

# HEALTH DATA HUB



# Introduction.

Les techniques de traitement de masse des données et d'intelligence artificielle laissent aujourd'hui entrevoir des progrès sans précédent dans l'amélioration de la qualité des soins, l'accompagnement des patients et la gestion du système de santé.

Au niveau international, la course est lancée : de nombreux pays investissent massivement dans le domaine, et des grandes entreprises américaines et chinoises communiquent massivement sur de premiers succès.

En France, le Président de la République a lancé en mars 2018 un plan « Intelligence Artificielle » avec pour objectif de faire de la France un leader dans ce domaine.

La France a déjà de nombreux atouts : une recherche en médecine et en mathématiques appliquées parmi les meilleures du monde, une base de données médico-administratives exceptionnelle, de nombreuses cohortes, registres et données hospitalières, ainsi qu'un écosystème de start-ups très actif dans ces domaines.

Il manquait néanmoins une infrastructure clef, un système de base de données et de services liés : c'est l'enjeu du Health Data Hub (HDH). Il permettra ainsi de croiser ces bases de données de santé et de faciliter la réalisation des projets d'intérêt public par les nombreuses équipes de recherche académiques et industrielles dans un respect total de la vie privée des usagers du système de santé.

Pierre angulaire de la démarche, un catalogue construit de manière progressive et itérative permettra de mettre à disposition des chercheurs, mais aussi des associations de patients et citoyens, des institutions, des start-ups, et des différentes parties prenantes du secteur de la santé les bases de données les plus prometteuses. La plateforme technologique mobilisée à cet effet garantira un très haut degré de sécurité et ne pourra héberger que des données non nominatives.

Le 24 juillet dernier, la loi relative à l'organisation et la transformation du système de santé, dont l'article 41 crée la structure a été promulguée. Quatre mois après, elle est officiellement créée et regroupe 56 parties prenantes, en grande majorité issues de la puissance publique.

Les premiers axes de la feuille de route stratégique 2020-2022 sont d'ores et déjà fixés: il s'agira de décloisonner et d'améliorer la qualité de notre patrimoine de données, de renforcer l'usage par une offre de service attractive, de positionner la France comme un leader dans l'usage des données de santé, et de garantir la participation de la société civile au projet et l'exercice de l'ensemble de leurs droits. Le Health Data Hub s'appuiera pour cela sur des partenariats, prioritairement en France et en Europe, avec des instituts de recherche et des producteurs de données.



# Quelques cas d'usage.

## **Apporter des réponses aux patients atteints de pathologies rares.**

Le sarcome est une catégorie de tumeurs rares pour laquelle l'efficacité des essais cliniques arrive à ses limites. Depuis près de 40 ans, ils ont été conduits sans pouvoir trancher sur la pertinence de la chimiothérapie avant ou après la chirurgie. Dans ce contexte, l'analyse des données collectées auprès de patients dans le cadre de cohortes, associées aux données de consommations de soin issues de l'Assurance Maladie offre une réelle opportunité d'évaluation des traitements et de leur performance.

## **Appuyer le professionnel de santé dans un contexte clinique qui se complexifie.**

Les logiciels de prescription utilisent déjà des outils d'alertes pour notifier le professionnel lorsque différents traitements pris par un même patient peuvent révéler des incompatibilités avec des risques pour sa santé. Largement perfectibles, ces systèmes aujourd'hui génèrent tant d'alertes qu'ils sont souvent désactivés ou ignorés par les professionnels. L'analyse des données permet de prioriser les alertes et de filtrer celles qui doivent impérativement être prises en compte.

## **Prendre en charge les patients le plus tôt possible.**

Les crises d'insuffisance cardiaque causent près de 5% des hospitalisations en France. En corrélant les données issues des pacemakers avec les données relatives aux hospitalisations, il est possible de concevoir des outils d'alerte qui détectent en amont des signaux faibles et permettent de prendre en charge les patients le plus tôt possible.



### **Prédire les trajectoires individuelles des patients et améliorer les actions de prévention.**

Les données de cohortes de patients suivis dans les centres experts Parkinson rapprochées de celles de l'Assurance Maladie permettent d'étudier la progression de la pathologie et d'ainsi améliorer le suivi des patients. Des outils pourraient ainsi être développés et fournis aux neurologues.

### **Améliorer la compréhension et la transparence du système de santé.**

En 2017, près de 82000 effets indésirables médicamenteux ont été déclarés pour 12000 médicaments commercialisés. L'analyse des données de santé permet d'aller plus loin dans la compréhension en calculant les taux de déclaration, c'est à dire le nombre d'effets indésirables déclarés rapporté au nombre de patients réellement exposés. Ce chiffre est stratégique pour les autorités, les professionnels de santé et les patients afin d'apprécier l'importance de chaque effet secondaire. Actuellement réalisé à la main, ce calcul pourrait être automatisé et facilité grâce à l'accès aux grandes bases de données par l'ANSM.

### **Gagner du temps médical, améliorer le dépistage et réduire les délais de diagnostic.**

Avec 11 883 décès estimés en 2017, le cancer du sein tue plus de femmes que tout autre cancer en France. Le dépistage précoce permet une baisse de 21% de la mortalité. Le recours à l'intelligence artificielle pourrait encore améliorer le dispositif de dépistage organisé du cancer du sein en réduisant le taux de faux positifs et de faux négatifs.

### **Proposer aux patients les meilleurs traitements sur le long cours.**

Les informations concernant les liens entre la prise de médicaments et leurs effets à long terme sont aujourd'hui très lacunaires, dans le cas général comme en transplantation d'organe. Dans le cadre d'un traitement à vie, la connaissance de ces relations est indispensable pour optimiser les stratégies thérapeutiques, le choix des doses mais aussi des formules de ces médicaments.





## **Notre vision.**

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


# Notre offre de services.

L'offre de services du HDH s'articule autour de 4 axes s'adressant aux différentes parties prenantes de l'écosystème :

- **Soutenir les responsables de données dans la collecte, la consolidation et la mise en valeur de leur patrimoine**
  - Appui à la collecte, la standardisation et la documentation des données
  - Fourniture d'un hébergement à l'état de l'art sécurisé et accompagnement dans la mise en conformité RGPD
  - Valorisation et rétribution du travail et de l'expertise du responsable des données
- **Offrir à tous les porteurs de projets un accès simplifié et accéléré aux données**
  - Consolidation d'un catalogue de données documentées
  - Accès à un guichet unique et à des procédures d'accès simplifiées
  - Mise à disposition d'une plateforme technologique et d'outils à l'état de l'art, au plus haut niveau de sécurité
- **Garantir à la société civile le respect des droits des citoyens**
  - Co-construction et consultation citoyenne
  - Ethique, transparence et traçabilité des usages
  - Information et accompagnement des citoyens dans l'exercice de leurs droits
- **Innover aux côtés des acteurs de la recherche et de l'industrie**
  - Organisation d'appels à projets et d'événements fédérateurs
  - Développement de partenariats académiques, industriels, nationaux et internationaux
  - Appui à la dynamique de développement d'outils open source et de l'open data

• INVENTORS  
• CREATIVITY  
• DISCOVERY





# Soutenir les responsables de données dans la collecte, la consolidation et la mise en valeur de leur patrimoine.

## **Appui à la collecte, la standardisation et la documentation des données.**

- Le Health Data Hub appuie financièrement et humainement les responsables des données dans la conduite des différentes actions nécessaires au transfert et à la mise à disposition des données : extraction, dé-identification, standardisation, documentation, création d'échantillon, mise en place des flux d'ingestion.
- Aux côtés des responsables des données, il contribue à la diffusion de standards internationaux et à une meilleure interopérabilité, afin de permettre des agrégations et chaînages de qualité.
- Il promeut enfin les initiatives visant à coordonner les acteurs locaux d'un territoire dans le but de créer de nouvelles bases.

## **Fourniture d'un hébergement à l'état de l'art sécurisé et accompagnement dans la mise en conformité RGPD.**

- Le Health Data Hub est secrétaire du CESREES, comité éthique et scientifique des recherches, études et évaluations en santé, qui évalue les demandes d'accès au catalogue de données.
- A ce titre, il assure le lien avec les comités scientifiques des bases concernées afin de leur permettre de formuler toutes leurs remarques sur un dossier précis. Cela doit garantir que le CESREES ne donne pas un avis contrevenant aux finalités convenues avec les volontaires lors du recueil de leurs données dans le cas des cohortes par exemple. De même, il s'agira de ne pas donner un avis favorable à un projet qui semble pertinent mais qui ne serait pas en réalité faisable du fait du protocole de collecte initial des données, de leurs biais potentiels ou de leur qualité plus généralement.



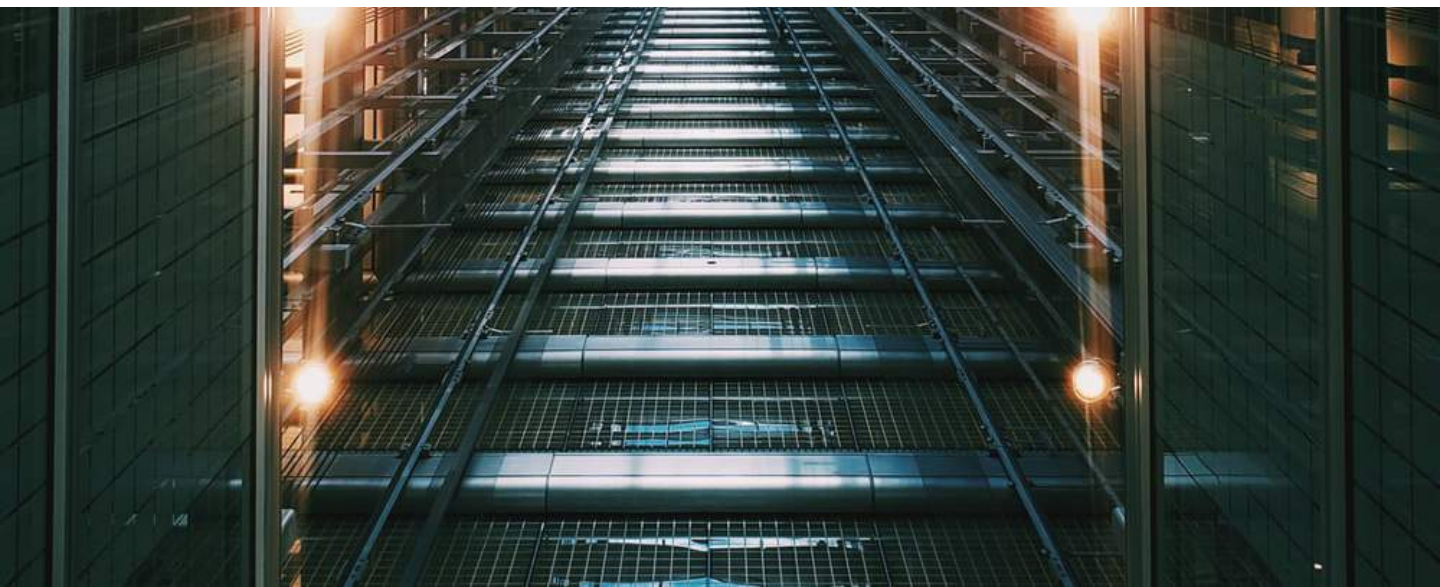
- Le Health Data Hub fournit une plateforme technologique sécurisée garantissant un hébergement certifié “hébergement de données de santé” et conforme aux textes applicables, avec une gouvernance de cybersécurité impliquant le haut fonctionnaire de défense et de sécurité des ministères sociaux et l’ANSSI (Agence Nationale de Sécurité des Systèmes d’Information).
- Il sera notamment capable de tracer toutes les actions réalisées sur les données afin d’en permettre l’audit et s’articulera avec les responsables de données pour rendre effectif l’exercice des droits des citoyens sans compromettre la sécurité des données.

### **Valorisation et rétribution du travail et de l’expertise du responsable des données.**


- Le Health Data Hub assure l’enrichissement systématique des

bases du catalogue avec les données de l’assurance maladie (SNDS historique), accélérant ainsi des appariements qui peuvent prendre aujourd’hui plusieurs années.

- Il renforce la visibilité des bases de données de recherche, et augmente ainsi les citations mentionnant leur ré-utilisation. Il garantit aux équipes en charge de la collecte un avantage compétitif et leur propose une information systématique à chaque publication de résultats obtenus grâce à leurs données.
- Il favorise voire accompagne la mise en place de partenariats entre les responsables des données et les utilisateurs, et leur offre la possibilité de tarifier l’accès direct aux données, de l’expertises ou des analyses réalisées par le responsable lui-même.







# Offrir à tous les porteurs de projets un accès simplifié et accéléré aux données.

## Consolidation d'un catalogue de données documentées.

- La loi Organisation et Transformation du Système de Santé élargit le SNDS, initialement composé des données médico administratives telles que les feuilles de soin, la facturation hospitalière et les causes médicales de décès, à l'ensemble des données associées à un remboursement de l'Assurance Maladie. Ce périmètre est très large puisqu'il couvre les données de registres, de cohortes, les données hospitalières, de médecine de ville, etc. Pour autant, l'ensemble des données concernées ne va pas être rassemblé au sein d'un fichier – ou même d'une plateforme – unique. Une sélection des bases les plus pertinentes sera répliquée et mise à jour en contenu au niveau du Health Data Hub afin de permettre leur
- partage à une communauté plus grande de porteurs de projets.
- L'agrégation des sources et/ou leur chaînage est essentiel pour obtenir la masse critique ou la perspective nécessaire au développement de certains cas d'usage. Les responsables des données seront libres de mettre à disposition ces données via d'autres systèmes, pourvu qu'ils soient tout aussi sécurisés.
- Afin d'aider les porteurs à explorer le patrimoine de données disponible et d'évaluer facilement la faisabilité d'un projet, le Health Data Hub développera en outre des outils de catalogage, de métadonnées, de documentation et de création de données synthétiques en collaboration avec les responsables des données.

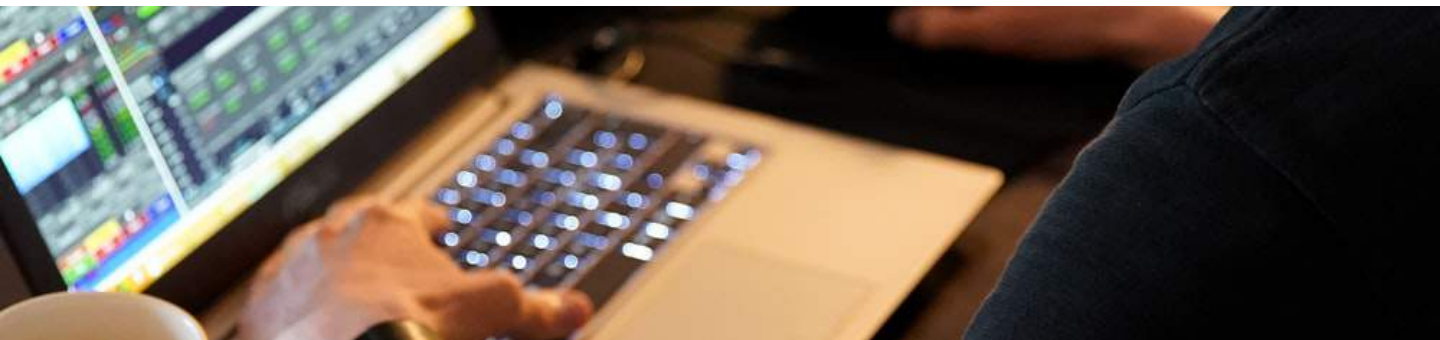
## **Accès à un guichet unique et à des procédures d'accès simplifiées.**

- Le Health Data Hub appuie les porteurs de projets nécessitant une autorisation de la CNIL à chaque étape de leur demande d'accès. Il assure à ce titre le secrétariat unique auprès de qui les dossiers sont déposés. Le secrétariat les transmet au comité éthique et scientifique national (CESREES) qui apprécie l'intérêt public de la finalité poursuivie, l'adéquation des données demandées avec cette finalité, la méthodologie et le cas échéant, la pertinence scientifique. Interlocuteur privilégié, le secrétariat assure le lien entre les porteurs de projet, les responsables de données, le CESREES et la CNIL afin de permettre la l'accès aux données dans les meilleurs délais. Tout au long de ces étapes, les chargés de mission, juristes et experts des données répondent aux questions des porteurs de projet, les appuient dans leurs démarches, et ce jusqu'à la mise à disposition des données.
- Le Health Data Hub est également force de proposition pour l'élaboration de procédures simplifiées d'accès aux données en étroite collaboration avec la CNIL.

- Le Health Data Hub maintient le répertoire public des projets mobilisant les données du SNDS. Il permet ainsi aux porteurs de produire et diffuser des informations relatives à leurs traitements et à leurs résultats auprès de la société civile et des responsables de données concernés.

## **Mise à disposition d'une plateforme technologique et d'outils à l'état de l'art, au plus haut niveau de sécurité.**

- Entièrement administrée par des agents du Health Data Hub, aux rôles et responsabilités bien définis, la plateforme technologique permettant l'analyse innovante des données mobilise des composants industriels choisis pour leur performance, leur sécurité et leur élasticité.
- L'offre logicielle à l'état de l'art permettra le développement des codes et algorithmes sans contrainte (R, python, spark..)
- En cible, une place de marché permettra également aux éditeurs de partager les outils pertinents avec la communauté et d'ainsi favoriser la mutualisation.





# Garantir à la société civile le respect des droits des citoyens.

## Co-construction et consultation citoyenne.

- Le Health Data Hub est un outil et un service construit pour et avec les citoyens, dans une logique affirmée de co-construction.
- Des consultations seront organisées à échéance régulière pour recueillir le point de vue et les besoins des citoyens sur des problématiques spécifiques.
- Les différents groupes de travail associent par ailleurs des représentants d'associations pour porter au mieux la voix des patients et plus généralement des citoyens.

## Ethique, transparence et traçabilité des usages.

- L'intérêt public de chaque projet et le caractère strictement nécessaire des données seront évalués par un comité éthique et scientifique indépendant, incluant des experts scientifiques, éthiques, juridiques, et

des représentants des associations de patients.

- La CNIL est l'autorité indépendante chargée de la protection des données à caractère personnel en France ; elle intervient dans la procédure d'accès aux données après l'avis du comité afin d'autoriser ou refuser les projets. Au-delà des autorisations qu'elle délivre, la CNIL exerce un pouvoir de contrôle sur l'utilisation des données et peut sanctionner tout mésusage.
- Le Health Data Hub mettra par ailleurs à disposition sur son site internet et de manière facilement exploitable la liste des projets de réutilisation des données de santé, avec l'identité de leur porteur et les thématiques couvertes, ceci conformément à l'obligation de transparence qui incombe à tous les projets menés sur des données du SNDS. Seront également partagés les résultats de ces projets dans le respect de la concurrence et de la compétitivité académique.



## **Information et accompagnement des citoyens dans l'exercice de leurs droits.**

- La loi a confié au Health Data Hub la mission "D'informer les patients, de promouvoir et de faciliter leurs droits, en particulier concernant les droits d'opposition". Cela se traduit par la publication sur son site internet d'informations claires et directement compréhensibles sur les droits des citoyens et la façon de les exercer. Dans la mesure où le Health Data Hub ne stockera jamais

d'information susceptible d'identifier directement une personne, l'exercice des droits des personnes se fera en concertation avec l'ensemble des responsables des données sources, sous la coordination du délégué à la protection des données du Health data Hub.

- Dans la même optique, le Health Data Hub produira une information sur son site internet régulièrement mise à jour en particulier sur les bases de données dont il assure la mise à disposition.







# Innover aux côtés des acteurs de la recherche et de l'industrie.

## Organisation d'appels à projets et d'événements fédérateurs.

- Le Health Data Hub se construit de façon progressive en s'appuyant sur l'écosystème de la santé. En particulier des appels à projets, prévus à un rythme annuel, permettront d'accélérer des projets spécifiques tout en permettant d'éprouver et d'approfondir l'offre de service du Health Data Hub.
- D'autres événements fédérateurs sont organisés tels que le colloque "Données de santé et intelligence collective" dont la première édition a eu lieu fin 2019 avec comme toile de fond les enjeux éthiques pratiques de l'utilisation des données de santé et des usages futurs de celles-ci.

## Développement de partenariats académiques, industriels, nationaux et internationaux.

- Le Health Data Hub rassemble l'écosystème de santé : les

organismes de recherche, les établissements et professionnels de santé, les associations de patients, les institutions, les industriels, les start-up... Conformément à la volonté de co-construction du Health Data Hub, différents partenariats pourront être développés qui devront prévoir la mise à disposition de jeux de données, de documentation, d'outils ou d'algorithmes créés pouvant servir à l'ensemble de la communauté.

- La mise en place d'un réseau de chercheurs affiliés au Health Data Hub ou encore l'organisation de la summer school "AI 4 Health" en collaboration avec les instituts 3IA, au cours de laquelle interviendront des experts de renommée internationale, permet de tisser des partenariats multidisciplinaires forts. Une fondation pourrait à terme être mise en place afin de promouvoir et multiplier ces partenariats.

## Appui à la dynamique de développement d'outils open source et de l'open data.

- Si l'ambition du Health Data Hub est bien de permettre un accès plus rapide aux données de santé, il n'est pas toujours garanti que les données obtenues permettent d'aboutir à des résultats probants, fiables, valorisables. Afin de tirer le meilleur parti des données, un travail considérable est nécessaire pour comprendre leur organisation, la façon dont elles sont collectées, les biais et les limites qu'elles présentent. Cet investissement ne saurait être entrepris par chacun des porteurs de projet. A ce titre, le Health Data Hub animera l'écosystème afin de permettre le partage et l'entretien des connaissances et des outils entre les projets.
- Le Health Data Hub inscrira également son action dans la démarche open data. A titre d'exemple, il co-organisera avec la société française de pathologie un Data Challenge en intelligence artificielle sur des lames virtuelles, dont les données pourront à terme rester ouvertes au-delà, sous réserve d'anonymisation totale, ce qui serait une première en France pour des données médicales.



# Pour plus d'information:

<https://www.health-data-hub.fr/>

[Health Data Hub sur Twitter](#)

[Health Data Hub sur Youtube](#)



**Stéphanie COMBES**

Directrice du Health Data Hub

[stephanie.combes@health-data-hub.fr](mailto:stephanie.combes@health-data-hub.fr)



**Bruno MAQUART**

Président du Health Data Hub

[Bruno.maquart@health-data-hub.fr](mailto:bruno.maquart@health-data-hub.fr)