Science Laboratories Tokyo et Tokyo Medical and Dental University.

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 l'impliquer dans les projets.

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

Le Health Data Hub repose sur un principe fondateur d'écoute, d'ouverture et de dialogue avec la société civile qu'il impulse dans chacun de ses projets via, notamment une direction dédiée. Pour cela, toute une série de dispositifs sont mis en place : des études sur les perceptions, des ateliers participatifs, des consultations et des concertations. Ainsi, en mars 2022, le HDH, en partenariat avec France Assos Santé, et le soutien de l'Association nationale des étudiants en pharmacie de France (ANEPF), a lancé le premier baromètre sur la connaissance des données de santé et de leur usage, auprès d'un panel de 1 000 personnes, représentatives de la population française. De plus, tout au long de l'année, des ateliers de vulgarisation scientifique se sont tenus pour expliquer, dans des termes simples, les projets pilotes que le HDH accompagne.

En mai 2022, le HDH a lancé, avec France Assos Santé, une concertation sur la place des usagers dans les comités ayant trait aux données de santé. Près de 40 personnes ont été entendues, dont des représentants des usagers, et d'autres participants siégeant dans des comités ou désireux de créer prochainement un comité. Parmi les structures représentées se trouvaient le groupement de coopération sanitaire (GCS) G4, les CHU de Rouen,

Lille, Bordeaux et Toulouse, l'AP-HP, l'Institut national du cancer (INCa), l'Inserm, ainsi que les plateformes interrégionales des données de santé HUGO et PAISaGE, la cohorte nationale Hépathér, la Banque Nationale de Données Maladies Rares (BNDMR), ainsi que le CESREES et le comité stratégique du ministère de la Santé et de la Prévention.

Une e-consultation et des workshops pour impliquer les citoyens dans le futur espace européen des données de santé

Lancée le 1^{er} février 2021, l'action conjointe « Vers un espace européen des données de santé » (TEHDAS) a pour but de développer le futur cadre politique, juridique et technique pour le partage et l'utilisation secondaire des données de santé dans le futur espace européen des données de santé (EHDS). L'un des groupes de travail, piloté par le HDH, aux côtés de Sciensano, de la NHS Confederation et de France Assos Santé, avait pour objectif d'appréhender la perception des citoyens sur l'utilisation secondaire de leurs données de santé, ainsi que la manière dont ils souhaitent être impliqués dans l'EHDS. Pour produire les recommandations attendues en janvier 2023, plusieurs phases d'écoute et de dialogue ont été mises en place, en particulier une e-consultation ayant permis de recueillir, en avril 2022, 5 932 contributions, ainsi que des workshops nationaux, dont un français, et un workshop européen.

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

La France a lancé, en 2021, le troisième plan d'action national du Partenariat pour un gouvernement ouvert. Un plan avec des engagements concrets pour l'ouverture et la transparence de l'action publique, auxquels le HDH participe. Le HDH a contribué à cette vaste initiative en proposant des actions concrètes : mise à disposition d'un programme d'écoute citoyenne, construction d'une culture de la donnée de santé, promotion de l'ouverture des données aux citoyens grâce aux Data Challenges et développement de la démarche d'ouvertures de codes sources. La mise en place de ces engagements et de ces divers chantiers permettra, entre autres, de rendre accessibles à tous des connaissances sur les bases de données qui sont uniquement exploitables par des experts actuellement. Ces engagements ont été co-construits avec la société civile et en particulier avec France Assos Santé. D'autres acteurs soutiennent ces actions comme l'ANEPF et la Fédération Française des Diabétiques.

Engagé dans une démarche citoyenne pour associer les citoyens à ses projets et démarches, le HDH a souhaité aller plus loin en 2022. Il a ainsi lancé une étude pour mieux saisir comment d'autres institutions publiques s'y prenaient pour impliquer les citoyens dans leur gouvernance, grâce notamment à des comités de dialogue pérennes. Portée par le HDH, cette étude a été pilotée par un comité rassemblant : le directeur de projets Amdac de la direction de la recherche,

des études, de l'évaluation et des statistiques du ministère de la Santé et de la Prévention, le président de France Assos Santé et le secrétariat général de la Conférence nationale de santé. Les dispositifs de dialogue pérennes de sept organismes publics ont été étudiés : Santé Publique France, la Haute Autorité de santé (HAS), l'Agence nationale de sécurité du médicament, l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (ANSES), l'Institut national de l'environnement industriel et des risques (INERIS), l'Institut français de recherche pour l'exploitation de la mer (IFREMER), l'Institut de radioprotection et de sûreté nucléaire (IRSN).

Enfin, le HDH met en œuvre un système d'information national. Cet outil, qui est en cours de développement, associe les utilisateurs finaux. Il permettra aux citoyens qui le souhaitent d'exercer leurs droits sur leurs données de santé. Il sera aussi un espace d'information et de sensibilisation sur la seconde vie des données de santé, et sur les projets utilisant des données afin que les droits soient exercés de façon éclairée.

Action 4.3 Mettre en place une politique de communication et de diffusion de la culture de la donnée de santé

Pour informer et diffuser une culture de la donnée de santé, plusieurs projets ont été lancés en 2022 :

- Une infolettre sur les données de santé ouverte à tous, en partenariat avec France Assos Santé.
- Une campagne de sensibilisation à la seconde vie des données en Normandie, avec France Assos Santé, le CHU de Rouen Normandie, le Data-lab Normandie, le LINC (laboratoire d'innovation numérique de la CNIL) et les étudiants de l'école Designlr.
- Un nouveau diplôme inter-universitaire (DIU) « Réutilisation des données pour la recherche en santé », en partenariat avec les Universités de Bordeaux, Aix-Marseille et Paris-Cité.
- Un projet d'évolution du répertoire public des projets permettant de restituer les résultats des études effectuées à partir des données du SNDS par les porteurs de projets.

Le HDH conduit par ailleurs des actions de formations. L'une d'elles s'adresse aux citoyens pour les sensibiliser au SNDS, avec le soutien de France Assos Santé, de l'Assurance Maladie, de la CNIL et la DREES.

Enfin, le HDH propose aussi des accompagnements ciblés :

- Avec France Assos Santé pour la mise en œuvre de son étude Vivre-Covid19. Cet accompagnement a permis d'organiser un séminaire soutenu par la présidence française du Conseil de l'Union européenne.

- Avec l'afa Crohn RCH France pour qu'elle dispose d'un tableau de bord rassemblant quelques indicateurs caractérisant les patients atteints de maladies inflammatoires chroniques de l'intestin (MICI) et leur prise en charge.
- Avec l'Association Nationale des Etudiants en pharmacie.



Vivre-Covid 19 : une étude associative ouvre ses données pour la recherche d'intérêt public

Vivre-Covid19 est la première grande enquête de France Assos Santé avec ses membres, dont France Handicap, Unaf, FFAAIR, afa Crohn RCH France, EFAPPE, Fibromyalgie France, France Rein, Unafam, France Alzheimer, Aides, FFD, FNAR, et la Ligue contre le cancer. Avec l'aide d'un cabinet d'étude expert, Sanoia e-Health, plus de 5 400 inscrits s'y expriment chaque mois sur leur niveau d'anxiété et leurs difficultés ou non d'accès aux soins durant la crise sanitaire.

En mai 2020, France Assos Santé et le HDH ont signé un partenariat pour élaborer la doctrine éthique de l'étude et mettre en œuvre les principes qui la structurent. Dès le lancement, pour répondre au premier principe de transparence, le protocole de cette étude a été mis en accès libre.

En 2022, pour répondre au principe d'ouverture à la recherche, un nouveau partenaire a été intégré à la convention : Santé publique France, pour exploiter les 2,5 millions de données de l'étude.

Avec l'étude Vivre-Covid19, c'est la première fois qu'un collectif associatif ouvre ses données à un opérateur public pour exploiter ses données, collectées de façon organisée et selon des procédures garantissant une réutilisation possible. France Assos Santé participe ainsi à la dynamique des sciences ouvertes, du partage et de l'utilisation collective des données.



5 000
participants à l'étude
Vivre-Covid19

Action 4.4 Développer la mise à disposition de données strictement anonymes en libre accès

En complément de son offre de formation, et de toute sa documentation pour aider les porteurs de projets et les responsables de données dans leur démarche d'accès à des données ou de partage de bases, le HDH publie des guides pratiques.

Le guide sur l'anonymisation des données de santé, publié en 2022 en est un exemple. Il explique ce que recouvre le concept d'anonymat et permet d'éviter une confusion fréquente entre la donnée anonyme et la donnée pseudonymisée. Il permet également d'évaluer si les données sont bien réellement anonymes, ce qui est un pré-requis sous la responsabilité du porteur de projet. En effet, il appartient au porteur de projet de s'assurer du caractère anonyme des données de santé qu'il souhaite traiter et de documenter l'analyse menée dans ce cadre.

Un second guide a été mis en accès libre en 2022, sous la forme d'un document à renseigner qui peut ensuite être soumis à l'avis du délégué à la protection des données.

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 accéder aux données ouvertes dans le cadre des Data Challenges accompagnés.

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 règlementé

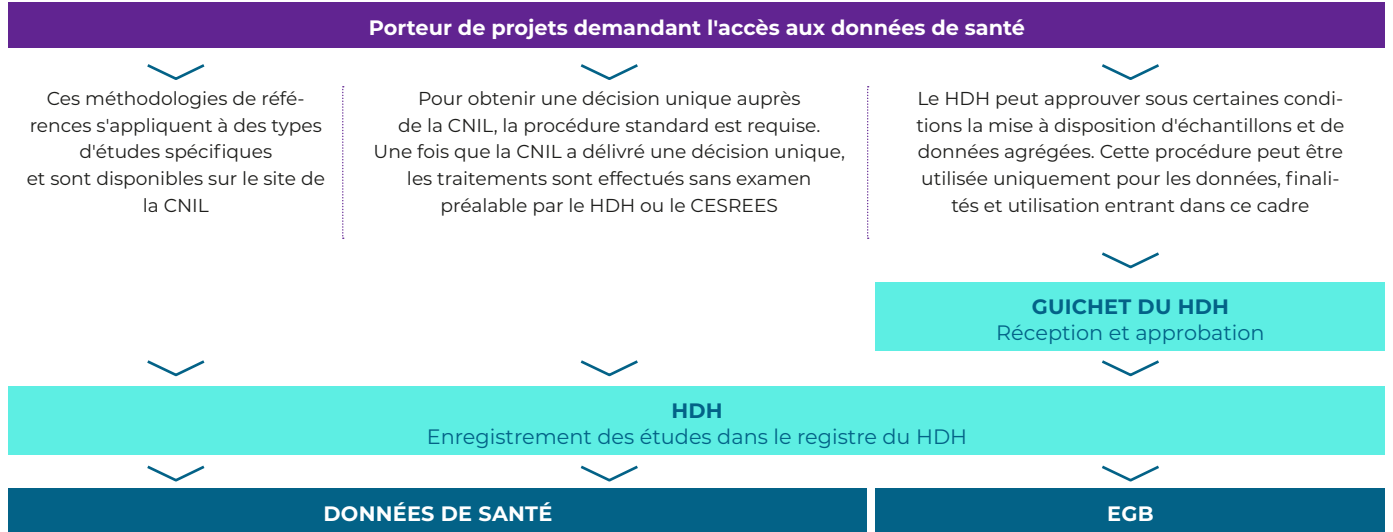
Dans la plupart des cas, il est nécessaire d'obtenir une autorisation auprès de la CNIL pour traiter des données personnelles de santé. En amont de cette autorisation, la procédure prévoit l'intervention de deux institutions :

- Le HDH, point d'entrée unique des projets mobilisant des données de santé : 266 projets y ont été déposés en 2022.
- Le CESREES qui rend un avis sur la finalité, la pertinence et la qualité scientifique du projet.

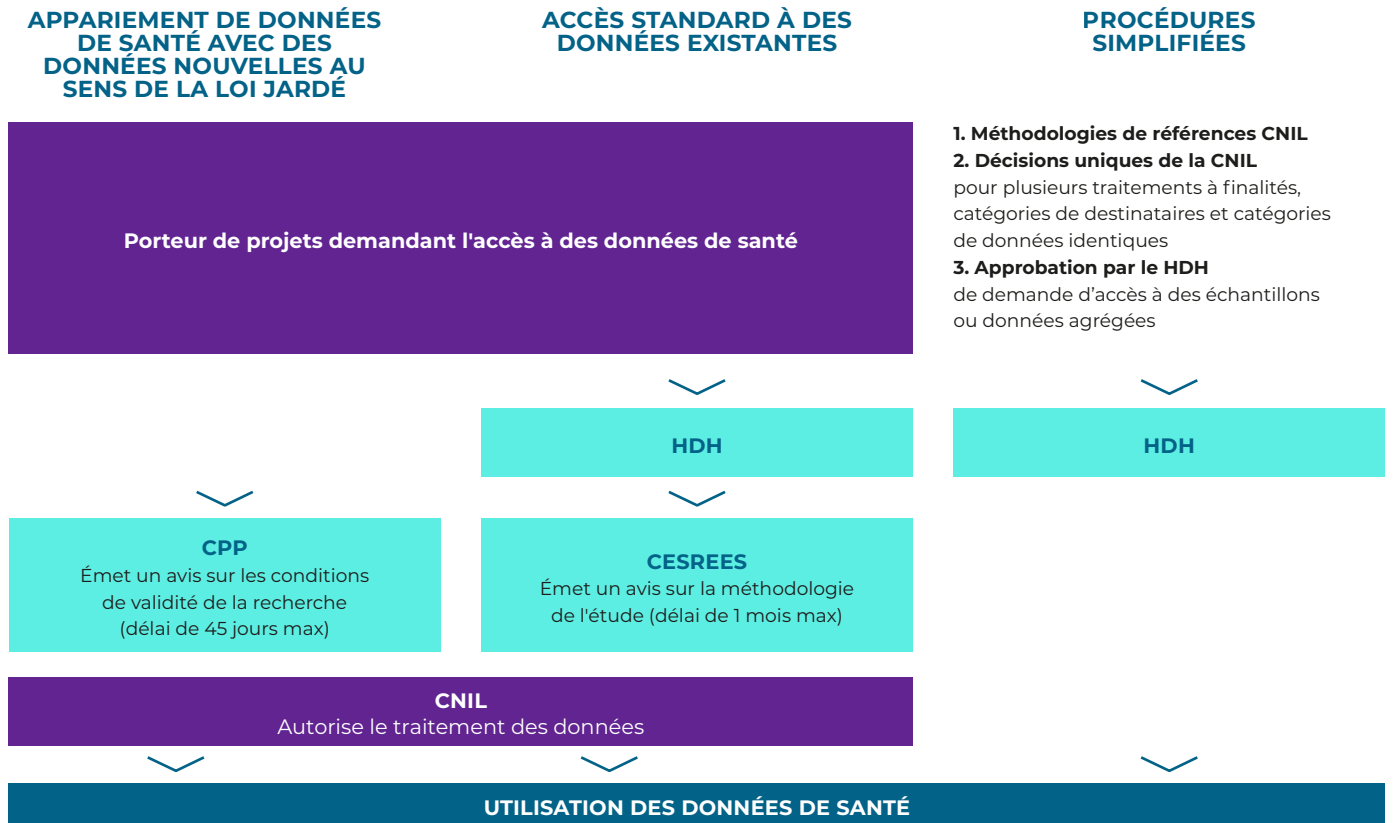
Afin de faciliter l'accès aux données et pour que les porteurs de projet puissent réaliser leur traitement dans les meilleurs délais, le temps imparti à chacune de ces institutions est encadré. Le HDH dispose ainsi de sept jours pour transmettre le projet complet au CESREES. Ce dernier a ensuite un mois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

3 PROCÉDURES SIMPLIFIÉES

- 1 MÉTHODOLOGIES DE RÉFÉRENCE** TELLES QUE MENTIONNÉES À L'ARTICLE 73 DE LA LIL
- 2 DÉCISIONS UNIQUES** TELLES QUE MENTIONNÉES DANS L'ARTICLE 66 DE LA LIL
- 3 ÉCHANTILLONS ET DONNÉES AGRÉGÉES** TELS QUE MENTIONNÉS AU II DE L'ARTICLE 66 DE LA LIL



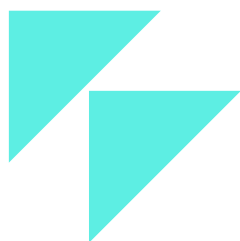
SCHÉMATISATION DU CIRCUIT DES DEMANDES D'ACCÈS AUX DONNÉES DE SANTÉ



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 2022, il s'est ainsi prononcé sur deux projets de mise à jour de méthodologies de référence encadrant l'accès simplifié aux données du PMSI (MR-005 et MR-006), les évolutions proposées visant à élargir leur périmètre et revoir les modalités d'accès.

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. Il a dans ce cadre, en lien avec la CNAM, saisi la Commission en mai 2022 pour adapter le référentiel d'accès simplifié pour l'EGB au nouvel échantillon du SNDS. Grâce à cette procédure ajustée, les porteurs de projet peuvent désormais accéder à un échantillon de la base principale du SNDS, de taille plus importante, sans solliciter une autorisation de la Commission, dès lors que leur projet est conforme au référentiel.

Nombre total de projets déposés au guichet en 2022	266
Nombre de projets déposés au guichet en vue d'une autorisation CNIL	259
dont transmis au CESREES	195
dont réorientés MR ou CPP	48
dont non complets	16
Nombre de demandes d'accès à l'EGB simplifié	7
Nombre de projets relevant d'une MR (méthodologie de référence)	1216
MR004	1096
MR005	42
MR006	78



Une offre de service destinée aux porteurs de projets

Favoriser l'utilisation des données de santé nécessite de documenter ce qu'elles contiennent et les possibilités qu'elles offrent, rendre lisibles les procédures d'accès et favoriser le partage des connaissances. C'est pourquoi le HDH élabore en permanence des outils opérationnels et pédagogiques destinés aux porteurs de projets. Ces outils les accompagnent à toutes les étapes de leur projet et sont améliorés chaque année. En 2022, deux nouveaux guides traitant de l'anonymisation sont ainsi venus enrichir la documentation partagée par le HDH :

- Un premier guide explicite les notions d'anonymisation des données et d'identification des personnes pour aider les porteurs de projets à cerner ce que recouvre le processus d'anonymisation, son périmètre d'application et les possibilités qu'il ouvre, notamment par rapport la pseudonymisation.
- Le second guide propose une méthode pour évaluer le risque de réidentification d'un jeu de données et attester que le risque est bien nul. L'évaluation du caractère anonyme des données étant de la responsabilité des porteurs de projets, il est important qu'ils disposent des clefs leur permettant de démontrer l'impossibilité de réidentifier les personnes à partir des données qu'ils souhaitent mobiliser.

Usage des données de santé et transparence

Le répertoire public des projets, prévu par la loi, s'est enrichi en 2022 de 1200 projets supplémentaires. Au total, plus de 7 000 projets mobilisant des données personnelles de santé y étaient répertoriés à fin 2022. Pour chacun d'eux, une fiche présente le bénéfice attendu, les données mobilisées, la méthode, les coordonnées du porteur de projet et de son DPO, ainsi qu'un calendrier de mise en œuvre. Un recueil des résultats issus de ces projets a en outre été initié, conformément au code de la santé publique. Pour que les acteurs concernés puissent mieux comprendre cette obligation, le HDH a conçu un guide pédagogique qui sera publié prochainement sur son site internet. Les résultats des projets pourront par ailleurs être téléchargés depuis le répertoire public dans les prochains mois.

En 2022, la réflexion pour faire évoluer la plateforme de soumission des demandes d'accès aux données s'est poursuivie. Elle s'est traduite par la mise en place au cours de l'été 2022 d'une plateforme web collaborative. Objectif : simplifier la gestion des demandes d'autorisation CNIL soumises au HDH au titre de sa mission de guichet, les échanges entre le secrétariat du CESREES (assuré par le HDH) et les rapporteurs des dossiers désignés par le président du comité, la gestion des conflits d'intérêts, ainsi que l'organisation de ces comités. Cette réflexion se poursuivra en 2023, pour mieux répondre aux besoins des acteurs de cette procédure et prendre en compte les évolutions à venir sur les formulaires de soumission de la CNIL.



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

9 rue Georges Pitard, 75015 Paris

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

Édition : Health Data Hub

Conception et réalisation : Parimage

Mars 2023

