

HEALTH DATA HUB

FAIRE DE LA FRANCE UN LEADER DE L'UTILISATION DES DONNEES DE SANTE

Javier Nicolau

Lab santé

Direction de la recherche, des études, de l'évaluation et des statistiques - ministère des solidarités et de la santé

Plateforme Nationale pour la recherche sur la fin de vie – 25/10/2019



Le Health Data Hub Pourquoi ?

Dynamiser l'utilisations des données de santé est indispensable pour améliorer la qualité des soins

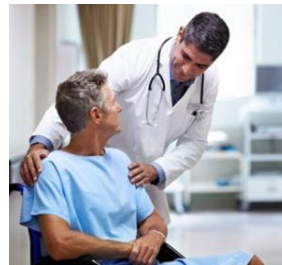
L'usage des données, en particulier des données de santé, connaît une explosion à l'échelle mondiale. Des nouvelles innovations émergent chaque jour. Ne pas saisir l'opportunité d'utiliser les données pour la prise en charge constitue une perte de chance pour le patient.

Recherche médicale



- Mettre en évidence des biomarqueurs et des prédispositions génétiques à certains cancer, identification des états précancéreux

Appui aux professionnels de santé



- Prédire les risques de complications liés à une opération et adapter la prise en charge pour un patient, notamment dans le cas d'une forte obésité

Suivi et information des patients



- Développer pour les patients des outils d'aide à l'interprétation, par exemple des examens biologiques

Pilotage du système de santé



- Prédire le nombre de ré-hospitalisations potentiellement évitables pour améliorer l'offre de soin

La France a les atouts pour être un leader de l'utilisation des données de santé mais des verrous subsistent.



Patrimoine de données

un patrimoine de données riche, varié,
de qualité

une absence de cartographie des sources de données, des règles d'accès hétérogènes, des données cloisonnées sans identifiant unique, des réflexes propriétaires, un contexte juridique complexe



Compétences

une excellence médicale et en
mathématique appliquée

des outils d'analyse restrictifs qui ne permettent pas toujours l'émergence d'usages innovants en comparaison des outils à l'état de l'art, une difficulté d'attirer et conserver des compétences du digital



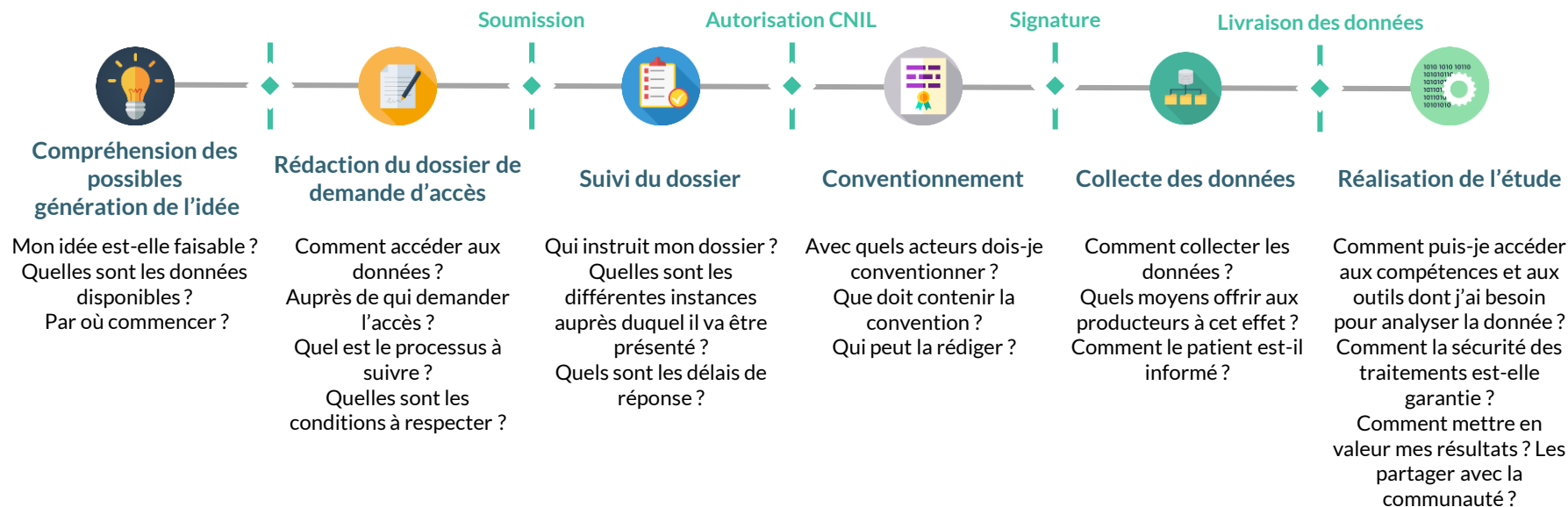
Financements

il existe quelques programmes tels que hop'en, e-parcours, article 51

des contraintes structurelles de financement de la recherche et de la collecte des données et l'absence de modèle économique viable pour les innovations

Dit autrement, mener un projet d'analyse des données de santé est complexe aujourd'hui en France

La consultation des parties prenantes réalisée dans le cadre de la mission de préfiguration menée à l'été 2018 a fait émerger des freins à chacune des étapes de la réalisation de projets d'exploitation de données de santé.





La situation suite à la loi de 2016

Aperçu des différentes sources de données de santé en France

Données médico-administratives



Registres épidémiologiques et de pratiques

Données issues de la recherche et des grandes cohortes nationales

Données des cabinets de médecine de ville

Données d'hospitalisation et dossier patient informatisé hospitalier

Données des laboratoires de biologie médicale

Données des laboratoires d'imagerie de ville

Dossier médical en santé au travail

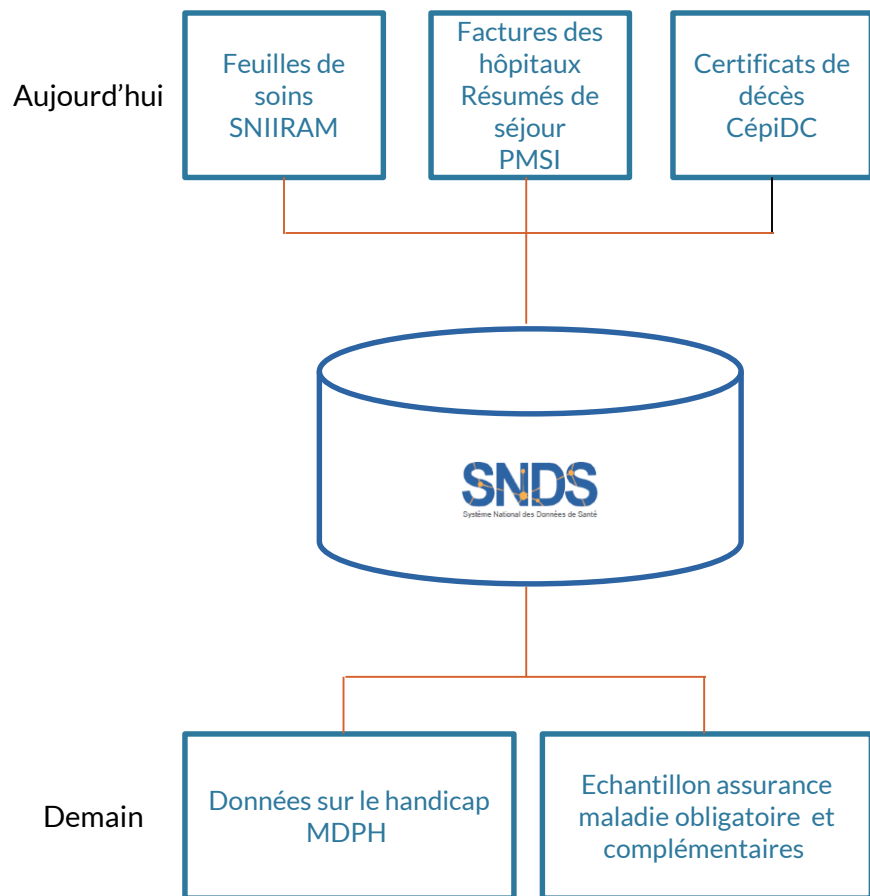
Données issues du secteur médico-social

Données d'évaluation par les patients

Données issues de la télémédecine et des dispositifs médicaux connectés

Données contextuelles

Créé dans la loi de 2016, le SNDS rassembleait plusieurs sources de données ne constituant qu'une petite partie du patrimoine existant



Une base de 67 millions de personnes avec une vision complète du parcours de soins du patient, sans de biais de sélection, sans perdus de vue, avec une bonne homogénéité du codage et 12 ans de profondeur d'historique



L'absence de données cliniques ou paracliniques, d'antécédents ou de facteurs de risques


Ce que la loi de 2016 a mis en place :


- une doctrine d'accès explicite : usage du SNDS permis à tous sous conditions de finalités et de sécurité
- une gouvernance unifiée des différentes bases de données le constituant
- des conditions et procédures d'accès claires, transparentes, des critères identiques pour tous, des délais de réponse garantis jusqu'à l'autorisation CNIL
- la perspective de procédures simplifiées
- un usage facilité du NIR
- un organisme facilitateur chargé de fluidifier les accès : l'institut national des données de santé (INDS)





Quelques exemples d'utilisation du SNDS

A l'heure actuelle, les demandes reçues par l'INDS pour en majorité sur :

 **Etude sur des pathologies spécifiques** : dépression résistante, maladie de Fabry, cataractes chez les patients atteints de glaucomes...

 **Epidémiologie** : quantification de populations atteintes, fardeau des maladies...

 **Etudes sur la prise en charge et les impacts des traitements** : antalgiques paliers II et III, couverture vaccinale, enfants et adolescents asthmatiques...

 **Recherche de facteurs prédictifs / facteurs de risque** : facteurs prédictifs de la non adhésion au traitement, facteurs déterminants du parcours des patients transplantés rénal, facteurs de risques de récurrence des tumeurs méningées et de survie des patients ...

 **Coût de la prise en charge et modèles médico-économiques** : évaluation du coût global du cancer en France, impact économique de la pénétration des biosimilaires, ...

De façon croissante dans un avenir proche :

 **Etudes "IA", fouille de données, datascience**

 **Etudes sur données appariées**



Le Health Data Hub Offre de service

- ✓ Suite au Rapport de Cédric Villani, volonté présidentielle de faire de la France un leader de l'IA - Discours du Président de la République le 29 Mars 2018 au collège de France, annonce de la mise en place d'un Hub de données de santé
- ✓ Mission de préfiguration lancée en juin 2018 avec 3 co-pilotes : Marc Cuggia, Dominique Polton, Gilles Wainrib
- ✓ Groupe plénier, 115 auditions, 5 ateliers, un benchmark international, une expertise juridique
- ✓ Rapport remis à la Ministre en Octobre 2018
- ✓ Des convictions fortes issues du dialogue avec l'ensemble de l'écosystème mobilisé, une ambition et une feuille de route

- ✓ **Mieux connaître** et pouvoir plus **facilement accéder** au patrimoine de données de santé
- ✓ **Décloisonner** les données et **faciliter leur rapprochement, dans le respect du droit du citoyen**
- ✓ Avoir accès à des **moyens adaptés et suffisants** (technologiques, humains et financiers) pour consolider et valoriser le patrimoine de données
- ✓ Fédérer l'écosystème autour d'un **modèle économique d'ensemble** favorisant le partage et reconnaissant les efforts de chacun

La loi OTSS crée le Health Data Hub, issu de la transformation de l'INDS dont il élargit les missions

Le Health Data Hub sera créé d'ici le 31 octobre. Il devrait rassembler une cinquantaine de personnes. Il reprend et élargit les missions de l'INDS. Celles-ci sont inscrites dans l'article 41 de la loi organisation et transformation du système de santé.

Secrétariat de l'accès aux données de santé



- Le HDH conserve la mission de secrétariat unique des demandes d'accès, d'expertise du SNDS et des facilitation des procédures d'accès
- Le HDH continue de mettre à disposition les informations relatives aux études sur son site internet (actuellement sur celui de l'INDS)



- Le HDH a également la mission d'aider les citoyens à exercer leur droit d'opposition s'ils le souhaitent

Animation et promotion de l'innovation



- Le HDH promeut l'innovation, il met à disposition de la documentation et des outils d'aide à la manipulation de données (meetup, faq, etc.)
- Le HDH organise des appels à projet et des challenges dans le but de concevoir de nouveaux services et de mettre en avant des thématiques particulières.
- Il met en relation les acteurs de l'écosystème



Plateforme sécurisée



- Le HDH met à disposition les jeux de données demandés par les utilisateurs dans une plateforme sécurisée avec des outils d'analyse à l'état de l'art pourvu que l'utilisateur dispose d'une autorisation CNIL.
- Si nécessaire, il réalise les appariements et apporte son soutien au producteur dans la collecte

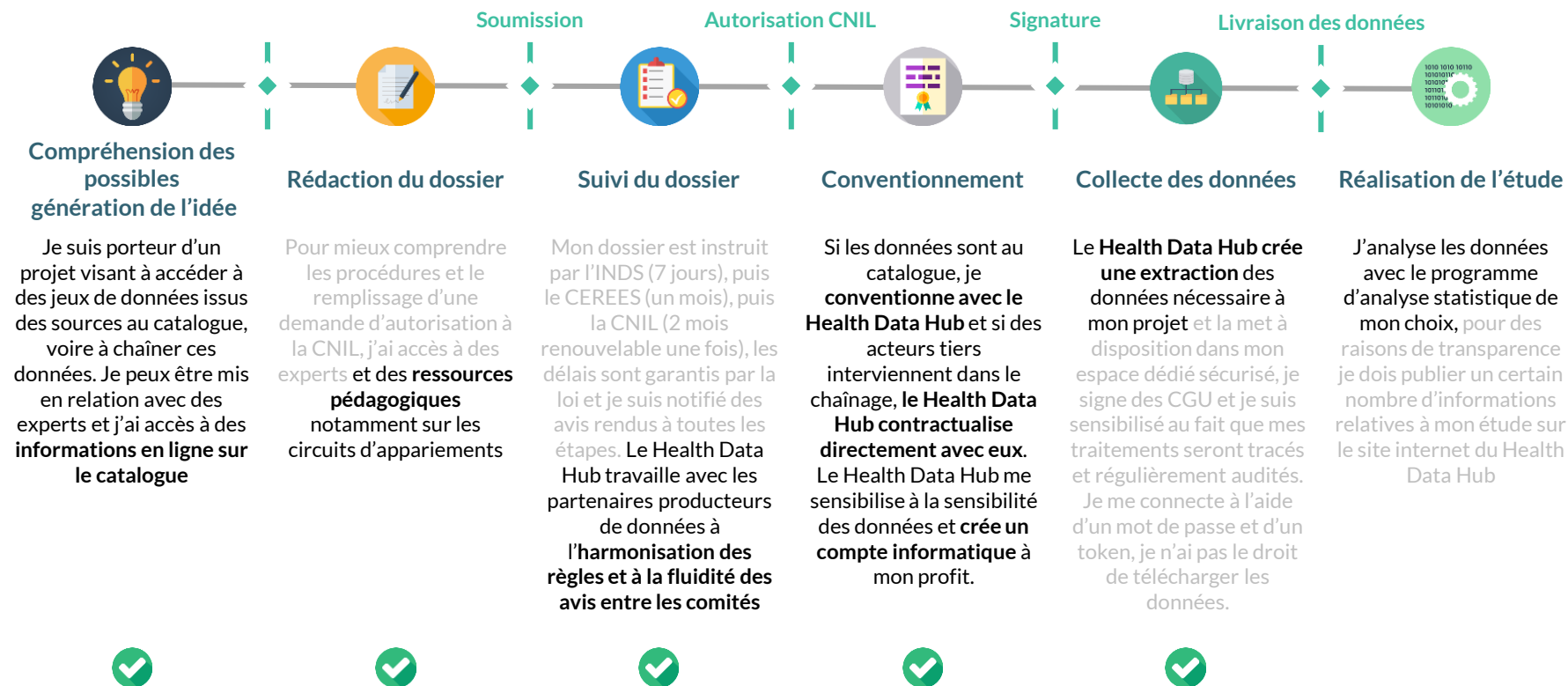
Catalogue de données



- Après le vote de la loi OTSS qui élargit le périmètre des données du SNDS à l'ensemble des données associées à un remboursement de l'Assurance Maladie, le HDH constitue de manière progressive et itérative un catalogue de données documenté pour mettre à disposition de la communauté scientifique les données stratégiques (SNDS historique, cohortes, registres, données hospitalières, données de ville...), ce catalogue est décrit dans un texte applicatif (arrêté).
- Le Hub participe à la clarification et à l'uniformisation des règles d'accès aux données .

Le Health Data Hub devrait faciliter les accès aux données du catalogue et leur chaînage

En renforçant les forces de l'INDS, le suivi des procédures sera complété par la réalisation des chaînages, la préparation des données et leur mise à disposition pour les utilisateurs du Health Data Hub qui souhaitent accéder à ces services





La mise en place du Health Data Hub prévoit des travaux d'identification de bases prioritaires à intégrer à son catalogue à destination des utilisateurs

Objectif



L'objectif du catalogue est de lister les bases stratégiques pour la communauté scientifique afin de faciliter leur accès et leur chaînage. Ce catalogue sera initialisé en 2020 et régulièrement mis à jour.

Il pourrait comprendre dans un premier temps le SNDS historique, certaines cohortes de recherche, certains registres, ainsi que d'autres bases de données identifiées par un groupe d'experts en cours de mise en place. Un travail spécifique devra être mené avec les établissements hospitaliers qui mettent en place des entrepôts sur l'ensemble du territoire et les acteurs liés à la production de données de médecine de ville.

Exhaustivité



Des données qui ne seraient pas au catalogue à un instant t pourraient toutefois être mises à disposition à la demande des utilisateurs, moyennant les délais nécessaires pour les rapatrier.

Le Health Data Hub n'a pas l'exhaustivité de la mise à disposition des données, et les producteurs qui le souhaitent ou qui ont les moyens en place peuvent, de toute évidence, assurer également ce service.

Metadonnées

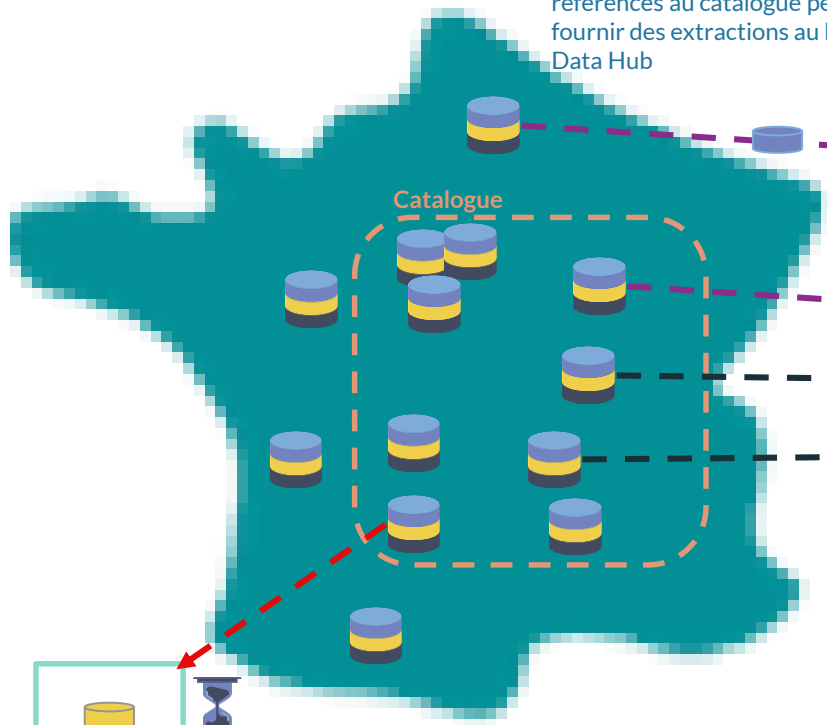


Seront incluses dans le catalogue des métadonnées et un certain nombre d'indications sur les bases relatives à leur contenu, leur qualité, leurs contributeurs afin d'initier une certaine harmonisation de la qualité des bases diffusées et faciliter la compréhension de leur contenu et de ce qu'il est possible de faire avec pour les utilisateurs.

Catalogue et mise à disposition des données - illustration



Les producteurs de données non référencés au catalogue peuvent fournir des extractions au Health Data Hub

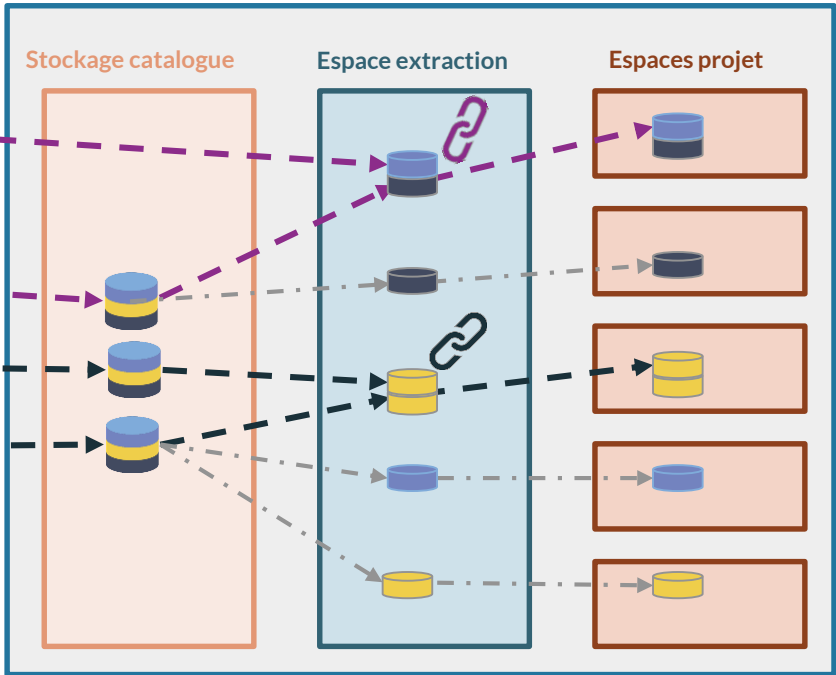


Bulle sécurisée ou portail



Les producteurs de données non référencés au catalogue peuvent fournir des extractions à l'utilisateur s'il a mis en place une bulle ou avec un prestataire ou son propre portail s'il en dispose

Plateforme sécurisée du Health Data Hub



Copie des bases référencées au catalogue
Le stockage est chiffré et sécurisé, l'hébergement est HDS, des audits et tests de pénétration sont réalisés, suivis par l'ANSSI

Ressources dédiées pour la préparation des extractions et la réalisation d'appariements

Mise à disposition sécurisée (double authentification, tracé des traitements)
Outils d'analyse à l'état de l'art (scala, R, python...)



Le Health Data Hub est régi par les règles de protection des données définies dans le RGPD et la LIL. Le recours au Hub ne donne pas accès à un régime dérogatoire ou une exemption en la matière.

Accès réglementé



L'autorisation d'accès et de traitement de ces données n'est accordée aux porteurs de projets qu'après habilitation auprès les instances compétentes, en particulier de la CNIL.

Lorsqu'un utilisateur obtient l'accès à un ou plusieurs jeux de données sur le Hub, ces dernières sont mises à disposition dans un espace cloisonné dédié se limitant à son besoin d'en connaître.

Données pseudonymisées



Les données stockées dans le Health Data Hub sont pseudonymisées, elles ne contiendront ni données nominatives, ni directement identifiantes.

Sécurité des données



Le référentiel de sécurité particulièrement strict encadrant le SNDS est étendu à l'ensemble des données de santé hébergées. Ce référentiel prévoit des dispositions techniques et organisationnelles pour couvrir les risques entourant les données parmi lesquelles la confidentialité, la traçabilité, l'intégrité.

Sensibilisation



L'accès aux données est nominatif et repose sur une double authentification. Les usagers sont systématiquement sensibilisés aux dispositions qui encadrent les traitements en amont de la mise à disposition des données.



Le fonctionnement du Health Data Hub est régi par les règles de protection des données définies dans le RGPD et la LIL. Le recours au Hub ne donne pas accès à un régime dérogatoire ou une exemption en la matière.



Information

On se situe dans la réutilisation secondaire de pour lesquelles il n'est pas possible de revenir au patient directement (données pseudonymisées). Dans ce contexte, la CNIL est l'autorité en charge d'arbitrer la légitimité des accès, la proportionnalité des données demandées et les conditions de sécurité des traitements. Une analyse d'impact sur la vie privée doit être menée et une information auprès des patients doit être prévue.



Consentement

Le consentement peut être demandé pour la constitution de certaines bases (cohorte de recherche) et pour certaines finalités.



Droit d'opposition

Le Health Data Hub a pour mission notamment de faciliter l'exercice du droit d'opposition du citoyen (en le mettant en relation avec les DPO des producteurs notamment).



Toute personne ou structure, publique ou privée, à but lucratif ou non lucratif, pourra accéder aux données de santé sur autorisation de la CNIL, en vue de réaliser une étude, une recherche ou une évaluation présentant un intérêt public.



Intérêt public

Aujourd'hui le CEIP qui juge de l'intérêt public suivant une doctrine publiée sur le site de l'INDS. L'intérêt public n'est pas incompatible avec l'intérêt commercial. Les missions du CEIP ainsi que celles du CEREES seront bientôt reprises par un unique comité éthique et scientifique.



Finalités

Seules quelques finalités de traitement sont autorisées, deux sont strictement interdites :

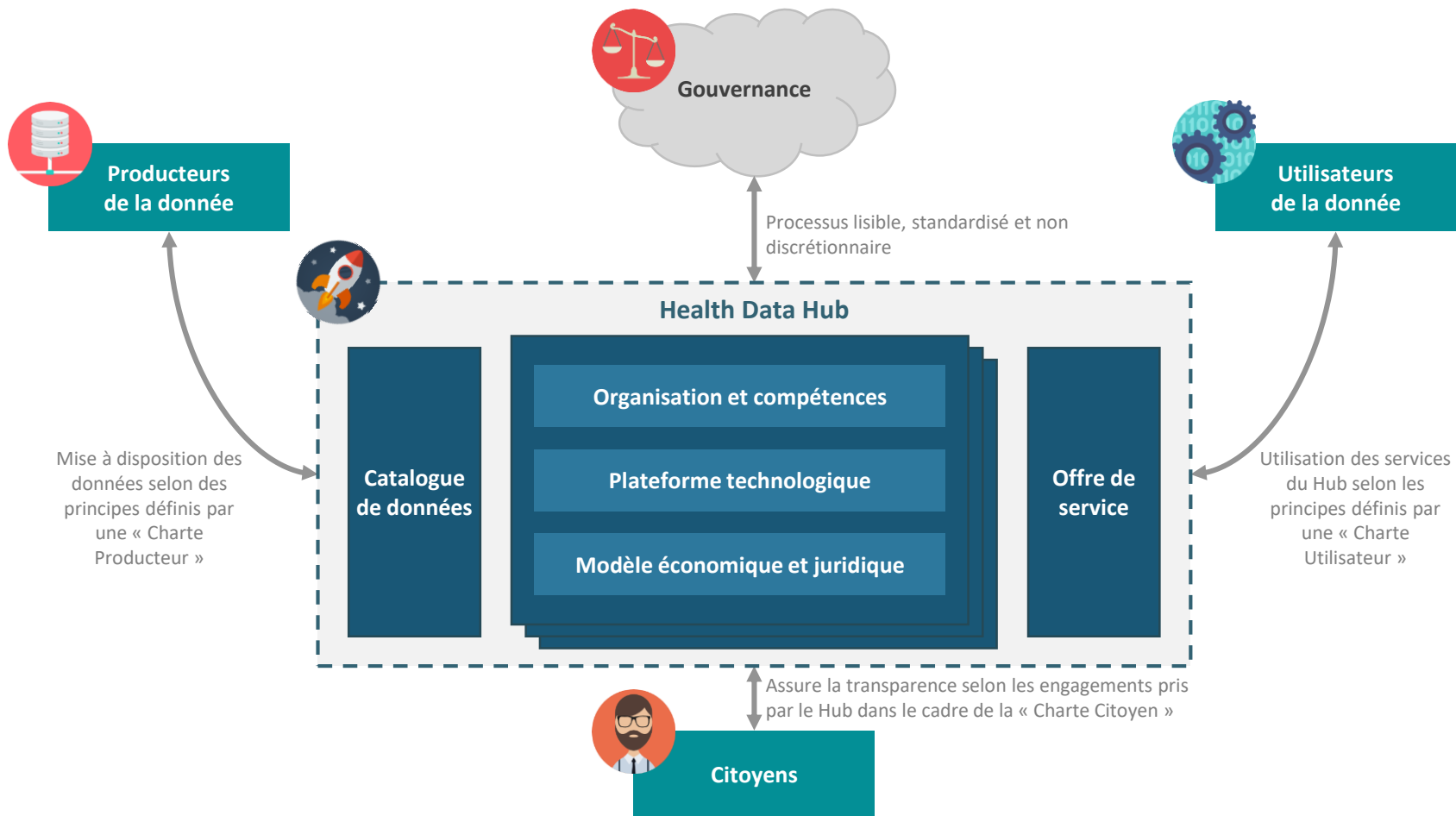
- la promotion en direction des professionnels de santé ou des établissements des produits de santé ;
- l'exclusion de garanties des contrats d'assurance ou la modification de cotisations ou de primes d'assurance pour un individu ou un groupe d'individus.



Accès direct

Les compagnie d'assurances ou industriels du produit de santé qui souhaitent accéder aux données doivent prouver qu'elles ne poursuivent pas ces finalités interdites.

Health Data Hub : la cible



Feuille de route 2019, les prochaines étapes

Fin 2018 - S1 2019

- La ministre confie à la **DREES de mettre en œuvre le projet** en novembre et à l'**INDS la mission de préfigurer le Health Data Hub** en février
- L'article 41 de la loi OTSS est **voté en première lecture à l'Assemblée Nationale** en mars. Il crée le Health Data Hub et élargit le SNDS
- **Spécification de la plateforme technologique** & gouvernance de cybersécurité au 1^{er} semestre 2019

- **Mise en place de la stratégie d'animation et engagement de l'écosystème de la recherche** & mise en place d'un GT "logiciels"

- **Des projets pilote** sont sélectionnés par appel à projet (publié le 31.01) et annoncés le 16 avril

- Publication du **starter kit** conçu grâce à l'accompagnement des projets pilote pour faciliter les démarches d'accès aux données de santé à l'ensemble des demandeurs (**documentations, programmes, données synthétiques**) associé à une offre de formation

S2 2019 - Début 2020

- Création de la **structure juridique et de ses instances de gouvernance** avant le 31 octobre 2019

- Colloque **intelligence artificielle en santé du 18 novembre**, pour expliquer la cohérence entre les actions Health Data Hub, espace numérique de santé & bouquet de services et le grand défi "comment améliorer le diagnostic avec l'intelligence artificielle", **annonce d'un nouvel appel à projet**

- **Test de la plateforme technologique** et de l'offre de service par les projets pilote au 4^{ème} trimestre 2019

- **Mise en place du CESREES**, restitution du GT données et établissement d'une première version du **catalogue de données**, et mise en place d'un GT "producteurs" pour déterminer la **gouvernance d'accès aux données**, et échanger autour du modèle économique

- **Premiers résultats des projets pilote** & ouverture des services à tous début 2020