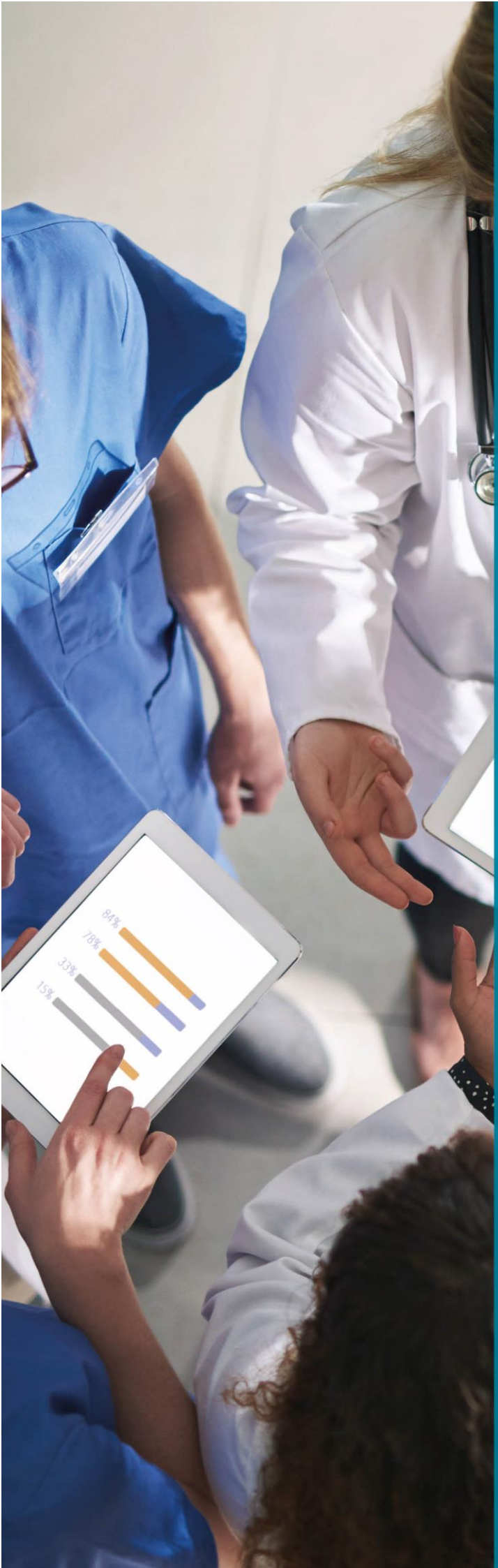


HEALTH DATA HUB

MISSION DE PRÉFIGURATION

Une mission pilotée par Marc Cuggia (CHU Rennes), Dominique Polton (INDS), Gilles Wainrib (OWKIN) et rapportée par Stéphanie Combes (DREES)

LA MISSION

A la suite de la remise du rapport Villani, le Président de la République a annoncé la création d'un « Health Data Hub » comme un des points forts de la stratégie Intelligence Artificielle française (IA). La Ministre des Solidarités et de la Santé a lancé le 12 juin 2018 une mission de préfiguration de cette plateforme d'exploitation des données de santé.

La mission, pilotée par trois experts, réunit également des représentants de la recherche, de l'écosystème des start-ups, de l'industrie, des professionnels et établissements de santé, de l'administration et de l'Assurance Maladie (cf. annexe pour une liste plus détaillée). Un très grand nombre d'acteurs de l'écosystème ont été entendus et/ou ont activement contribué à ces réflexions (cf. annexe). La mobilisation de la communauté a été très appréciée par la mission qui remercie vivement toutes les personnes avec qui elle a interagi.

Le rapport propose une feuille de route pour la mise en œuvre opérationnelle du « Health Data Hub », ainsi que des recommandations, notamment sur les aspects organisationnels et réglementaires pour que cette feuille de route puisse se dérouler dans un contexte favorable. Il s'appuie sur les entretiens réalisés, cependant ses conclusions n'engagent que ses auteurs.

INTRODUCTION

Déterminer des prises en charge adaptées et efficaces pour les maladies rares en agrégeant des observations de sources multiples. Dépister ou caractériser les états précancéreux grâce à l'intelligence artificielle. Doter les professionnels de santé d'outils pour accompagner le choix des meilleures options de prises en charge dans le contexte personnel du patient. Développer les essais cliniques virtuels. Suivre, en vie réelle et dans la durée, les impacts des innovations diagnostiques ou thérapeutiques et les effets croisés des prescriptions médicamenteuses.

Les 114 auditions menées dans le cadre de notre mission ont fait émerger près d'une centaine d'idées concrètes d'usages pour mettre notre patrimoine de données de santé au service de notre recherche, de nos praticiens et établissements de soin, des citoyens, de nos start-ups, de nos laboratoires et *medtechs* et de la puissance publique.

Un formidable potentiel est aujourd'hui à notre portée. Les freins ne sont pas d'ordre technique : les technologies de traitement sont matures et les données existent. Les acteurs consultés expriment au premier chef le besoin d'un guichet unique, assurant un accès simplifié, effectif et accéléré aux données. Qu'elles soient produites par le système de soins, les professionnels ou les patients eux-mêmes, les données de santé constituent un patrimoine extrêmement fragmenté. L'accès aux gisements de données de santé est régi par des gouvernances chaque fois différentes – parfois complexes. Les demandes d'accès et d'appariements peuvent impliquer des délais de traitement de l'ordre de plusieurs années. Du fait de la dispersion des moyens, les capacités technologiques et humaines requises pour valoriser les données dans un cadre sécurisé font souvent défaut et la capitalisation sur les méthodes employées reste l'exception. Les barrières sont aussi d'ordre culturel. Les nouvelles technologies, législations, méthodes et modes de travail intimident. Enfin, la valeur de la donnée – et surtout de son partage - restent méconnus et la plupart des acteurs témoignent de l'existence de réflexes propriétaires chez les producteurs.

Surmonter ces obstacles nécessite en premier lieu de prendre acte d'une responsabilité collective autour d'un principe fondateur : les données de santé financées par la solidarité nationale constituent un patrimoine commun. Ces données doivent donc être mises pleinement au service du plus grand nombre dans le respect

de l'éthique et des droits fondamentaux de nos concitoyens. Il est donc primordial d'en garantir l'accès aisé et unifié.

Le « Health Data Hub » doit être l'instrument de l'Etat au service de cette ambition.

En tant que tiers de confiance, il facilitera le partage en mettant en relation les producteurs et les utilisateurs publics comme privés selon un processus standardisé, lisible et non discrétionnaire. Il proposera un guichet d'accès unique à l'intégralité des données de santé soutenues par la solidarité nationale, accompagnera les procédures d'habilitation et réalisera les opérations d'appariements pour mettre à disposition des jeux de données documentés avec un engagement sur les délais. Il soutiendra la collecte et la consolidation des données, d'une part en mobilisant l'écosystème pour la mise en place de normes et standards, et d'autre part en veillant au financement et à la juste rétribution des efforts des producteurs. Il proposera des capacités technologiques, et un accès à des compétences rares à la demande – permettant ainsi aux acteurs ne disposant pas d'une taille critique de bénéficier de tous les moyens requis pour exploiter les données. Il assurera la transparence vis-à-vis de la société civile et des citoyens à travers un portail permettant de consulter les sources de données disponibles et leurs réutilisations. Enfin, des projets sélectionnés par le biais d'un appel à manifestation d'intérêt bénéficieront d'un accompagnement prenant la forme d'une mise à disposition de compétences et de capacités techniques, ou directement d'un financement.

Cette offre de service sera délivrée par une structure centrale appuyée par des pôles, opérant dans une logique de proximité géographique. Afin de garantir la soutenabilité du modèle économique d'ensemble et de rétribuer les coûts supportés pour la collecte des données et leur mise en qualité, l'accès aux services pourra notamment être facturé aux acteurs privés sous la forme d'un abonnement fixe et d'une part variable associée à l'usage. Des projets visant à consolider le patrimoine de données pourraient par ailleurs faire l'objet d'un co-financement entre acteurs institutionnels et privés.

Le présent rapport propose à la Ministre des Solidarités et de la Santé, des modalités d'organisation du « Hub », des principes d'intervention et des modalités juridiques et opérationnelles pour régir le partage de données. Nous avons aussi eu le souci d'esquisser une gouvernance, conçue comme le point de départ d'un projet qui a vocation à évoluer avec agilité au fur et à mesure de ses réalisations. Il propose également d'ouvrir un chantier législatif afin de renforcer le Système National des Données de Santé (SNDS) en élargissant son périmètre. Ce chantier sera l'occasion de définir un horizon clair à tous les acteurs et d'affirmer le principe d'un partage effectif des données.

Nous proposons d'engager sans attendre une phase de test visant à constituer et éprouver l'offre de service et la plateforme technologique autour de jeux de données ciblés et de cinq projets pilotes qui pourraient être financés, entre autres, par le grand défi « Comment améliorer les diagnostics médicaux par l'intelligence artificielle ? » également lancé dans le cadre de la stratégie Intelligence Artificielle du Président de la République. Le « Health Data Hub » enrichira ensuite son catalogue de manière progressive pour y intégrer, à terme, les principaux gisements du patrimoine de données de santé financés par la solidarité nationale. Rapidement et au vu des enjeux de recherche et de compétitivité, ces efforts gagneront à être inscrits dans la perspective d'une collaboration d'échelle européenne.

En constituant le Health Data Hub, nous réunirons les conditions d'un cercle vertueux : permettre aux acteurs du système de santé et aux citoyens de comprendre la valeur du partage des données, et par là même les convaincre de l'intérêt de faire grandir et d'entretenir notre patrimoine commun. L'enjeu est de taille : l'excellence de notre système de soins, son indépendance face aux intérêts étrangers, et la compétitivité de la France et de l'Europe dans un domaine économique critique. La mobilisation et l'engagement massif de l'ensemble des parties prenantes en appui à la réalisation de ce rapport – que nous tenions à saluer ici – nous ont convaincus que les conditions sont aujourd'hui réunies pour faire de la France un acteur de référence de la donnée de santé.

SOMMAIRE

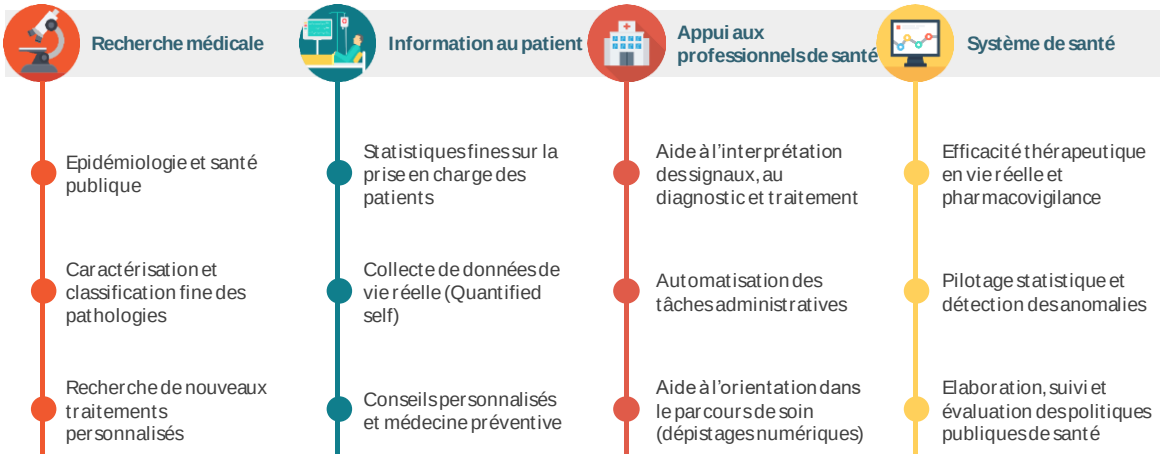
1

LES ENJEUX

1 LES ENJEUX

Le patrimoine de données de santé est une richesse nationale et chacun en prend peu à peu conscience.

Les sources de données sont de plus en plus nombreuses, riches, variées. Les techniques permettant de les exploiter arrivent à maturité et se diversifient. Pour l'ensemble des acteurs de l'écosystème consultés pendant les 114 auditions réalisées dans le cadre de la mission, un vaste champ de perspectives nouvelles – dont nous sommes encore probablement aux prémices - s'ouvre à nous.



Un renouveau des approches préventives, diagnostiques et thérapeutiques

Avec l'allongement de la durée de vie et les progrès de la médecine, les contextes cliniques des patients se complexifient. Par exemple, les personnes âgées souffrent de plus en plus d'une pluralité de pathologies et sont amenées à être suivies par différents professionnels de santé et structures médicales. Le praticien doit alors arbitrer pour offrir le meilleur traitement tout en prenant garde à la qualité de vie du patient, notamment lorsque les traitements sont lourds et admettent d'importants effets secondaires. Analyser d'importants jeux de données en mobilisant les approches dites d'« intelligence artificielle » peut alors éclairer sur la stratégie thérapeutique à adopter, patient par patient.

On constate également d'importants progrès dans la prise en charge du cancer. De nombreuses molécules innovantes dont l'indication dépend des anomalies moléculaires détectées dans les tumeurs sont actuellement à l'étude et arrivent sur le marché. Les patients connaissent ainsi plusieurs phases de traitement qu'il faut orchestrer. Si le choix d'une première stratégie thérapeutique est souvent aisé, cibler et évaluer les lignes de traitement suivantes devient déterminant pour la recherche et conditionne l'accès des patients à ces thérapies innovantes et coûteuses. Pour garantir les meilleures chances de succès, il faut mobiliser une expertise médicale beaucoup plus large, prendre en compte la réaction du patient aux premiers traitements reçus et ses spécificités individuelles ; et trouver des patients pour lesquels l'efficacité

d'un traitement a été constatée dans des conditions similaires. A cette fin, il devient indispensable de constituer de grands jeux de données mobilisant des dossiers patients de plusieurs centres hospitaliers pour avoir une masse critique permettant de réaliser des inférences performantes et précises. De plus, dans un contexte de médecine ambulatoire, la recherche ne peut plus reposer uniquement sur les données des dossiers médicaux hospitaliers, elle doit également mobiliser ceux de la médecine de ville, ainsi que les données produites par les patients eux-mêmes. Par conséquent, elle sera de plus en plus difficile à réaliser sans recourir à une part d'automatisation pour extraire de l'information pertinente dans de grands jeux de données hétérogènes. L'enjeu est de taille pour la prise en charge du cancer, puisque cette pathologie est encore responsable en France de 30% des décès (plus de 160 000 par an) avec 400 000 nouveaux cas par an et un coût annuel de 15 milliards d'euros.

Les techniques d'intelligence artificielle visent également à prédire des événements au sein d'une séquence. La capacité à anticiper est un enjeu essentiel pour la prise en charge tout au long du parcours de soin, par exemple pour anticiper, voire éviter un recours aux urgences dans le cas d'un patient insuffisant cardiaque. Dans la prise en charge de l'obésité, on pourra, par exemple, être en mesure de prédire, pour un patient donné, si l'opération n'est pas trop risquée. Dans le même ordre d'idée, on peut prévenir un rejet de greffon, une récidive de tentative de suicide ou une ré-hospitalisation, etc. Passer d'une médecine curative de masse à une médecine personnalisée, et ce pour toutes les spécialités, nécessite d'être capable de rassembler des grands jeux de données issus de plusieurs sources (remboursements de l'assurance maladie, données cliniques hospitalières ou de ville, données de population issues notamment des cohortes, objets connectés…) et des expertises pour les exploiter (cliniciens, spécialistes en informatique médicale, épidémiologie, biologie humaine…).

Des outils d'aide au diagnostic, à la décision et à l'interprétation de plus en plus performants

Dans la pratique, la mobilisation de ces approches sur des données médicales se concrétise par la publication d'articles scientifiques mais aussi par le développement de nouveaux outils qui mobilisent cette connaissance pour la restituer de façon opérationnelle au professionnel en fonction du contexte clinique. On parle alors d'outils d'aide à la décision diagnostique ou thérapeutique. Les options de prise en charge suggérées par ces outils pourront contribuer à enrichir les raisonnements des professionnels. Plus les données seront disponibles et de qualité, moins les outils développés à partir de celles-ci seront biaisés et plus ils seront efficaces. Cela signifie qu'ils prendront davantage en compte le contexte et produiront des recommandations plus pertinentes. A court terme, ce sont surtout les outils « d'aide à l'interprétation » que l'on voit émerger, car ils reposent sur des sources de données homogènes. A titre d'exemple, des start-ups travaillent au développement d'outils permettant de détecter une anomalie sur un électrocardiogramme et de gagner des heures précieuses pour venir en aide à un patient dans une situation d'urgence et en absence de cardiologue. De manière plus anecdotique, l' « Apple Watch » a récemment reçu l'aval de la *Food and Drug Administration* (FDA) grâce à sa fonctionnalité visant à détecter les problèmes de rythme cardiaque via des électrodes placées sur le côté de la montre. Encore rudimentaires, ce dernier type de mécanisme doit être appréhendé comme un complément mais pas comme une alternative à la consultation médicale lorsqu'elle est requise.

En analyse d'image, les algorithmes témoignent également d'une efficacité très intéressante pour le secteur de la santé. La FDA a ainsi autorisé le premier dispositif médical utilisant de l'intelligence artificielle pour dépister la rétinopathie diabétique à partir de photos de la rétine. La commercialisation du dispositif est autorisée aux Etats-Unis où il est utilisable en première ligne par des professionnels de santé non spécialisés en ophtalmologie, permettant ainsi de donner accès à ces soins au plus grand nombre. En automatisant

certaines analyses routinières, on permet ainsi aux experts de passer davantage de temps sur des cas plus complexes où ils devront interpréter des signes non présents dans les données. Aujourd'hui, l'élaboration de ces algorithmes requiert une annotation manuelle des signaux ou images par des spécialistes, mais on peut imaginer que le chaînage avec d'autres sources de données pourrait faciliter la qualification des données et dynamiser l'innovation en la matière. D'autres tâches peuvent être facilitées par des algorithmes ; plusieurs expériences visent à mobiliser les comptes-rendus médicaux pour automatiser le codage des actes et diagnostics à l'hôpital dans le cadre de leur remboursement par l'Assurance Maladie.

Un système de santé plus efficient

L'analyse des données de santé permet également de mieux piloter le système dans son ensemble. Pour les décideurs et ses relations avec les différents acteurs, il est indispensable de pouvoir prendre des décisions éclairées et objectivées par une utilisation accrue de la donnée. Les analyses de données de « vie réelle » (par opposition aux données collectées dans le cadre d'essais cliniques) permettent ainsi de mesurer l'efficacité thérapeutique avérée des traitements, dans le contexte de vie des patients et pour une population plus large que celle sélectionnée dans le cadre de l'essai clinique (profil fragile ou âgé, pathologies simultanées, traitements non-indiqués conjointement, suivi plus long). Ces analyses permettent également de détecter des effets indésirables et doivent être mobilisées au maximum dans l'exercice des missions des agences sanitaires, en particulier pour garantir la sécurité sanitaire et éviter des crises telles que celle du Médiator® ou encore celle des prothèses mammaires PIP®. Enfin, elles constituent un matériau indispensable à la formulation de recommandations en matière de politiques publiques en santé et à l'évaluation de leur implémentation.

On l'a vu, les cas d'usage sont nombreux et dans ce secteur sans doute plus que dans d'autres, l'enjeu de souveraineté nationale est prégnant. L'Etat a un devoir de garantir au citoyen une médecine de pointe tout en s'assurant de l'éthique de l'usage des données et de la fiabilité des innovations dont il bénéficiera. Compte-tenu de la concurrence internationale très forte, on peut s'attendre à l'émergence d'applications et d'innovations thérapeutiques en dehors du territoire qu'il ne sera pas possible de réguler et dont on ne pourra pas priver le citoyen. Une solution est à trouver pour encourager les efforts d'innovation consentis par les acteurs français, en particulier en leur favorisant l'accès à de grandes bases de données. Dans certains domaines, il est déjà naturel pour les acteurs de partager leurs données pour accélérer la science. On peut citer par exemple le TCGA, *The Cancer Genome Atlas* (initiative nord-américaine) ou l'*International Cancer Genome Consortium* (ICGC), initiative internationale rassemblant près d'une vingtaine de pays (dont la France et les Etats-Unis, une partie des données du TCGA sont incluses dans l'ICGC). Dans les deux cas, l'idée est de donner accès à des dizaines de milliers d'observations correspondant à quelques dizaines de types de cancers. Chaque fois, les partenaires se sont rapprochés pour partager leurs données, car elles étaient si chères à produire que la seule manière d'en avoir davantage était de renoncer à un droit d'exclusivité.

2

LES ATTENTES DES ACTEURS

2 LES ATTENTES DES ACTEURS

Les nombreuses opportunités proposées par l'écosystème le confirment : la valorisation des données de santé présente un formidable potentiel. Le patrimoine de données français représente un atout de taille. La France a été visionnaire il y a une quinzaine d'années avec la construction du système national inter-régimes d'assurance maladie (SNIIRAM), apparié depuis avec la base des résumés de séjour hospitaliers (PMSI). La création du Système National de Données de Santé (SNDS) en 2016 a constitué un nouveau jalon fort, dont chacun reconnait la valeur.

A l'étranger, de nombreux Etats, notamment les Etats-Unis et la Chine, font valoir des premiers succès emblématiques et un cadre facilité d'usage attirant ainsi les talents et entreprises innovantes. Les acteurs privés, comme les GAFAM et les BATX[1], avancent également et développent – en dehors de la sphère publique – des innovations qui vont demain être amenées à intégrer le quotidien de nos concitoyens. La souveraineté et l'indépendance de notre système de santé face aux intérêts étrangers, ainsi que la compétitivité de notre recherche et de notre industrie dépendront de la vitesse de la France à s'emparer du sujet.

Quels sont les verrous à lever pour accélérer et tirer le plein potentiel des données de santé financées par la solidarité nationale ?

Le tableau ci-après présente une synthèse des attentes exprimées par l'ensemble des parties prenantes rencontrées dans le cadre de la mission.


Offreurs de soins


Ecosystème de la recherche


Agences et autorités sanitaires


Représentants de la société civile, citoyens et patients


Industries de santé


Start-ups


Assurances

[1] GAFAM désigne les géants du web américains, à savoir Google, Apple, Facebook, Amazon et Microsoft, et BATX les géants du web chinois (Baidu, Alibaba, Tencent et Xiaomi).

Les freins rencontrés	Les attentes de l'écosystème
1. Mieux connaître et pouvoir plus facilement accéder au patrimoine de données de santé	
- Fragmentation du patrimoine de données ; - Manque de documentation, de modèles de données, ou d'échantillons pour apprécier les possibilités et la qualité des données a priori, entrainant de nombreux aller-retours pour les acteurs nouveaux.	**Pouvoir explorer, identifier, comprendre et apprécier a priori les possibilités offertes par les données disponibles** 
- Méconnaissance d'ensemble du cadre réglementaire et juridique, en évolution rapide ; - Complexité des procédures d'accès aux données avec des gouvernances discrétionnaires, spécifiques à chaque source et organisées en silo, absence d'obligation de réponse et d'engagement sur les délais, manque de lisibilité sur les critères de refus, absence de standard juridique européen homogène ; - Complication accrue pour les projets multi-sources, difficultés pour chaîner sur le NIR ; - Absence de dispositif d'ensemble permettant d'assurer la transparence vis-à-vis des citoyens ; - Une législation pouvant donner lieu à des interprétations restrictives, notamment l'article 193 de la loi de modernisation de notre système de santé, selon lequel le SNDS peut être utilisé pour des études, recherches ou évaluations, ce qui interdirait la constitution d'entrepôts de données appariés avec le SNDS et ouverts à de multiples usages potentiels.	**Harmoniser, simplifier et rendre lisible le processus d'accès aux données** 
- Éclatement des jeux de données, difficulté de constituer des bases atteignant une taille critique en nombre d'observations, par exemple pour les maladies rares ; - Absence d'un tiers de confiance national pour la mise à disposition de données chainées.	**Accéder à des grands jeux de données** 

Les freins rencontrés	Les attentes de l'écosystème

2. Avoir accès à des moyens adaptés et suffisants pour consolider et valoriser le patrimoine de données

- Données hétérogènes impliquant des efforts importants pour les numériser, les collecter, les rassembler, les harmoniser ;
- Difficultés d'extraction depuis les systèmes informatiques des éditeurs ;
- Faible interopérabilité sémantique des données.

Rétribuer les producteurs pour la constitution et la mise en qualité des bases

- Coût élevé pour traiter au bon niveau de sécurité ;
- Inadéquation des solutions actuelles face aux besoins d'innovation (GPU, R, Python, APIs, …) ;
- Mise en place d'expériences artisanales ou « shadow IT »
- Rareté des expertises (data science et informatique médicale, éthique et juridique sur les sujets de Big data et d'IA en santé), et des compétences requises pour les appariements ;
- Tensions sur le marché du travail et défaut d'attractivité des acteurs institutionnels.

Avoir accès aux capacités technologiques et aux compétences requises pour la valorisation

3. Fédérer l'écosystème autour d'un modèle économique d'ensemble favorisant le partage et reconnaissant les efforts de chacun

- Financement insuffisant des coûts de production, mais également de l'amélioration de la qualité, de l'adoption de standards internationaux et de documentation ;
- Absence de modèle économique pour les start-ups et les fournisseurs privés d'algorithmes ;
- Absence de modèle de référence sur le partage de la valeur et sur les valorisations scientifiques ou industrielles.

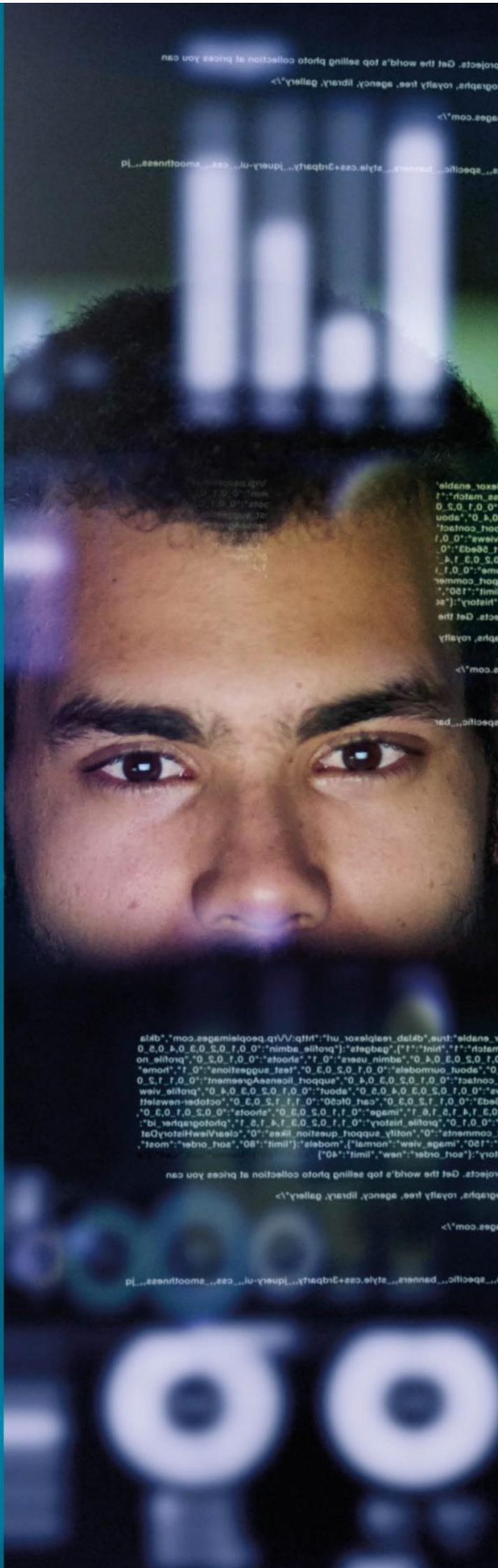
Établir un modèle de partage de la valeur entre producteur et utilisateurs

- Manque de sensibilisation aux enjeux de partage des données et des algorithmes, aux technologies et aux méthodes liées à l'intelligence artificielle ;
- Absence de politique de structuration pour des données de qualité ;
- Besoin de canaliser les efforts autour d'une vision nationale des enjeux stratégiques.

Au-delà du partage de la valeur, orchestrer l'ensemble de l'écosystème

3

NOS AMBITIONS



3 NOS AMBITIONS

Les recommandations de la mission procèdent d'un principe fondateur : les données de santé financées par la solidarité nationale constituent un patrimoine commun et doivent être mises pleinement au service du plus grand nombre dans le respect de l'éthique et des droits fondamentaux des citoyens.

Consolider et renforcer notre patrimoine de données

Le potentiel de réutilisation et la valeur des données découlent directement de leur qualité. Il est primordial de responsabiliser les producteurs sur la mise en qualité des données sources, des bases de données associées et d'allouer les moyens nécessaires, tout en évitant la dispersion des moyens collectifs. Ainsi, un grand nombre de registres et de cohortes sont financés par la collectivité. Ces bases de données sont stratégiques pour la France, mais leur production coûte très cher (plusieurs millions d'euros par an pour une cohorte) en raison d'une intervention humaine indispensable pour atteindre le haut niveau de qualité requis. Il est essentiel alors de s'interroger sur les mutualisations possibles, et notamment **d'utiliser les sources de données collectées dans le cadre des soins ou de leur remboursement pour venir enrichir systématiquement cohortes et registres**. Recourir davantage aux données médico-administratives permettrait de réduire significativement le coût de collecte tout en diminuant le délai de mise à disposition, grâce à une collecte manuelle limitée au strict nécessaire. Ces enrichissements permettent en outre d'augmenter la profondeur des données mais également de contrôler des biais de sélection, et de réduire les perdus de vue (entre 50 à 80% de patients de certains registres que l'on ne parvient plus à recontacter au bout d'un an). Parallèlement, la médicalisation des données médico-administratives permettrait un changement de paradigme. Au lieu de construire une nouvelle cohorte pour chaque pathologie ou axe d'analyse, il pourrait devenir possible de construire des cohortes à façon, avec une surface intéressante, sans s'exposer à un recrutement de plus en plus difficile.

Les gisements de données collectées dans le cadre des soins et de leur remboursement n'échappent pas, par ailleurs, à une nécessaire stratégie de mise en qualité. Non produites à des fins de recherche, ces données ne sont pas forcément utilisables à l'état brut. Elles doivent être nettoyées, structurées, mises en forme, voire enrichies. Les données de l'Assurance Maladie présentent l'avantage de la centralisation. Les données hospitalières résultent, en revanche, pour la plupart, du renseignement par les professionnels de santé de logiciels hétérogènes et multiples. Le constat s'accentue s'agissant de la médecine de ville. Pour tirer le meilleur parti de ces données**, il est indispensable de consentir les moyens financiers et humains nécessaires à l'harmonisation des systèmes d'information et à leur modernisation, d'opérer des choix de standards/terminologies pour certaines catégories de données et de mettre en place des politiques incitatives à la structuration et au partage des données** (de telles incitations pourraient par exemple trouver place dans la Rémunération sur Objectifs de Santé Publique pour les médecins généralistes). A titre d'exemple, la Suisse

investit entre 200 et 250 millions de francs suisses dans sa politique quadriennale de recherche en e-santé avec notamment un million de francs suisses par établissement hospitalier universitaire. En France, la mise en place des groupements hospitaliers de territoire est l'occasion d'une convergence des systèmes d'information des établissements concernés. Cette convergence ne doit cependant pas s'imaginer à un niveau local mais bien à un niveau national pour éviter des investissements nouveaux qui aboutiraient à un paysage national encore morcelé dans les années à venir.

A partir du moment où les données ont déjà bénéficié d'un financement pour les produire, le principe de gratuité doit prévaloir ; en revanche, il est légitime que le producteur de donnée reçoive une rétribution dès lors qu'il consent des efforts techniques pour les mettre à disposition et pour améliorer leur qualité. On pense, par exemple, à l'énorme travail qui consiste à retourner manuellement au dossier pour trouver certaines informations spécifiques mais indispensables (tâche qui pourra peut-être être, à terme, en partie automatisée par le biais des techniques de traitement automatique du langage). **Ces données ont par ailleurs un fort potentiel de valorisation économique. Il est essentiel de prévoir un juste partage de la valeur qui résulterait de leur utilisation, tout en gardant à l'esprit l'illégalité de la vente de données de santé non parfaitement anonymes**[2]. Enfin, au-delà des aspects purement financiers, les chercheurs, promoteurs d'une base de données de recherche, doivent pouvoir obtenir un retour sur le plan académique dans un contexte de forte compétition où les publications scientifiques sont indispensables pour candidater à certains financements ou accéder à certains postes. A ce titre, le Plan National pour la Science Ouverte, annoncé par la ministre Frédérique Vidal le 4 juillet 2018, prévoit une meilleure valorisation du partage de la donnée et de l'effort de constitution des bases sur le plan académique. Via la création d'identifiants uniques pour référer aux bases de données, il sera possible de construire des indicateurs de citations qui valoriseront les données de bonne qualité et largement diffusées, et bénéficieront au promoteur de la base au même titre que ses propres publications. **Il est essentiel d'accélérer ce mouvement dans la santé**, sans doute plus fermé que d'autres secteurs, qui ouvrent leurs données depuis des décennies (par exemple en astronomie).

[2] Article L1111-8 du code de la santé publique, VII.-Tout acte de cession à titre onéreux de données de santé identifiantes directement ou indirectement, y compris avec l'accord de la personne concernée, est interdit sous peine des sanctions prévues à l'article 226-21 du code pénal.

Faire du partage la règle, de la fermeture l'exception

Les données financées par la solidarité nationale doivent être partagées avec tous les acteurs, publics comme privés, et bénéficier ainsi au système de santé, à la recherche, au tissu industriel et à l'assurance du maintien de la souveraineté nationale sur un secteur stratégique. Ce partage doit se faire dans le respect de l'éthique et des droits fondamentaux du citoyen, notamment en pseudonymisant[3] les données.

Il est parfois utile de rappeler que **la donnée de santé n'est pas la propriété du producteur** aux yeux de la loi. Son traitement est notamment encadré en France par le Règlement Européen relatif à la protection des données personnelles (RGPD) et un chapitre spécifique de la loi informatique et libertés modifiée pour intégrer les dispositions de ce règlement. Le RGPD et la loi informatique et liberté prévoient l'usage de données de santé pseudonymisées pour le pilotage du système de santé, la réalisation d'études ou de recherches, pour la sécurité sanitaire, mais aussi lorsque la personne concernée donne son consentement explicite pour une finalité déterminée. La réutilisation secondaire des données de santé, notamment à partir de bases de données massives, est possible sans recueillir un consentement, à condition que le traitement présente une finalité d'intérêt public et que des garanties appropriées pour les droits et libertés des personnes concernées soient mises en place. Ces garanties passent par la mise en œuvre de mesures techniques et organisationnelles, notamment pour assurer le respect du principe de minimisation des données conservées. Pour ces traitements, l'autorisation de la CNIL est nécessaire[4]. Cette réutilisation nécessite en principe une information individuelle des personnes concernées sur les finalités de ce nouveau traitement, lorsque c'est possible, afin en particulier de leur permettre d'exercer un droit d'effacement de leurs données. Le RGPD permet toutefois de déroger à cette obligation d'information lorsqu'elle se révèle impossible ou demanderait des efforts disproportionnés. Il permet également de déroger au droit d'effacement pour des motifs d'intérêt public, ou à des fins de recherche, dans la mesure où ce droit est susceptible de rendre impossible ou de compromettre gravement la réalisation des objectifs du traitement.

La loi du 26 janvier 2016 pour la modernisation de notre système de santé a créé le Système National des Données de Santé (SNDS) qui regroupe pour l'instant les données des feuilles de soin (le SNIIRAM), les données utilisées pour la facturation des établissements hospitaliers (le PMSI) et la base des causes médicales de décès. Les données médico-sociales des MDPH (maisons départementales des personnes handicapées) et l'échantillon représentatif des données de remboursement par bénéficiaires transmises par les organismes d'assurance maladie complémentaire sont également attendus. L'inscription de ce dispositif dans la loi est un signal fort. L'Etat se porte ainsi garant d'un accès égalitaire pour tous les acteurs, publics comme privés, pour des finalités d'intérêt public et selon des règles de fonctionnement claires, transparentes, non discrétionnaires et opposables. La gouvernance du SNDS prévoit notamment la création

[3] On dit que les données sont pseudonymisées lorsque les variables directement identifiantes ont été retirées.

[4] La nouvelle loi «Informatique et Libertés » maintient en effet, un régime d'autorisation pour les recherches, études et évaluations dans le domaine de la santé qui contraste avec l'esprit d'allègement des formalités porté par le RGPD. Ce régime d'autorisation vient en complément de l'auto-régulation des pratiques qui incombe désormais aux acteurs qui doivent être en mesure de démontrer à tout moment en cas d'audit, la conformité de chaque traitement à la réglementation, ce qui pose de façon aigue la question du modèle économique de la mise en œuvre du texte et des moyens nécessaires à cette double exigence.

de l'Institut National des Données de Santé (INDS), chargé d'accompagner les utilisateurs dans le processus d'habilitation et d'un comité scientifique (le CEREES[5]) pour l'instruction des demandes[6].

La mise en œuvre de la loi pose un cadre qui devrait être appliqué à toutes les autres sources de données financées par la solidarité nationale. En élargissant le périmètre du SNDS, les acteurs de l'écosystème gagneraient en lisibilité, les droits et devoirs de chacun seraient clarifiés, les règles d'accès aux différentes bases seraient harmonisées, enfin le rapprochement des différents flux de données ne poserait plus de difficulté d'ordre juridique ou technique. Ce dernier point constitue aujourd'hui une grande source de frustration auprès des acteurs dissuadés, pour la plupart, d'entreprendre des projets qui s'avèreront longs et complexes, entraînant des délais peu compatibles avec le temps commercial ou de l'innovation. **Une évolution législative apparaît donc indispensable** à la mission et permettrait, par ailleurs, de valider sur le plan juridique l'enrichissement du patrimoine de données français (cohortes, registres et données cliniques) par des données médico-administratives.

Au-delà du cadre strictement légal, il est nécessaire de procéder à **la large diffusion d'une culture de la donnée**. Dans un contexte de forte concurrence scientifique, les efforts de partage des producteurs de données doivent être récompensés au travers de politiques d'incitation. Ainsi les financements publics devraient être systématiquement conditionnés à la reconnaissance et au respect du principe de partage. Les producteurs jouent incontestablement un rôle clé dans l'écosystème par leur expertise pointue, catalyseur de l'émergence d'innovations, et présentant des garanties éthiques et scientifiques élevées. Cet état de fait ne doit cependant pas être instrumentalisé pour justifier d'un droit d'exclusivité sur la valorisation d'un patrimoine collectif. Un effort de sensibilisation au potentiel des données de santé doit être fait envers les citoyens, les professionnels de santé et plus généralement auprès de l'ensemble des acteurs qui seront amenés à utiliser ce patrimoine. Par exemple, un enseignement spécifique pourrait à court terme être inscrit dans le cursus des études en santé. Si les patients hésitent peu à partager leurs données à des fins de recherche et d'innovation, c'est qu'ils ont conscience que partager ses données peut sauver des vies et relèvent d'un acte citoyen. N'attendons pas d'être souffrants pour épouser cet état d'esprit. La sensibilité de cette donnée appelle, certes, à un accès régulé, sécurisé et pour un usage responsable, mais elle nous oblige à faire évoluer fortement nos organisations pour en tirer le plus grand bénéfice pour la santé des patients.

[5] Pour une recherche impliquant des données du SNDS et la personne humaine (réglementation Jardé), c'est le CPP (comité de protection des personnes) qui aura vocation à se prononcer au lieu du CEREES. La procédure d'autorisation prévue fait intervenir des guichets et comités distincts selon que la recherche implique ou non la personne humaine, ce qui peut entraîner des difficultés lorsque les chercheurs veulent apparier des sources impliquant les deux comités de part la juxtaposition des procédures que cela entraîne.

Mettre en synergie les moyens techniques et humains et soutenir les initiatives prometteuses

Des initiatives valorisant ces données de manière innovante et pointue se mettent en place sur tout le territoire. Au-delà des moyens nécessaires à la collecte, les acteurs témoignent de difficultés à rassembler des capacités à faire. Les compétences en architecture ou sécurité informatique sont rares et chères, de même que les compétences d'ingénierie et de science des données. Pour pouvoir mener des travaux d'analyse des données, les acteurs investissent dans la constitution d'infrastructures technologiques parfois sous dimensionnées, voire artisanales, faute de moyens. **Il semble essentiel de mettre à disposition de ces acteurs des capacités technologiques et humaines mutualisées, afin d'atteindre une taille critique permettant une industrialisation et une sécurisation des processus**. S'agissant des données de santé, cette mise en synergie à travers un « Hub données de santé » offre en outre la possibilité de mieux maîtriser leur circulation et de faciliter l'audit, la traçabilité et la transparence des traitements[7].

La mise à disposition de capacités à faire dynamisera l'écosystème. La promotion de *l'open source*, de standards ouverts peut renforcer ce phénomène en diffusant plus rapidement la connaissance et le progrès au sein de la communauté. Les grands acteurs institutionnels publics pourraient d'ailleurs prendre à leur compte la responsabilité de créer un élan de cette nature. En effet, la loi pour une République Numérique prévoit que tout algorithme produit par une administration est un document public devant être versé à la collectivité dès lorsqu'une personne le demande[8].

Au-delà des capacités à faire, certains cas d'usage pourraient bénéficier d'un accompagnement de bout en bout. En effet, assurer des succès rapides contribuerait à la fois à améliorer le système de santé et à diffuser la culture de la donnée. **Des projets démontrant l'intérêt du partage et de l'exploitation de nos gisements de données pourraient favoriser l'obtention de premiers résultats rapidement dans des domaines où les enjeux en termes de santé publique ou de promotion d'une filière industrielle sont forts**. La sélection de ces projets pourrait s'effectuer **via un appel à manifestation d'intérêt qui documenterait la valeur intrinsèque de la candidature, sa faisabilité et sa contribution à la politique de santé et à la politique de partage de la donnée.** Cette mise en avant de sujets « phare » pourrait se faire en lien étroit avec la nouvelle feuille de route du comité stratégique de filière et les défis Intelligence Artificielle et Santé pilotés par le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation et le ministère de l'Economie et des Finances.

[7] Dans l'avis n° 129 qu'il vient d'émettre préalablement à l'engagement de la révision bioéthique, le Comité consultatif national d'éthique indique que « la création d'une plate-forme nationale sécurisée de collecte et de traitement des données de santé constitue une piste intéressante pour articuler entre eux les différents enjeux éthiques afférents aux données de santé. La définition du mode d'alimentation de cette plate-forme relèverait d'un niveau législatif si le choix était fait de mettre en œuvre un mécanisme de consentement présumé dans le cas d'un intérêt public pour la santé du type de celui existant en matière de prélèvement d'organes. Un tel choix permettrait plus de lisibilité et d'efficacité dans le fonctionnement du dispositif dans son ensemble. »

[8] Le principe est posé par l'article Art. L. 311-3-1 et les exceptions par l'article L311-5 2°

Permettre la structuration d'une filière Intelligence Artificielle (IA) et Santé

Les avis des analystes convergent sur une croissance à deux chiffres du marché des produits et services basés sur l'IA jusqu'à au moins 2025 et une taille de marché dépassant les 10 milliards de dollars à l'horizon 2020. L'IA pourrait ainsi devenir le segment le plus dynamique du secteur de l'économie numérique. La France manque toutefois aujourd'hui d'entreprises ayant une forte notoriété en IA et de taille suffisante pour avoir un effet d'entrainement sur une filière technologique ou sectorielle. La plupart des grandes entreprises de référence françaises ne semblent pas avoir construit, en matière d'IA, de stratégies proactives nécessitant le développement d'écosystèmes autour d'elles. Concernant les startups françaises de l'IA, aucune n'a, pour l'instant, une rapidité de croissance suffisante pour avoir un effet d'entraînement.

Il est donc **nécessaire de s'interroger sur la manière de faire émerger, dans un contexte concurrentiel dominé par de grands acteurs étrangers du numérique qui investissent massivement et attirent les talents du monde entier, des *leaders* français** en la matière. Il conviendrait a minima de développer, en France, un cadre favorisant, d'une part la sensibilisation de l'ensemble des filières aux enjeux et aux potentiels de l'IA, et d'autre part, les coopérations entre acteurs (recherche publique, startups, grandes entreprises, incubateurs, investisseurs, etc.) en matière d'IA. **Sur le champ de la Santé, le Hub pourrait favoriser la mise en relation de ces acteurs, et faciliter la création de consortiums** à l'instar de l'initiative Hu-PreciMED pour la médecine de précision ou pour répondre à des appels d'offre européens ou encore contribuer à des projets qui seront portés dans le cadre du nouveau comité stratégique de filière.

En particulier, **le Hub pourra, en tant que tiers de confiance, permettre le partage de données publiques, comme privées, dans un espace sécurisé et neutre facilitant la contractualisation de partenariats entre les acteurs**. Le développement de l'IA dépend beaucoup de l'accessibilité de données massives et soulève de multiples questions éthiques, législatives et de valorisation économique auxquelles le Hub entend apporter des premières réponses. Sans ce dispositif, associé à des appels à projets, les entreprises iront investir dans d'autres pays. Certains s'engagent aujourd'hui activement pour faciliter les partenariats publics-privés en IA et santé et deviennent à ce titre très attractifs (Angleterre, Danemark, Israël, etc.). Bien que les données soient estimées d'une grande valeur potentielle, cette valeur n'est ni certaine ni garantie dans la durée. A mesure que les innovations se développeront et exploiteront les données disponibles, les gisements perdront de leur valeur. Il y a donc également un enjeu économique fort à être réactif et à dépasser les blocages liés à la quête d'un partage parfaitement juste d'une valeur potentielle.

A noter toutefois, qu'au-delà des cas d'usage permis par une plateforme de cette nature, la transformation de l'essai ne sera possible qu'en prévoyant un réel modèle économique aux innovations en santé. Ce point doit s'inscrire dans la réflexion en cours dans le chantier « modes de financement et de régulation » de la Stratégie de Transformation du Système de Santé (STSS).

Pour l'atteinte de ces quatre grandes ambitions, la mission recommande donc d'adopter une politique de la donnée de santé ambitieuse et de se munir de ressources nécessaires à son implémentation au niveau du Ministère des Solidarités et de la Santé. Les piliers de cette politique pourraient être :

- Le principe de partage de la donnée suivant une gouvernance unifiée et garantie par la Loi ;
- La sécurisation d'importants moyens pour consolider le patrimoine de données français et les initiatives qui structurent ces gisements sur le territoire national tout en conditionnant ces moyens à une responsabilisation des acteurs en termes de mise en qualité et d'adhésion à un principe de partage ;
- La création d'un tiers de confiance national sous la forme d'une structure partenariale publique-privée accompagnant producteurs et utilisateurs, favorisant l'accès aux données, leur valorisation à l'état de l'art et le développement de cas d'usage pertinents. Ses activités reposeraient en particulier sur la mise en place d'un environnement technique répondant aux problématiques de sécurité informatique, de confidentialité des données, de non dispersion des moyens et garantissant un partage des données en conformité avec le cadre réglementaire[9] ;
- L'identification des projets pilotes prometteurs en termes de santé publique et de promotion de la filière industrielle ;

Certains grands acteurs institutionnels se sont d'ores et déjà positionnés comme partenaires et pourront être associés aux chantiers de mise en œuvre opérationnelle, notamment les projets « pilotes » : l'Institut National du Cancer (INCa), l'Assistance Publique – Hôpitaux de Paris (AP-HP), le réseau des centres de données cliniques du Grand Ouest, l'Institut national de la santé et de la recherche médicale (Inserm), la Caisse Nationale de l'Assurance Maladie (CNAM), Santé Publique France et la Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques (DREES). D'autres pourraient les rejoindre : Unicancer, les Hospices Civils de Lyon (HCL), les futurs Instituts interdisciplinaires d'Intelligence Artificielle (Instituts 3IA), ainsi que de nombreux autres producteurs de données et grands organismes et opérateurs de recherche français (INRIA, CNRS, CEA…) voire européens.

[9] En lien avec la délibération de la Commission Nationale de l'Informatique et des Libertés n°2016-316 du 13 octobre 2016 portant avis sur le projet de décret en Conseil d'Etat relatif au SNDS, et le rapport de la Cour des comptes sur les données personnelles de santé gérées par l'assurance maladie rendu le 3 mai 2016

4

LE HEALTH DATA HUB

4 LE HEALTH DATA HUB

La mission propose une première cible d'ensemble et des premières hypothèses concernant les principes, l'organisation, les moyens et les modes de fonctionnement du Health Data Hub. Ce modèle sera amené à s'affermir dans les phases de concertation à venir.

Vision d'ensemble

Le Health Data Hub pourrait être l'instrument de l'Etat au service d'une ambition : mettre le patrimoine des données de santé financées par la solidarité nationale au service du patient et du système de santé dans le respect de l'éthique et des droits fondamentaux de nos concitoyens.

- **Moteur et fédérateur**, le Health Data Hub représenterait - avant tout - un tiers de confiance dans le paysage des acteurs de la donnée de santé, et permettrait de ce fait le partage des données dans le respect du droit des patients et en assurant la transparence avec la société civile. Il proposerait, par ailleurs, une ambition et des objectifs communs par le biais de son conseil stratégique.

- Il constituerait une **vitrine nationale et internationale** à travers un patrimoine de données et une offre de service lisibles pour tous les acteurs publics comme privés.

- Il constituerait un **guichet unique** afin de faciliter l'accès aux données (avec des règles d'accès transparentes et non discrétionnaires pour les acteurs publics comme privés, un accompagnement dans les procédures de demande d'accès et dans la valorisation économique, une documentation des bases partagées, des échantillons de test, des capacités techniques et juridiques pour réaliser des appariements, etc.).

- Il constituerait un **garant de la qualité** de la donnée partagée : en lien avec les producteurs et les utilisateurs, il favoriserait la diffusion de standards de qualité internationaux, et implémenterait des circuits de valorisation de l'effort de collecte et de mise en qualité des producteurs.

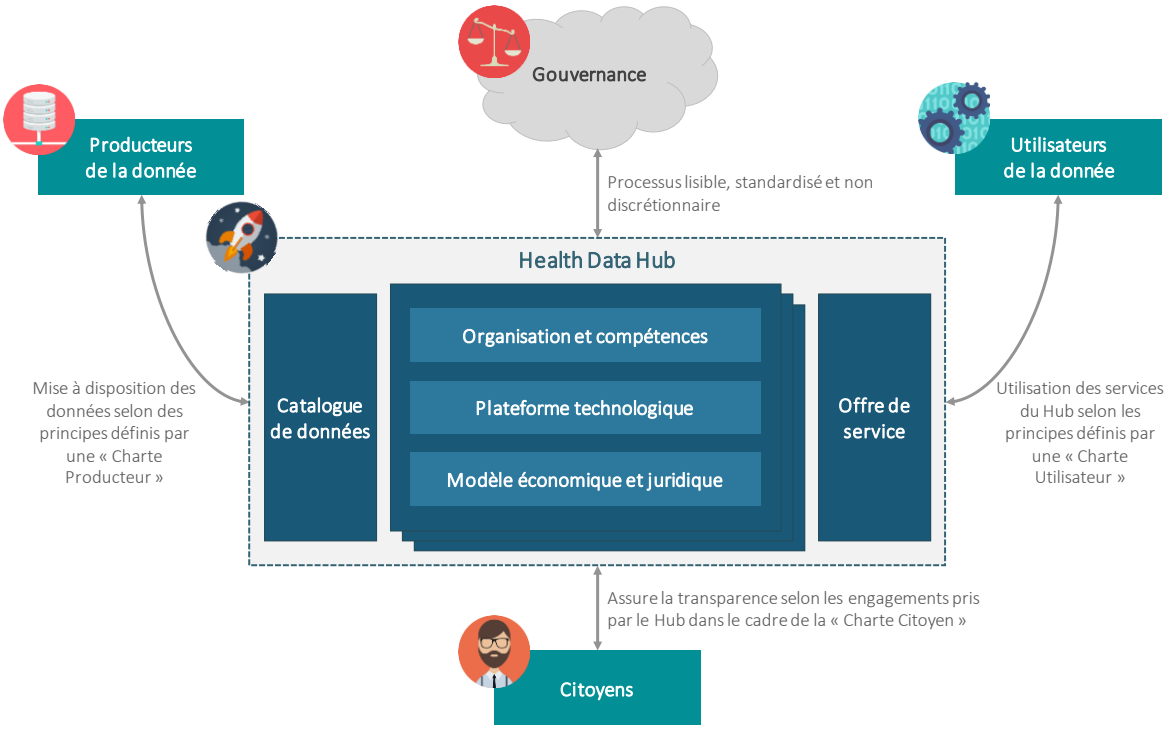
- Il jouerait un rôle essentiel dans la **promotion de l'innovation** avec le développement d'un environnement dans lequel les innovations peuvent prospérer : échantillons de données test, développements d'API, modèle économique favorisant l'innovation (gestion des questions de propriété intellectuelle et de partage du risque…), délais compatibles avec les contraintes des petits et gros industriels.

- Enfin, il permettrait la **mutualisation de ressources technologiques et humaines** avec une plateforme technologique sécurisée pour le partage des données et des expertises détenues en propre pour assurer le lien entre utilisateurs, producteurs et experts de la donnée. Ces expertises ne se substitueraient pas à celles de la communauté qui seraient valorisées via le rôle d'animation du Hub.

Fonctionnement général

Pour jouer son rôle de tiers de confiance entre les producteurs[10] de données publics et privés (acteurs institutionnels, offreurs de soins, organismes et opérateurs de recherche, autorités sanitaires, laboratoires, etc.), les acteurs souhaitant les valoriser, et les citoyens et représentants de la société civile, le Health Data Hub pourrait admettre le fonctionnement et l'offre de service décrits ci-après.

Ce fonctionnement et cette offre de service seraient de nature à répondre aux trois défis traditionnels des plateformes de données : le défi technique (sécurité, rapidité, format etc.), le défi légal et éthique (autorisations, etc.) et le défi de la confiance, du partage entre les acteurs.



[10] Une même entité pourra naturellement être à la fois productrice de données et utilisatrice des services du Hub. C'est par exemple le cas des CHU qui ont par leur nature universitaire un rôle de formation par et pour la recherche.

Son fonctionnement pourrait ainsi reposer sur :

- Une gouvernance de la donnée et des principes de collaboration standardisés, lisibles et non discrétionnaires, régissant les liens entre les producteurs, le Hub, les structures de gouvernance et les utilisateurs de la donnée ;

- Une offre de service à destination des utilisateurs publics, privés et des citoyens autour de quatre missions :
    - Donner accès aux données de santé ;
    - Soutenir la collecte et la consolidation des données ;
    - Accompagner la valorisation des données de santé ;
    - Soutenir l'écosystème et assurer le lien avec les citoyens et la société civile ;

- Un catalogue de données, qui sera enrichi progressivement pour référencer à terme les principaux gisements de données de santé financées par la solidarité nationale, et un espace « projets » ;

- Une organisation en réseau, articulant un Hub central et des Hubs « locaux » chargés de mettre en œuvre l'offre de service, dont un pourrait être dédié à la recherche sur les données de biologie et santé ;

- Une plateforme technologique, conçue et dimensionnée pour assurer l'offre de service dans les conditions requises de sécurité et pouvant être administrées par les Hub locaux sur leur périmètre de responsabilité. Cette plateforme présenterait un niveau élevé de sécurité compte tenu de la sensibilité des données ;

- Un modèle juridique et économique, qui doit permettre de garantir la soutenabilité du modèle économique d'ensemble, de compenser pour partie les coûts supportés pour la collecte des données, leur mise en qualité, le fonctionnement du Hub et de traiter de manière transparente la question de la propriété intellectuelle.

Gouvernance de la donnée

Les données de santé financées par la solidarité nationale constituent un patrimoine commun. En tant que tiers de confiance, le Hub aurait la responsabilité d'en faciliter le partage et la valorisation en mettant en relation les producteurs et les utilisateurs selon un processus standardisé, lisible et non discrétionnaire.

Vis-à-vis des utilisateurs, le Hub constituerait un « guichet unique » auprès duquel ils pourront solliciter l'accès à toutes les données du catalogue[11]. Ce processus pourrait en grande partie s'inspirer de celui mis en place avec la création du SNDS. Aujourd'hui, chaque projet de recherche, étude ou évaluation, mobilisant des données du SNDS et n'entrant pas dans le cadre des accès permanents, doit soumettre une demande auprès de l'INDS (sept jours maximum pour transmettre au CEREES), puis recueillir l'avis sur la méthodologie proposée, entre autres, auprès du CEREES[12] (avis donné dans le délai d'un mois), et enfin obtenir l'autorisation de traitement des données auprès de la CNIL (retour au bout de deux mois maximum renouvelables une fois). S'ensuit un délai de mise à disposition des données par la CNAM sur son portail. Les délais d'accès ont été considérablement réduits et sont aujourd'hui de l'ordre de trois à six mois (cf annexe). Bien que l'accès effectif puisse paraître encore important pour certains acteurs, les règles présentent l'avantage d'être lisibles[13]. Pour les autres sources de données, les procédures sont parfois moins transparentes. Chacune dispose de sa propre gouvernance (comité éthique, règles de refus, délai de réponse, règles de valorisation économique, conditions de partenariat…). Les acteurs témoignent du manque d'une cartographie des bases et des règles d'accès.

En cible, il serait souhaitable que les règles d'accès aux données de santé convergent, ce qui pourrait se concrétiser pas la mise en place d'un seul[14] comité éthique et scientifique (CES), à géométrie variable, adoptant des règles proches de celles retenues pour le SNDS. Ce comité pourrait, par exemple, se réunir une fois par mois pour statuer sur les demandes d'accès aux données mises au catalogue du Hub. Il mobiliserait un ou plusieurs experts selon la nature du projet (représentant des patients, éthiciens, experts des différentes sources mobilisées, experts en intelligence artificielle, etc.) et des représentants des producteurs des données visées pour juger de la faisabilité d'un projet si nécessaire. La composition du comité pourrait être organisée par le Hub pour chaque séance en visant de conserver un panel d'experts réduit. Si plusieurs jeux de données de même nature proviennent d'acteurs différents, le Hub pourrait sélectionner un unique

[11] Ce guichet n'exclut pas une demande d'autorisation CNIL si la demande d'accès ne s'inscrit pas dans le cadre d'une méthodologie de référence ou une procédure simplifiée.

[12] Le CEREES émet un avis sur la méthodologie retenue, sur la nécessité du recours à des données à caractère personnel, sur la pertinence de celles-ci par rapport à la finalité du traitement et, s'il y a lieu, sur la qualité scientifique du projet.

[13] La mise en place du Hub pourrait être de nature à réduire les délais d'accès au SNDS en contribuant à faciliter une partie de l'aval (conventionnement, extraction, mise à disposition). Le rôle de « guichet unique » joué par le Hub doit permettre de faciliter l'accès à des sources multiples. Il n'a pas vocation à être un guichet « exclusif ». En particulier, la création du Hub ne remet pas en cause le rôle de la CNAM dans la mise à disposition des données du SNDS. En mutualisant une infrastructure à l'état de l'art et conforme au référentiel de sécurité, le Hub pourrait également éviter aux acteurs détenteurs d'extraction du SNDS (ou systèmes fils) ou souhaitant en détenir de procéder à une mise en conformité de leurs systèmes d'information, procédure longue et onéreuse. La mise en place du Hub pourrait, enfin, justifier la mise en place de nouvelles méthodologies de référence pour accélérer encore l'accès.

[14] Afin de ne pas créer d'engorgement, ce comité pourrait avoir une réplication obéissant aux mêmes principes au niveau des « Hub locaux », référents sur les demandes d'accès aux données locales – voir 4.5.1 Organisation générale en réseau du Hub.

représentant compétent. Enfin, la procédure pourrait être adaptée lorsqu'une demande est très proche d'une demande similaire acceptée précédemment, qu'elle ait été déposée par la même personne ou par une autre[15].

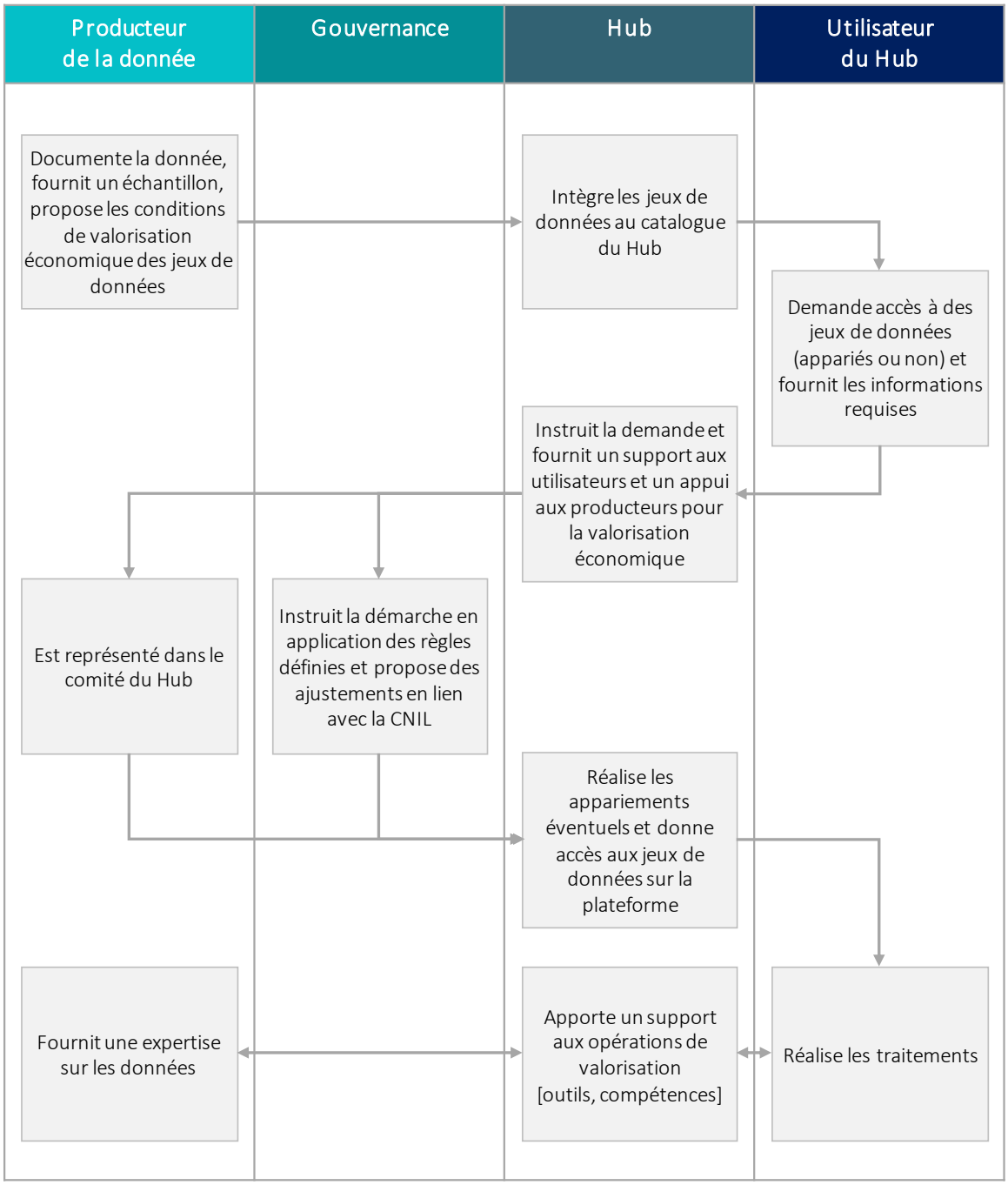
A court terme, il semble raisonnable à la mission que l'activité des divers comités scientifiques et éthiques organisés par les différents partenaires du Hub se maintienne dans une période transitoire, sous réserve de l'application de règles communes et transparentes d'évaluation des dossiers dans des délais fixés et de la justification des refus le cas échéant. Ces règles communes et transparentes seraient élaborées en concertation avec les acteurs de l'écosystème dont certain disposent d'un savoir-faire en la matière[16]. Le Hub pourra accompagner les producteurs dans cette démarche progressive de convergence vers des principes communs. En cas de projet mobilisant les données de plusieurs partenaires, il est proposé que le Hub coordonne l'accès aux différentes sources via un seul comité éthique et scientifique qui correspondrait, dans son fonctionnement, au comité prévu en cible et décrit précédemment. L'idée est de se prémunir d'une accumulation de gouvernances et d'extension des délais de mise à disposition des données, dans un contexte où les appariements sont des procédures juridiquement et techniquement complexes. Une évolution législative, à mener en concertation avec la CNIL, pourrait être de nature à simplifier ce dernier point.

S'agissant de la valorisation économique entourant la mise à disposition des données auprès d'acteurs privés, le Hub pourrait fournir une assistance sous la forme d'expertise de valorisation et de modèles de contrat qui auraient vocation à faciliter les négociations et à garantir un accès effectif aux données dans des délais raisonnables et annoncés. Ces modèles seraient co-construits avec les partenaires du Hub lors de la phase de préfiguration et dans le contexte de projets pilotes.

Pour les données partagées via le Hub en dehors du catalogue et dans le cadre de projets précis, les règles pourraient temporairement différer. Pour des contraintes de confidentialité ou de compétitivité de la recherche ou industrielle, par exemple pour un jeu de donnée n'ayant pas fait l'objet d'une publication, le chef de projet pourrait demander une fermeture temporaire de la donnée : les données ne feraient alors pas l'objet d'un partage par défaut, sauf validation explicite du chef de projet et producteurs pour un usage bien déterminé. Cette classification ferait l'objet d'un réexamen à échéance régulière par la gouvernance du Hub.

[15] Les accès permanents au SNDS pour les acteurs institutionnels ne seraient pas remis en cause. Cet accès permanent n'exempte pas les utilisateurs de documenter les traitements conformément au RGPD.
[16] A titre d'exemple, à l'échelle nationale, les CHU mènent également des travaux d'harmonisation sous l'égide de la conférence des Directeurs Généraux (GT Données Massives en Santé).

Le schéma suivant synthétise le processus associé à la collecte, l'accès et à l'usage des données et le partage des rôles associé :



Principes de collaboration

Les engagements respectifs dans leurs rôles de producteurs et/ou d'utilisateurs et du Hub pourraient être formalisés sous la forme de trois chartes (« Producteur », « Utilisateur » et « Citoyen ») qui seraient établies en concertation avec les parties prenantes, incluant notamment la CNIL, et prévoiraient de manière transparente et harmonisée les conditions d'accès au service du Hub, les engagements, rôles et responsabilités de chacun des acteurs et les modes de fonctionnement communs :

- La charte « Producteur » décrirait notamment la gouvernance d'accès aux données et les exigences communes à toutes les bases en termes de qualité et de structuration des données, de documentation et d'identifiants communs, de manière assez similaire aux principes « FAIR » (Faciles à trouver, Accessibles, Interopérables et Réutilisables). Elle indiquerait aussi les ressources et services du Hub dont les producteurs pourraient bénéficier pour atteindre ces critères (enrichissement de leurs données, hébergement, valorisation économique, information autour de l'utilisation des données, monétisation de leur expertise autour de l'interprétation et de l'utilisation scientifique des données, mutualisation des compétences rares, etc.). Les avantages compétitifs dont disposeraient les producteurs pour l'utilisation des bases qu'ils produisent à des fins de recherche devraient être proportionnés, définis et transparents.

- La charte « Utilisateur » poserait le cadre de référence et les principes généraux d'utilisation et de traitement des données partagées dans le Hub. Elle couvrirait les conditions et règles d'accès au service, ainsi que les modalités de partage des résultats de recherche et de certains traitements opérés (notamment si ces traitements visent à améliorer la qualité des bases de données du patrimoine) après écoulement d'une période d'exclusivité. Ce cadre de collaboration s'inscrirait dans le respect de règles claires relatives à la protection des données personnelles d'une part, et à la propriété intellectuelle et industrielle d'autre part. La charte « utilisateurs » poserait aussi des principes de notification des producteurs lors des publications et de crédits relatifs à la donnée, en cohérence avec le Plan National pour la Science Ouverte qui promeut la création d'indicateurs de citations autour de la réutilisation des données et non restreints aux seules publications.

- La charte « Citoyen » formaliserait les engagements de transparence, d'éthique et de respect des droits fondamentaux que l'ensemble des acteurs du Hub prennent vis-à-vis des citoyens et de la société civile.

Offre de service

Le Hub proposerait aux utilisateurs publics et privés une offre de service standardisée autour de quatre axes :

- Donner accès aux données de santé ;
- Soutenir la collecte et la consolidation des données de santé ;
- Accompagner la valorisation/l'analyse des données de santé ;
- Soutenir l'écosystème et assurer le lien avec la société civile.

Chacun de ces services fera l'objet de conditions d'accès qui seront détaillées dans les chartes Producteur et Utilisateur du Hub. Les engagements de résultat du Hub vis-à-vis des utilisateurs prendraient la forme d'engagement de service, par exemple sur les délais de mise à disposition des données.

En complément de ces services à la demande, le Hub accompagnerait de bout en bout des projets prioritaires, sélectionnés pour leur potentiel de contribution à la recherche, à l'innovation ou encore à la constitution du patrimoine de données partagées via le Hub.

DONNER ACCES AUX DONNEES DE SANTE

Le Hub permettrait aux utilisateurs potentiels de la donnée de santé d'explorer et d'identifier les données disponibles au sein du patrimoine national, d'être accompagnés dans les procédures d'habilitation et d'accéder – sous réserve d'habilitation - à des jeux de données (simples ou appariés) documentés.

Explorer et identifier les données disponibles

« Je suis interne en médecine intéressé par la recherche ou chef de clinique en cours de réflexion sur mon projet scientifique, je souhaite connaitre le contenu précis des bases de données auxquelles je peux avoir accès via le Hub pour accélérer mes recherches. »

Les utilisateurs auraient à leur disposition des outils publics permettant d'explorer le patrimoine de données partagées via le Hub et d'identifier les jeux de données pertinents au regard du cas d'usage considéré.

Ils disposeraient d'informations sur la structure et le contenu des jeux de données disponibles (par exemple les champs disponibles, couverture, profondeur historique, niveau de qualité, fréquence de mise à jour, contraintes de confidentialité, documentation etc.) et auraient la possibilité d'accéder à des échantillons anonymisés et/ou synthétiques (par exemple pour permettre de comprendre la structuration d'une base de données et d'évaluer son potentiel avant de faire une demande d'accès aux données plus complètes). Sur ce dernier point qui pourrait s'avérer complexe, le Hub pourra mobiliser l'écosystème pour accompagner les producteurs dans leurs travaux, en particulier les grands organismes et opérateurs de recherche français (Inria, Inserm, CNRS, CEA).

Le Hub pourra également mettre en relation les utilisateurs qui le souhaitent avec les experts concernés afin de les aider à préciser et mieux qualifier leurs demandes.

Accompagner les procédures d'habilitation et les demandes d'appariement

« Je suis médecin chercheur dans un CHU, je fais partie d'un réseau européen de recherche sur une maladie rare, et j'aimerais accéder aux données d'autres centres hospitaliers pour compléter ma cohorte de patients. »

Les utilisateurs du Hub trouveraient sur une interface publique les conditions et les procédures d'accès et d'usage des données qu'ils souhaitent utiliser. Ils auraient la possibilité de formuler et de documenter leurs demandes d'accès et d'appariement sur l'interface et pourront suivre le statut de celles-ci au moyen d'un « workflow » donnant de la visibilité sur le statut et les étapes en cours.

Le Hub pourrait être en charge de la collecte des demandes d'accès, du suivi et de l'outillage du processus d'habilitation et pourrait fournir une expertise ciblée à la demande pour accompagner les cas les plus complexes (ex : expertises technique, juridique, etc.), en lien avec l'actuel Institut National des Données de Santé qui incarne ce rôle pour le SNDS.

Le Hub devrait prévoir dans ce processus une formation à la sensibilité des données de santé. Des ressources pourront être identifiées (comme des cours en ligne proposés par l'ANSSI) ou d'autres contenus potentiellement proposés par les producteurs. Des foires aux questions pourraient également être prévues et un accent particulier devrait être mis au moment de la signature des conditions générales d'utilisation du

Hub par l'utilisateur, notamment sur la traçabilité et l'audit de ses requêtes et les risques encourus si des comportements abusifs étaient détectés.

Dans le cas le plus simple, l'utilisateur souhaite accéder à un jeu de données présent dans le « catalogue » du Hub, i.e. faisant partie de la bibliothèque de données régulièrement actualisées et considérées comme d'un grand intérêt pour la communauté. Mais l'utilisateur peut aussi vouloir rapprocher des sources qui ne seraient pas présentes dans le Hub et utiliser les services d'appariement et de mise à disposition du Hub. Dans ce contexte, le Hub pourrait mettre à disposition des expertises juridiques, expertises autour de la protection des données personnelles et de la valorisation économique pour accompagner l'utilisateur dans son dialogue avec certains producteurs de données, par exemple les DSI des établissements de santé. Il pourra également s'appuyer sur les expertises de la communauté scientifique (notamment l'Inserm, l'Inria, le CNRS et le CEA) dans le cas où le porteur de projet n'aurait pas une vision précise de la traduction scientifique de la question « métier » qu'il se pose, et/ou des sources de données permettant d'adresser celle-ci. Le recours à ces expertises pourrait être consolidé par la création d'un hub « local » de recherche sur les données de santé regroupant les acteurs des divers organismes et opérateurs de recherche cités ci-dessus.[17]

Mettre à disposition des jeux de données appariés et documentés

« Je suis statisticien à l'ANSM (agence nationale de sécurité du médicament et des produits de santé), j'ai un accès permanent au SNDS et j'ai obtenu une habilitation pour accéder à des données hospitalières sur le circuit du médicament. Je souhaite réaliser un appariement avec le SNDS pour mener une étude de pharmaco-épidémiologie et j'ai besoin d'accéder à ces données et de les croiser. »

Les utilisateurs disposeraient d'un espace personnel leur donnant accès aux jeux de données pour lesquels ils ont obtenu une habilitation via la gouvernance du Hub.

Si un utilisateur a demandé et obtenu le droit de réaliser un appariement, celui-ci est préparé par le Hub dans un espace cloisonné prévu à cet effet puis mis à disposition de l'utilisateur dans son espace personnel.

Les utilisateurs auront également la possibilité d'*uploader* sur leur espace des jeux de données propriétaires qu'ils pourront utiliser pour enrichir les données, dans la limite de ce qui est autorisé par la loi.

[17]La création d'un hub « local » recherche s'inscrit dans une démarche d'approfondissement de l'offre du Health Data Hub vers le monde de la recherche.

SOUTENIR LA COLLECTE ET LA CONSOLIDATION DES DONNEES

Le Hub doit accompagner les producteurs dans leur effort de collecte et de mise en qualité des données. Les données sont souvent silotées et faiblement normalisées. Décloisonner les données et les harmoniser constitue un travail lourd et coûteux, d'ingénierie et parfois de recherche, que le Hub pourrait soutenir, et accompagner sur le plan humain et financier.

Appuyer la définition et l'implémentation de terminologies, formats et standards communs

En tant que carrefour entre producteurs et utilisateurs, le réseau Hub est idéalement placé pour :

- Mobiliser l'écosystème pour participer aux réflexions sur les normes et standards, de préférence internationaux, en lien avec des agences, telles que l'ASIP ou l'ANAP, et des acteurs tels que les collectifs d'industriels de promotions de standards (InteropSanté ou Phast par exemple) ;
- Inscrire ces réflexions dans un contexte international en prenant contact avec des Hubs étrangers qui mèneraient des démarches similaires ;
- Diffuser les initiatives prometteuses des acteurs de l'écosystème et les faire passer à l'échelle.

Ces réflexions pourront alimenter la décision politique d'imposer certains cadres d'interopérabilité, standards sémantiques ou encore terminologies pour des jeux de données ciblés. Le Hub pourrait se faire l'intermédiaire pour passer des messages vers la puissance publique. Le Hub pourrait apporter son soutien auprès des acteurs concernés dans la mise en œuvre opérationnelle de feuilles de route pour l'intégration de ces standards.

Soutenir le financement des démarches de collecte

Extraire les données des logiciels – notamment pour les données issues des soins –, et les mettre dans des formats ou des standards communs (cf supra) est long et coûteux. Il faut créer des connecteurs et mettre en place des entrepôts pour agréger les différents flux et parfois reconstruire les modèles de données, en particulier dans les établissements hospitaliers. Cette étape est un passage obligé pour constituer un patrimoine de données plus large et le mettre à disposition des utilisateurs.

Le Hub pourrait soutenir le financement des initiatives de collecte et de consolidation du patrimoine de données de plusieurs manières :

- Via l'accompagnement ou le financement de projets sélectionnés, entre autres, pour leur contribution à la construction du Hub (patrimoine, outils de collecte mutualisables, etc.) ;
- En se faisant l'intermédiaire avec la puissance publique pour que des financements soient prévus dans le cadre de plans nationaux tels que Hop'EN.

Tous les financements devraient toutefois être adossés au principe de partage des données et à l'utilisation des normes communes identifiées par la communauté.

Soutenir l'investissement des producteurs dans la mise en qualité

Une fois collectées, rassemblées, accessibles, les données sont utilisables mais leur plein potentiel de réutilisations ne peut être atteint qu'en investissant de manière conséquente dans la « qualité » de ces données, par exemple, par l'annotation, le redressement, le retour au dossier, etc...

Le Hub doit soutenir l'investissement des producteurs dans la qualité, il pourrait :

- Mobiliser l'écosystème[18]pour participer aux réflexions sur les normes et standards en matière de qualité, en s'inscrivant dans un contexte international ;
- Mobiliser l'écosystème, producteurs, utilisateurs, recherche (notamment l'Inserm, l'Inria, le CNRS et le CEA) pour développer des outils pouvant automatiser la mise en qualité et diffuser ces outils ;
- Mobiliser l'écosystème et la puissance publique pour participer, pour les bases de recherche, à la mise en œuvre opérationnelle du Plan National pour la Science Ouverte. Ce plan prévoit de considérer les bases de données comme un *asset* scientifique devant être cité au même titre que les publications scientifiques, et par conséquent, pouvant contribuer au cumul de points (par exemple sigaps) des chercheurs/promoteurs de base qui accèdent à des financements et évolutions de carrière sur la base de ces scores ;
- Rétribuer les efforts de mise en qualité à travers le modèle économique du Hub.

ACCOMPAGNER LA VALORISATION DES DONNEES DE SANTE

Le Hub proposerait des capacités technologiques et des compétences à la demande pour accélérer les projets de valorisation. Il donnerait également accès à un catalogue d'algorithmes et de méthodologies construites par la communauté.

Mettre à disposition des capacités technologiques à la demande

« Je suis une start-up du monde digital ou un chercheur souhaitant mobiliser des approches de machine learning, je souhaite tester des méthodes de sélection de variables pour mettre en évidence l'impact de la prise concomitante d'un médicament sur la réponse à un traitement dans le cadre de la prise en charge d'un cancer. Pour cela, j'ai besoin de ressources de calcul et de logiciels de traitement des données adaptés, tel que python par exemple.»

Les utilisateurs du Hub auraient accès dans un environnement sécurisé à des capacités technologiques leur permettant d'exploiter les données mises à disposition par le Hub :

- Capacité de stockage, ressources de calcul (CPU, GPU) ;

[18] Mobilisation des Centres d'Investigation Clinique (CIC) et de F-CRIN notamment, qui ont vocation à faciliter et à promouvoir la mise en œuvre de méthodologies, de la standardisation et de l'interopérabilité des données de recherche clinique.

- Langages et outils de requête et d'analyse statistique (par exemple : Python, R, Jupyter, RStudio, Spark, SAS) ;
- Outils de visualisation (par exemple : R Shiny, Tableau) ;
- Librairies pour l'apprentissage automatique (par exemple : Tensorflow, Keras) ;
- Outils de développement d'applications (SDK)

Mettre à disposition des briques documentaires ou techniques clés en main

« Je suis statisticien à la DREES (direction de la recherche, des études, de l'évaluation et des statistiques du ministère de la santé), je souhaite évaluer l'impact de la tarification à l'activité sur le taux de réadmission à 30 jours qui peut être, selon les cas, interprété comme un marqueur de non qualité des soins. Pour cela, j'ai besoin des données du PMSI et de construire une vue "parcours" pour identifier les admissions successives pour un même patient. J'ai besoin également de distinguer les cas par grandes familles de pathologies et de traiter différemment les réadmissions programmées ou non programmées. Je suis donc preneur d'une vue "parcours" de ces données et d'algorithmes d'inférence des pathologies par patient développés par la Cnam. Je construis un indicateur de taux de réadmission à 30 jours, je partage ma requête à la communauté, je notifie la Cnam de ma publication. Je réponds aux questions sur la FAQ (foire aux questions) pour lesquelles je suis compétent. »

Les utilisateurs auraient accès à des librairies de codes et d'algorithmes de traitement documentés issus du monde de *l'open source* ou développés par la communauté. A titre illustratif et de manière non exhaustive : opérations standards de préparation de données, modélisations et algorithmes de *machine learning*, traitement du langage naturel et sémantique, traitement d'images, de sons, d'autres signaux physiologiques numérisés.

En complément, ces outils seraient accompagnés d'une documentation collaborative avec des outils de FAQ dynamiques permettant de solliciter les contributeurs et de partager les meilleures pratiques.

Accompagner à la demande des projets de valorisation de données

« Je suis chercheur en data science travaillant sur un projet de prédiction du rejet de greffe en collaboration avec une association de patients, je souhaite confronter les résultats de mes recherches auprès d'un panel d'experts de la discipline avant publication. »

Le Hub tiendrait à disposition (via une mise en relation des acteurs de l'écosystème et/ou de capacités propres) des expertises ciblées pour adresser une problématique ponctuelle :

- Expertise juridique et de valorisation économique ;
- Expertises médicales ;
- Expertises liées à la valorisation de la donnée (data science, data engineering, architecture, …).

Pour certains projets sélectionnés, le Hub pourrait mettre à disposition tout ou partie d'une équipe pluridisciplinaire pour accompagner le projet dans son ensemble, depuis son cadrage initial jusqu'à son développement.

Au-delà des équipes que le Hub détiendrait en propre, il s'appuierait sur le tissu d'expertise de l'écosystème tel que l'Inserm qui a développé des services d'appui aux chercheurs afin de faciliter l'accès aux données et leurs appariements[19], ou encore le réseau des Centres de Données Cliniques du Grand Ouest qui se déploie dans des CHU ou CLCC (centre de lutte contre le cancer).

SOUTENIR L'ECOSYSTEME ET ASSURER LE LIEN AVEC LA SOCIETE CIVILE

Le Hub pourrait avoir vocation à animer l'ensemble des parties prenantes en encourageant le partage et l'ouverture au sein de la communauté, en soutenant et accélérant les initiatives innovantes et en renforçant la transparence vers la société civile et les citoyens.

Permettre le partage et le développement d'une culture de la donnée au sein de la communauté

« Je suis un chercheur, une entreprise ou une startup, je cherche à développer un outil de traitement automatique du langage permettant le codage automatique des diagnostics à partir des comptes rendus médicaux. Pour réaliser mes travaux, je souhaiterais accéder à des corpus de comptes rendus médicaux désidentifiés et être mis en relation avec les collèges des professionnels de santé ou des experts en informatique médicale pour préciser mon projet. Une fois ma solution développée, je souhaite la mettre à disposition des utilisateurs du Hub de manière payante si je suis un acteur privé ou gratuite. »

Les utilisateurs du Hub auraient accès à un catalogue de formations thématiques liées à la valorisation de la donnée. Certaines de ces formations pourront être animées et proposées par les producteurs de données partenaires du Hub.

En complément, le Hub pourra également collaborer avec les instituts 3IA, les organismes et opérateurs de la recherche tels que l'Inria, l'Inserm, le CNRS et le CEA, et les doyens de facultés dans toutes leurs démarches visant à former les populations médicales sur les enjeux, les méthodologies et les outils liés à l'utilisation des données de santé.

Pour augmenter l'efficacité des travaux de recherche et d'innovation, le Hub pourra également renforcer l'esprit de collaboration, de partage et d'ouverture au sein de la communauté au travers d'outils collaboratifs :

- Un réseau social, qui permettrait d'identifier les acteurs (expertises, réalisations, publications…), de communiquer sur des sujets spécifiques et de mettre en valeur les utilisateurs du Hub et leurs réalisations ;
- Une plateforme collaborative offrant un accès ouvert et gratuit à des études, des publications, des résultats de recherches, des algorithmes etc. Cette plateforme s'inscrirait dans le mouvement

[19] L'Inserm a été identifié comme « coordinateur national d'infrastructures de recherche utilisant des données de santé » par le décret du 26 décembre 2016 pour réaliser des extractions et la mise à disposition des données pour les traitements mis en œuvre à des fins de recherche, d'étude ou d'évaluation et dans le cadre d'une convention conclue avec la CNAM (code de la santé publique, art R. 1461-3).

« *open science* » pour construire une connaissance collective et s'inspirera des initiatives existantes, telles que *OpenFDA* ou *Open Ontology Repository* ;
- Une plateforme d'applications sur laquelle les acteurs de l'écosystème pourraient publier les applications qu'ils auraient développées autour des données partagées via le Hub, définir les conditions d'usage (gratuit, *freemium*, etc.) et collecter les retours de la communauté.

Soutenir et accélérer les initiatives innovantes

« Je suis un étudiant, je recherche un sujet de thèse dans le domaine de la santé et de la data science. Je souhaiterai participer à un challenge coorganisé par le Hub pour rencontrer la communauté et identifier des sujets de thèse intéressants. »

Pour accélérer la recherche et l'innovation sur les enjeux stratégiques de santé publique, le Hub appuierait les porteurs de projets. A cet effet, il pourrait :

- Accompagner l'organisation de *challenges* et de *hackathons* au travers d'un soutien financier ou humain, de la mise en réseau avec des acteurs pertinents ou d'un accès facilité aux données. Le Hub pourrait aussi être à l'initiative d'un *challenge* international, à l'instar du *Solve's Global Challenge* organisé par le Massachussetts Institute of Technology (MIT), pour renforcer son *leadership* en matière d'innovation dans la santé ;
- Lancer, avec le soutien de la puissance publique, des appels à projets sur des thématiques clés autour de l'IA et de la santé, en vue d'attribuer des subventions publiques à des projets à fort potentiel d'innovation et de transformation. L'identification d'un nombre limité (trois ou quatre apparaît comme un maximum) de priorités nationales fortes avec un horizon de réalisation relativement court (trois à cinq ans) semble pertinent ;
- Accompagner les acteurs de l'écosystème qui souhaiteraient répondre à des appels à projets européens et internationaux dont les procédures sont souvent complexes. Le Hub pourra faire appel au réseau et à ses équipes internes pour appuyer les porteurs de projets dans leurs démarches.

Assurer la transparence vers la société civile et les citoyens

« Je suis un patient, je souhaite partager mes données pour accélérer la recherche dans un champ thérapeutique donné, le Hub me permet d'alimenter une base aux finalités précisées, dans des conditions de sécurité adéquates et avec la possibilité de consulter l'utilisation qui en est faite »

Pour garantir le lien de confiance et la transparence, le Hub communiquerait sur son portail autour de la nature des données collectées et de la façon dont celles-ci sont utilisées. Avec la rédaction d'une charte citoyen, il s'engagera vis-à-vis de la société civile sur une utilisation des données personnelles dans le respect de la réglementation et des principes éthiques.

Le Hub, en lien avec les associations de patients, aura aussi pour mission de sensibiliser les citoyens à l'importance du partage pour alimenter des études et les travaux de recherche pouvant contribuer à l'amélioration du système de santé, de la médecine, à la qualité des soins, la prévention, l'innovation, etc. Pour cela, il lancera des actions de communication avec des associations de patients, qui auront également la possibilité de partager les données qu'elles récoltent via le Hub. A noter que le patient peut être l'acteur

de la réutilisation de ses données, via les associations de patients qui sont parfois autant utilisatrices que productrices de données. Il pourrait à ce titre bénéficier de l'accompagnement du Hub pour leur valorisation analytique.

Au-delà des associations de patients, le Hub pourra s'appuyer sur d'autres acteurs pour sensibiliser, communiquer, ou animer l'écosystème. On peut citer le Conseil National de l'Ordre des Médecins, Epidemium, Ethik IA ou les Instituts interdisciplinaires d'intelligence artificielle, etc.

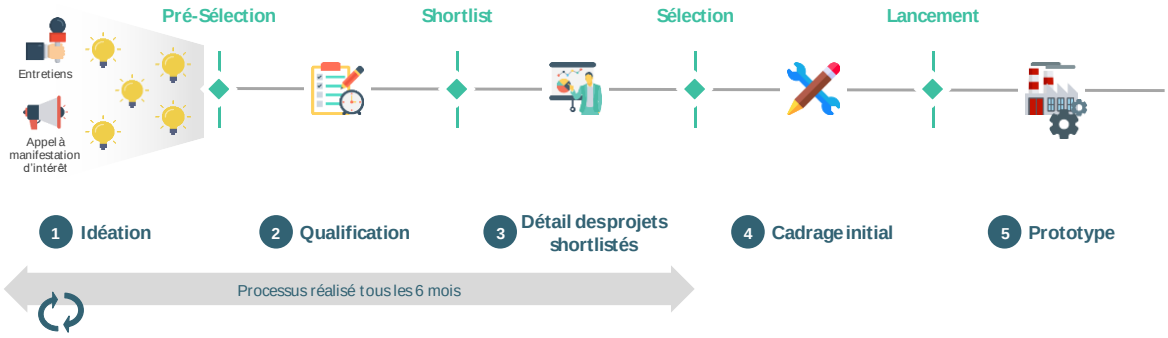
La mission recommande d'engager dès le début du projet une réflexion sur l'évolution de ce portail vers un accès nominatif où un citoyen pourrait exercer un consentement explicite ou un droit d'opposition, bien qu'elle ait conscience du besoin d'étudier précisément l'impact en termes de sécurité informatique et organisationnelle de détenir le lien entre les identités des patients et les données partagées dans le Hub. Certains projets de mise à disposition des données à des fins de recherche prévoient ce type de dispositif, c'est notamment le cas du projet Québécois PARS qui décrit différentes catégories de données pour lesquelles le consentement peut être demandé explicitement (dans l'absolu ou sous un certain délai) ou considéré au contraire comme donné par défaut sauf expression d'un droit d'opposition (dans l'absolu ou sous un certain délai). C'est aussi le cas en Corée où le *national disease registry* donne accès au patient à toutes ses données via une plateforme nationale où il peut exercer son consentement.

L'articulation de ce portail avec l'espace numérique de santé (préconisé dans le cadre du chantier « virage numérique » de la stratégie de la transformation du système de santé pilotée par Dominique Pon et Annelore Coury) pour tous les citoyens dès leur naissance, contenant entre autres les données médicales issues du DMP (dossier médical partagé) mais prévoyant également la possibilité de verser ces données personnelles et d'accéder à des applications certifiées.

ACCOMPAGNER DES PROJETS SELECTIONNES DE BOUT EN BOUT

Au-delà des besoins exprimés à la demande par les acteurs porteurs de projet ou utilisant les données du catalogue de manière opérationnelle, des séries de projets prioritaires pourraient être régulièrement sélectionnées et accompagnées dans leur réalisation de bout en bout. Ces projets seraient retenus selon des critères prédéfinis, dans le but de dynamiser les usages, et de tester et enrichir les fonctionnalités du Hub de manière itérative.

Dès son lancement, le Hub accompagnerait chaque semestre trois à cinq projets prioritaires, par exemple, pour lesquels les données seraient mises à disposition et appariées. Un financement et/ou des capacités (compétences, plateforme, locaux) seraient apportés pour la réalisation de bout en bout de ces projets. Les trois à cinq premiers projets à être lancés constitueront également des « pilotes » qui permettront de constituer et d'éprouver progressivement les outils et les modes de fonctionnement du Hub. Leur sélection pourrait se faire en lien avec les premiers partenaires du Hub.



En régime pérenne, l'identification des idées et opportunités pourrait s'opérer au moyen d'appels à manifestation d'intérêt (AMI) récurrents. A l'issue d'une pré-sélection, les projets pourraient ainsi être retenus sur la base d'une appréciation de différents critères, comme par exemple :

- La valeur intrinsèque - sur le plan de la recherche, du développement du tissu économique, de la politique publique de santé, etc.
- L'accessibilité et la qualité des données requises, la complexité technologique, le niveau de mobilisation des acteurs, etc.
- La valeur contributive à la constitution du Hub : les données annotées, qualifiées et partagées, les algorithmes et techniques mises en commun, etc.

Les hubs locaux ont vocation à élargir de manière multi partenariales et multi organismes cette démarche d'accompagnement de projet.

Les services du Hub seraient proposés sur l'ensemble du territoire via un Hub central appuyé d'un réseau de Hub « locaux ». Ce réseau, piloté par le Hub central, répondrait à une logique de regroupement géographique, qui permettrait d'offrir une meilleure efficacité opérationnelle en désengorgeant la structure centrale, tout en facilitant l'accompagnement des producteurs et utilisateurs des données.

ORGANISATION GENERALE EN RESEAU

Hub central ou local, intégrant une gouvernance, des compétences et un accès à la plateforme technologique

Expertise (médicale, juridique, données…)

Utilisateurs du Hub

Producteurs de données

Les Hubs locaux assureraient le lien avec les producteurs de données locaux. Ils réaliseraient à ce titre auprès d'eux le référencement des données au catalogue national et soutiendraient les producteurs dans la collecte et mise en qualité des données produites. Les demandes d'accès aux sources « locales » seraient instruites par un comité éthique et scientifique (CES) local, conformément au paragraphe 4.2 Gouvernance de la donnée, et qui associerait les producteurs locaux. Afin de fluidifier l'accès aux données, les Hub locaux seraient à ce titre compétents pour administrer les accès au patrimoine local de données.

L'ensemble du réseau Hub proposerait une offre de service standardisée, présentant les mêmes engagements de qualité de service (délais, qualité, etc.) vis-à-vis des utilisateurs. Chaque Hub local serait à ce titre compétent et dimensionné pour accompagner le processus d'habilitation et les opérations de préparation des jeux de données et d'aide à leur valorisation, et disposerait à cette fin d'un réservoir de compétences en propre. Il pourrait aussi s'appuyer sur des compétences présentes au sein de l'écosystème local, ou encore, à la demande, associer les producteurs pour une prestation d'expertise sur leurs données.

	Demande d'accès à des sources administrées par le hub local uniquement	Autres demandes (par ex. plusieurs sources)
Guichet unique et accompagnement de la demande	Hub local	Hub central en lien avec l'équipe locale
Autorité compétente pour arbitrer sur la demande	CES local[20]	CES central
Equipe en charge de préparer et mettre à disposition les jeux de données	Hub local	Hub central en lien avec l'équipe locale
Equipe en charge d'accompagner la valorisation	Hub local	Hub central en lien avec l'équipe locale

Les hubs locaux auraient également pour mission de mettre en réseau les compétences et animeraient localement l'écosystème local via des sessions de rencontres, la création des conditions d'échange, la mise en relation d'acteurs travaillant sur la même thématique, etc.

En miroir du Hub central, les Hubs pourraient lancer des appels à projets locaux et les faire passer au travers d'un processus récurrent de sélection selon les mêmes critères que ceux observés par le Hub central. Les projets prioritaires sélectionnés, par exemple au rythme d'un projet tous les six mois, feraient l'objet d'un

[20] Le comité éthique et scientifique ne se substitue pas à la décision de la CNIL, sauf mise en place d'une méthodologie de référence

accompagnement ciblé sur des expertises manquantes ou d'un accompagnement de bout-en-bout sur toute la réalisation du projet.

Afin d'éviter une dispersion des moyens, de veiller au haut niveau de sécurité requis pour la circulation et la valorisation des données de santé, et de fluidifier les interactions entre les Hubs locaux et le Hub central, l'ensemble des entités du réseau partageraient une plateforme technologique commune.

L'organisation en réseau dans son ensemble serait conçue pour trouver le meilleur équilibre entre l'enjeu de proximité des producteurs et des utilisateurs et la nécessaire non dispersion des moyens. La création d'un nouveau Hub local devrait notamment satisfaire aux prérequis suivants :

- S'engager à respecter au niveau du Hub local les engagements du réseau Hub (y compris les chartes producteurs, utilisateurs et citoyens) et les méthodologies associées ;
- Disposer d'une masse critique et couvrir toutes les compétences requises ;
- Disposer d'un portefeuille suffisant d'utilisateurs potentiels et de producteurs de sources distinctes.

ACTIVITES ET COMPETENCES

Pour délivrer l'ensemble des services proposés par le Hub, les équipes du Hub central et des Hub locaux devraient être dotés des compétences permettant d'assurer les grandes activités suivantes :

- La collecte et la gestion des données, par des *data managers, data engineers,* architectes bases de données, couvrant le co-développement des connecteurs avec les acteurs producteurs, l'intégration, la structuration et la documentation des données (dans un souci d'harmonisation des standards de qualité) ; la mise en application des règles relatives à l'accès et l'usage de la donnée (incluant l'audit) et la réalisation des appariements ;

- L'accompagnement dans l'utilisation des données, en tenant à disposition - à la demande ou dans le cadre de l'accompagnement de projets prioritaires - des compétences liées à la valorisation des données (*data scientists, data engineers, scrum masters, UX designers,* développeurs), des expertises liées à l'informatique médicale ou plus généralement aux données de santé ;

- La gestion de la plateforme technologique, assurée par des profils architectes et administrateurs IT, des développeurs et des spécialistes SSI couvrant la conception et déploiement du socle technologique, l'administration et la maintenance de la plateforme, et la mise en sécurité des systèmes d'information ;

- L'animation de l'écosystème, et notamment l'animation de la gouvernance locale, l'animation scientifique et du réseau d'experts, la communication et l'organisation d'événements avec l'écosystème ;

- La protection des données personnelles et l'accompagnement des procédures d'habilitation, assurée par un *Data Protection Officer* (DPO) appuyé par des compétences juridiques spécialisées dans les données de santé.

GOUVERNANCE DU RESEAU HUB

La gouvernance du réseau Hub devra être stabilisée dans des travaux de concertation à venir. La mission propose en première hypothèse une **gouvernance stratégique** composée :

- D'un **conseil stratégique,** réunissant à échéance semestrielle des représentants de l'Etat et des représentants des utilisateurs et des producteurs. Il serait compétent pour les questions relevant de la stratégie générale et du budget, et définirait les grandes orientations du Hub ;
- D'un **comité scientifique,** s'appuyant sur des personnalités reconnues, notamment dans les domaines de l'anonymisation, du traitement et de l'analyse des données, de la médecine, et de l'évolution des technologies. Ce comité sera chargé de conseiller le Hub autour des perspectives de long terme et des orientations et priorités à donner ;
- Un **comité d'administration** resserré, se réunissant entre 4 et 6 fois par an pour suivre le budget et assurer le pilotage des activités.

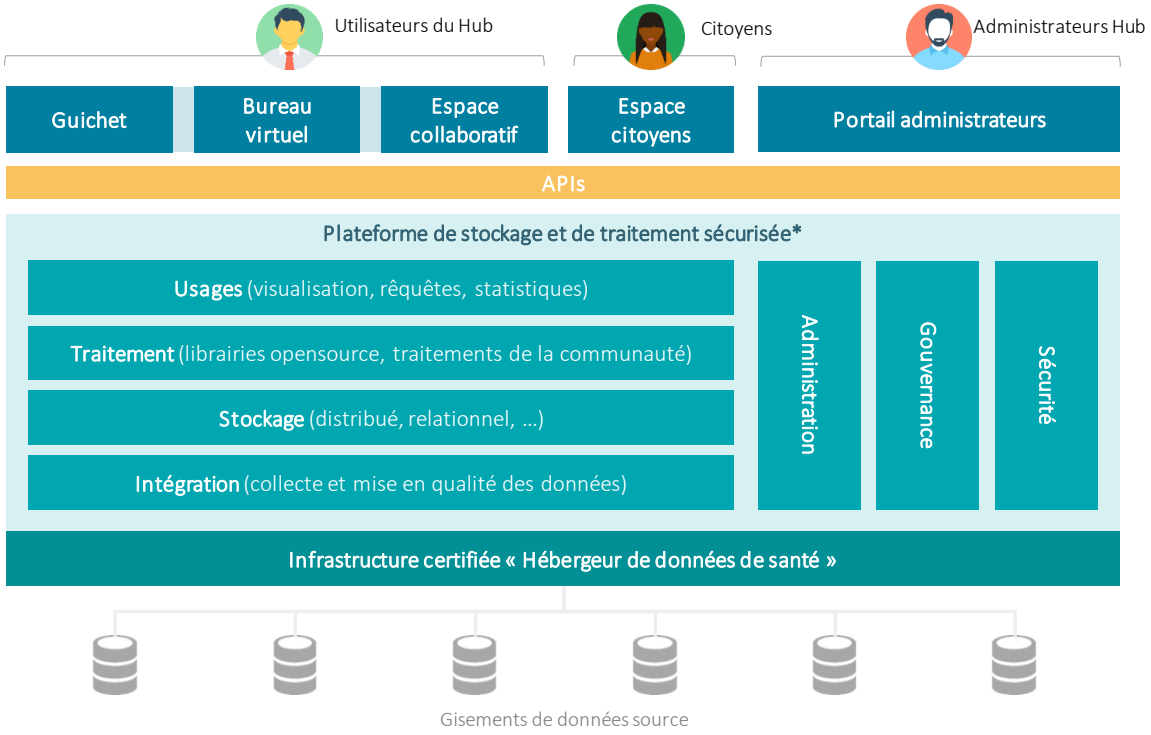
La **gouvernance opérationnelle** s'appuierait sur :

- Le comité **« donnée »,** réunirait des représentants des producteurs de données et des utilisateurs de données du réseau Hub, et des instances en charge de la protection des droits individuels. Il définirait les critères devant être vérifiés par les jeux de données en lien avec les retours utilisateur, les évolutions du catalogue, proposerait des conditions de valorisation économique et serait en charge du suivi de la conformité des usages ;
- Le comité **« technologies »** serait garant de l'alignement avec les besoins des utilisateurs des solutions d'architecture technique et définirait la feuille de route technologique. Il réunirait un responsable technologique, des experts techniques des principales bases du Hub et des utilisateurs de l'infrastructure. Un accent serait notamment mis sur les questions de sécurité ;
- Le comité **« projets »** assurerait le suivi des projets prioritaires : il lancerait les appels à manifestation d'intérêt (AMI), instruirait les dossiers, sélectionnerait les projets, et assurerait le suivi de leur réalisation.

Des points de pilotage réguliers ayant lieu 3 fois par an, en présence du directeur du Hub et des responsables de comités, se tiendront par ailleurs pour effectuer des arbitrages sur les questions d'intégration de nouvelles bases de données, les candidatures d'équipes pour constituer des Hubs locaux, les évolutions du socle technologique etc. sur la base des travaux d'instruction menés par les comités.

En complément, un groupe utilisateur associerait une fois par an des représentants du monde scientifique et de la recherche, des acteurs industriels et privés, et des agences sanitaires et régaliennes en charge des questions médico-économiques. Il représentera la voix des utilisateurs du Hub et constituera un canal de remontée de suggestions d'amélioration, besoins et avis.

PERIMETRE FONCTIONNEL



*pouvant être instanciée et administrée localement par une antenne du Hub

En appui de son offre de service, le Hub disposerait d'un environnement technologique à l'état de l'art permettant :

- De collecter, de stocker, et de traiter les données provenant des différentes sources dans un environnement sécurisé respectant à la fois les obligations relatives à l'hébergement des données de santé et le référentiel de sécurité du SNDS et permettant de garantir la bonne application des conditions de gouvernance des données prévues au titre du cadre législatif et des chartes du Hub, ainsi que la mise à disposition aux administrateurs centraux ou locaux d'un ensemble de fonctionnalités pour administrer le patrimoine de données **(« Plateforme de stockage et de traitement sécurisée ») ;**

- D'offrir aux utilisateurs :
  - Un portail public offrant la possibilité de consulter les informations sur les données disponibles sur le Hub, de formuler des demandes d'accès aux données et aux services du Hub et de suivre le statut de leurs demandes d'habilitation sur un espace personnel **(« Guichet ») ;**

- Un environnement sécurisé d'accès aux données pour lesquelles ils sont habilités accompagné d'un ensemble d'outils de requête et d'analyse statistique (type R, Python), de visualisation (type RShiny, Tableau), de développement (SDK, Git), de librairies data & IA issues de *l'open source* pour faciliter leur traitement **(« Bureau virtuel ») ;**
- Un espace collaboratif donnant accès à des informations sur le Hub, à la présentation des projets et des réalisations, et à des ressources documentaires partagées de type wiki, à un réseau social mettant en relation des acteurs de la communauté et un accès aux applications réalisées par la communauté **(« Espace Collaboratif ») ;**
- Des APIs offrant une façade d'échange permettant d'accéder en fonction des droits attribués aux ressources de la plateforme de stockage et de traitement **(« APIs ») ;**

- D'offrir aux citoyens la possibilité de suivre l'utilisation de leurs données personnelles, détaillé au paragraphe 4.3.4 (**« Espace Citoyen »**) ;

- D'offrir aux administrateurs de l'équipe centrale et des Hub locaux un portail permettant d'administrer et de superviser le système, de gérer les utilisateurs et les droits d'accès, de gérer les socles applicatifs et les APIs, et de définir et suivre les requêtes et usages facturables (**« Portail Administrateur »**).

PRINCIPES D'ARCHITECTURE

L'architecture de la plateforme devrait être conçue en tenant compte de trois enjeux structurants :

La protection et la maitrise des données de santé

La plateforme technologique et les processus associés devront respecter les exigences liées à l'hébergement de données de santé et respecter le référentiel de sécurité du SNDS (arrêté du 22 mars 2017 relatif au référentiel de sécurité applicable au Système national des données de santé) :

- A cette fin, des fonctionnalités assurant notamment une double authentification, la traçabilité, l'intégrité, la confidentialité et la disponibilité devront être implémentées. L'accès aux services se fera par un bureau distant à affichage déporté ou par APIs sécurisées. Certains types de données particulièrement sensibles feront l'objet d'une protection renforcée, cela pourra être le cas des données génomiques par exemple. Parmi ces principes, la traçabilité en particulier peut susciter une crainte vis-à-vis de la protection de la propriété intellectuelle et du secret industriel sur des algorithmes innovants. Elle devra donc s'accompagner de garanties techniques, organisationnelles et/ou juridiques offrant des gages aux acteurs privés ou aux chercheurs utilisant la plateforme.
- Les fonctionnalités d'administration du patrimoine de données devront en particulier permettre une traçabilité complète des accès et des opérations réalisées.

Si le Hub était désigné opérateur de service essentiel, il devrait également respecter les règles de sécurité associées et formulées dans l'arrêté du 14 septembre 2018 fixant les règles de sécurité et les délais mentionnés à l'article 10 du décret n° 2018-384 du 23 mai 2018 relatif à la sécurité des réseaux et systèmes d'information des opérateurs de services essentiels et des fournisseurs de service numérique.

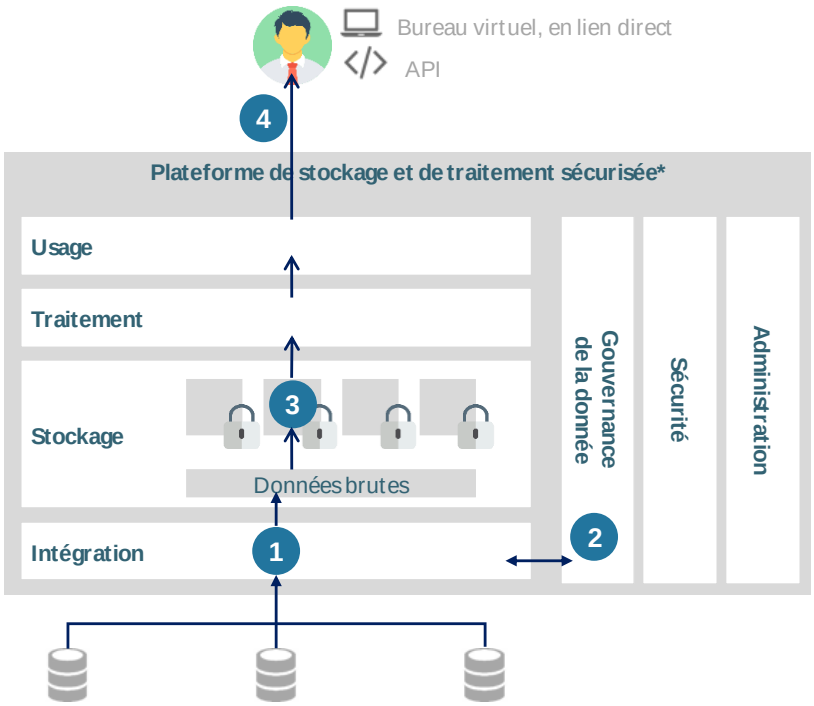
L'évolutivité et le maintien à l'état de l'art

- La plateforme est amenée à intégrer un patrimoine de données de plus en plus large et devra pouvoir faire face à des volumes et des besoins croissants. Elle doit aussi être à même de répondre dans la durée aux besoins les plus avancés en termes d'IA sans jamais être désuète au risque de ne pas être utilisée.
- L'architecture devra donc répondre à des besoins de scalabilité (volume, puissance de calcul, CPU, GPU), d'élasticité (notamment pour s'adapter de manière dynamique aux demandes contractualisées au moyen du catalogue de services), et de modularité (pour intégrer de manière dynamique de nouveaux composants et/ou en remplacer certains tombés en obsolescence).

L'ouverture

- La plateforme doit pouvoir faciliter la collaboration et la capitalisation sur les réalisations de la communauté. Le socle technologique privilégiera les solutions *open source*. Les librairies *open sources* associées à la *data science* (tensorflow, pandas, scikit, …) seront disponibles et tenues à jour par les administrateurs. Un outil de gestion de versions type « Git » sera proposé pour partager les algorithmes.
- La plateforme pourra également communiquer et s'interfacer avec des outils tiers via une communication au moyen des APIs, qui donneront accès aux ressources des plateformes, de gouvernance des données, de sécurité, dans le respect des réglementations et conditions d'accès et d'usage de la donnée.

CYCLE DE VIE DE LA DONNEE



Toutes les données financées par la solidarité nationale ne seront pas centralisées. Certaines pourront l'être : enrichissements en routine du SNDS avec des données complémentaires, appariements pérennes de plusieurs bases de données supposant une alimentation régulière, sélection de bases de données jugées d'intérêt pour la communauté des utilisateurs. Le chargement s'opèrerait alors selon un protocole convenu avec les producteurs à une fréquence régulière (batch de mise à jour) [1]. Le catalogue permettrait d'explorer l'intégralité du patrimoine de données partagées via le Hub mais d'autres sources pourront être mobilisées « à la demande », de manière ponctuelle, en lien avec les producteurs de données pour constituer des espaces projets temporaires.

Des fonctionnalités seront mises à disposition des administrateurs centraux et locaux de la plateforme pour administrer le patrimoine de données. Elles couvriront le cycle de vie de la donnée, tout au long de la chaîne de génération, d'alimentation et de traitement : Planification de l'alimentation, des traitements, du stockage, de l'archivage et durée de rétention ; Gestion du catalogue de données et des métadonnées, définition des règles métier applicables pour chaque attribut : accès, usage, qualité ; Traçabilité au long du cycle de vie : d'où vient la donnée, par où est-elle elle passée, qui y a eu accès…, et permettra audits et alertes vis-à-vis de la conformité aux politiques de sécurité en place ; Gestion des utilisateurs et des règles de sécurité [2].

Lorsqu'un utilisateur obtiendra l'accès à un ou des jeux de données, un espace cloisonné logiquement et/ou physiquement se limitant à son besoin d'en connaître est constitué sur la plateforme. Les appariements mobilisant des variables identifiantes seront réalisés sur un espace distinct et cloisonné [3].

L'utilisateur aura accès aux données et aux outils de traitement associés sur son bureau virtuel (affichage déporté sur un poste dédié, transitant par un lien direct) ou par API dans la limite des ressources d'infrastructure et applicatives (CPU, GPU, stockage, outils) qui lui auront été allouées [4].

TRAJECTOIRE

En cible, la plateforme technologique devra couvrir le spectre le plus large de besoins de l'écosystème, que ce soit pour accompagner le travail quotidien de la statistique publique, des opérateurs et agences sanitaires comme pour traiter des cas d'usages issus de la recherche et de l'industrie mobilisant des techniques de fouille de données ou d'intelligence artificielle.

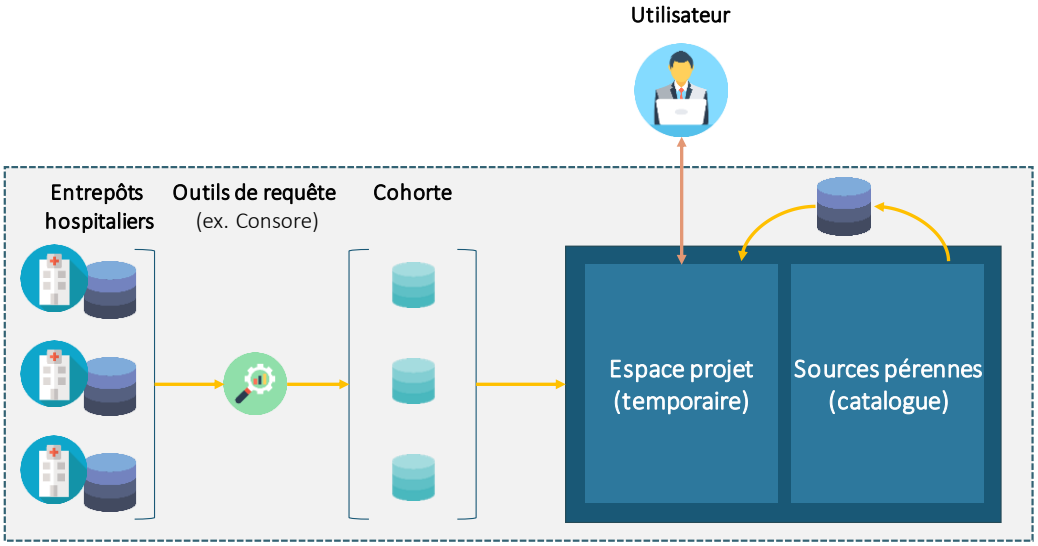
La logique de mutualisation des moyens à un niveau national et territorial vise à offrir à l'écosystème des producteurs des outils communs de mise à disposition des données à destination d'une communauté d'utilisateurs qui va s'accroître. Ces outils présenteront un haut niveau de transparence et une traçabilité complète, ce qui représente un important coût qu'il semble raisonnable de mutualiser. Ceci ne remet pas en cause la mise en place et le développement de systèmes d'information en propre pour agréger les données, les structurer et les mettre en qualité avant de les référencer, le cas échéant, au catalogue du Hub. A cet égard, on peut citer – entre autres – les entrepôts de l'AP-HP, des Hospices Civils de Lyon et ceux du réseau des centres de données cliniques du Grand Ouest. Les établissements devront, autant que possible, se doter de ressources humaines en propre pour réaliser ces activités, grâce notamment à la valorisation économique de la mise à disposition des données ou à des financements issus de programmes nationaux tels que Hop'EN. Les compétences locales seront également celles qui pourront contribuer à l'offre de service des Hubs pour accompagner si nécessaire les porteurs de projet dans leur besoin d'expertise des données. Les Hubs centraux et locaux pourront également consolider ces ressources grâce à leur réservoir de compétences pour des opérations ciblées.

La trajectoire de constitution sera pragmatique et progressive. Des versions successives seront ainsi étalées dans le temps pour enrichir et déployer progressivement les capacités demandées. Les premières étapes se focaliseront à ce titre sur l'analyse de données « à froid » (i.e. les sources de données seront rafraichies au moyen de « batch » réguliers mais pas en temps réel), qui correspond à la grande majorité des besoins exprimés lors de la consultation des parties prenantes. Ceci n'exclut pas la mise en place dans un second temps de capacité d'intégration et de traitement en temps réel, voir l'hébergement d'applications conteneurisées amenées à fonctionner en production sur l'infrastructure du Hub, qui seraient de nature à répondre à des besoins plus avancés.

De la même manière, l'architecture privilégiera dans un premier temps une solution centralisée. Pour le calcul sur d'importants jeux de données homogènes, la piste du calcul réparti au niveau des systèmes d'information de chaque établissement de santé, par exemple, nécessiterait pour être opérante des développements et des expérimentations, une mise à jour et un alignement long et coûteux de l'ensemble des systèmes sources et limiterait les possibilités d'usages aux algorithmes conçus à cet effet. En outre, cette approche ne permettrait pas l'appariement de sources complémentaires, or beaucoup de cas d'usage découleront de la capacité à rapprocher différents jeux de données. Evidemment, à long terme, le déploiement du DMP et son enrichissement en fera le candidat naturel pour abonder la plateforme du Hub, en ce sens où il apparie à la source les différentes données d'un patient donné, dans un cadre d'intéropérabilité bien déterminé. Les projets Hub et DMP doivent donc être étroitement liés pour que cette convergence soit opérationnelle et efficace le moment venu, en particulier les évolutions du DMP doivent prendre en compte cette perspective. Certains pays dont une partie des données sont chaînées par

construction (notamment parce qu'ils disposent de dossier patient de longue date, de grands registres et d'identifiants nationaux communs ou encore parce que les mêmes acteurs coordonnent différentes activités telles que l'assurance, l'hôpital, la télémédecine) prévoient leur mise à disposition à des communautés d'utilisateurs ciblées dans une infrastructure de calcul pour une réutilisation à des fins de recherche. C'est notamment le cas, par exemple, du Danemark où 30 registres nationaux ont été rassemblés au sein du Statens Serum Institut en mars 2012, de la Corée avec son Healthcare Big Data Hub ou du Clalit Research Institute en Israël.

La mutualisation de l'infrastructure n'est pas incompatible avec les outils de faisabilité tels que Consore (Unicancer) en oncologie, qui se développent et se déploient dans certains CLCC. Ces outils, aux mains du producteur, permettent à un clinicien d'identifier les patients qui pourraient constituer une cohorte dans le cadre d'un projet de recherche. De manière distincte, chaque CLCC (dans notre exemple) dispose de peu d'observations. Le clinicien qui souhaiterait donc agréger les données issues de plusieurs CLCC pour constituer une cohorte significative, et enrichir éventuellement ces données avec des données médico-administratives issues du SNDS, pourrait faire appel au Hub comme tiers de confiance. Ce dernier pourrait organiser la collecte des observations sélectionnées, les agréger et les mettre à disposition du porteur de projet dans son espace. Le soutien à des projets tels que Consore peut être une manière de favoriser la constitution de bases de données hospitalières pour des projets de recherche et de tester la faisabilité et la pertinence d'intégrer ces données au catalogue à moyen terme.



MODELE JURIDIQUE

Les missions du Hub seront hybrides **ou « à double visage »**[21] pour reprendre une terminologie habituelle. Le Hub cumulera en effet des missions de service public administratif et des missions industrielles et commerciales. Les **missions qui relèvent d'un service public administratif** sont globalement celles qui se rattachent à la mission de mise à disposition des données du patrimoine du Hub. Pour assurer la lisibilité de l'accès aux données pour les utilisateurs, l'articulation du Hub avec l'INDS devra être clarifiée. L'INDS est l'Institut National des Données de Santé, créé en 2016 par la loi de modernisation du système de santé et est chargé, entre autres, d'accompagner les utilisateurs dans leur processus d'accès aux données (cf annexe). D'autres activités se situent à l'inverse dans un cadre concurrentiel et constituent des activités marchandes au regard du droit communautaire[22], comme par exemple : la fourniture pour compte de tiers d'espaces d'hébergement des données ; la fourniture de technologie d'appariement des données ; la mise à disposition de services pour l'ingénierie et l'analyse des données…

Pour disposer d'une forme de permanence de moyens et d'une vraie capacité d'engagement, la mission considère que le Hub **devra rapidement être doté de la personnalité morale.** Cette solution semble indispensable pour garantir l'autonomie juridique et financière du Hub, pour sécuriser ses financements dans un cadre pluriannuel, pour nouer des partenariats durables et pour accompagner des dynamiques de valorisation et la diversification des ressources.

Les modalités envisageables pour créer cette personnalité morale sont potentiellement nombreuses. La mission considère que la formule juridique retenue devra bénéficier de trois qualités essentielles : la mutabilité, la souplesse en gestion et la sécurité juridique. Elle recommande ainsi de concentrer la réflexion autour de deux options :

- Celle d'un **établissement public** : cette option présente l'avantage de la plasticité au moment de la conception du projet, même si cette plasticité ne serait que provisoire ; le Hub constituerait en effet une nouvelle catégorie d'établissement public qui, en vertu de l'article 34 de la Constitution[23],

[21] Un autre exemple de ces organismes à double face est l'Office national des forêts, créé par une loi de 1964 et mis en place en 1966, qui est chargé de la gestion et de l'équipement des forêts et terrains à boiser ou à restaurer appartenant à l'État.

[22] Ce qui caractérise la notion d'activité marchande, c'est le fait d'offrir à titre onéreux des biens ou des services sur un marché (CJCE, 18 juin 1998, Commission c/ Italie, aff. C-35/96, R. p. I-3851, point 36). La jurisprudence communautaire est en effet attachée non à la forme juridique de l'entité, mais à la nature des activités que celle-ci mène. Elle considère ainsi comme une « entreprise » au sens du droit communautaire « toute entité exerçant une activité économique indépendamment du statut juridique de cette entité » (CJCE, 23 avril 1991, Höfner, n°C-41/90).

[23] Selon une formule inchangée depuis 1979 (CC, n° 79-108 L, 25 juillet 1979, Agence nationale pour l'emploi, R. p. 45), le Conseil constitutionnel considère que relèvent d'une même catégorie au sens de l'article 34 les établissements « dont l'activité s'exercice territorialement sous une même tutelle administrative » et dont la spécialité est « analogue ». Deux critères se cumulent donc pour apprécier cette notion de « catégorie d'établissement public » : celui du contrôle exercé sur l'établissement, et celui de la nature de l'activité. Ce second critère sera prépondérant dans le cas du hub. Il fait l'objet de jurisprudences évolutives : par exemple, l'Agence nationale de l'habitat (ANAH) a été considérée comme constituant à elle seule une catégorie dont les ressources devaient être définies

devrait voir ses missions, ses ressources et ses règles de gestion[24] établies directement par la loi. A cette source de rigidité s'ajouteraient les contraintes induites par le principe de spécialité des établissements publics, qui leur interdit d'engager des initiatives en dehors des missions définies par son texte constitutif.

- Celle du **groupement d'intérêt public (GIP)[25]** : le recours à cette forme juridique présente des avantages indéniables de souplesse ; elle met dans les mains des autorités constitutives du Hub, dans un cadre contractuel – la convention constitutive du GIP - la responsabilité de définir ses missions, ses règles de fonctionnement et ses modalités d'intervention ; la forme du GIP n'implique en effet pas le recours à la loi[26] ; elle est donc celle qui offre le plus de capacité d'évolution dans le temps. La mission alerte néanmoins sur le fait que la souplesse associée au GIP peut se perdre dans le temps en l'absence d'une gouvernance adaptée. En présence d'un grand nombre de partenaires et en l'absence d'une gouvernance adaptée du groupement, garantissant des prises de décision rapide, la convention constitutive du GIP peut vite devenir très compliquée à modifier. Parce qu'elle est souvent trop détaillée, elle peut être source de blocages multiples dans l'organisation d'une structure telle que le Hub.

En cohérence avec les options retenues dans le rapport la **formule du GIP présente davantage de souplesse, mais pour préserver cette souplesse,** il faut éviter le travers consistant à multiplier les parties prenantes décisionnaires (qui par exemple, peuvent être très nombreux dans la sphère publique)[27].

La création d'un GIP peut **se faire en dotant le futur hub d'un capital social**[28], mais cela ne relève pas d'une obligation. Le recours à un capital nous paraît de bonne pratique si une partie de la valorisation des activités du Hub peut se réaliser sous la forme de prises de participation.

Au-delà du choix de la forme juridique la plus appropriée, les règles de gestion du Hub dépendront du **régime juridique induit par la nature de son activité**. En effet, les règles juridiques ne sont pas directement liées à la forme juridique qui sera choisie dans la mesure où le droit communautaire de la concurrence et des aides

---

par la loi (section des travaux publics, 8 mars 2005, n° 371 188, Projet de décret relatif aux conditions d'attribution des aides à la construction, à l'acquisition et à la réhabilitation des logements et modifiant le code de la construction et de l'habitation). Il en a été de même de l'établissement chargé des sondages, diagnostics et opérations de fouille d'archéologique préventive, doté d'un monopole (AG, 4 février 1999, n° 363 144, Projet de loi relatif à l'archéologie préventive, EDCE 2000 p. 76)

[24] En vertu d'une décision de 1964, il s'agit des règles qui « fixent le cadre général de son organisation et de son fonctionnement » (CC n° 64-27 L, 17 et 19 mars 1964 Radiodiffusion-Télévision , R. p. 33).

[25] Pouvant avoir une activité de service public industriel et commercial.

[26] La loi n°2011-525 du 17 mai 2011 de simplification et d'amélioration de la qualité du droit, dite loi Warsmann, a clarifié les choses de ce point de vue : elle laisse les membres constitutifs libres de créer ce type d'organisme dès lors que son objet est d'exercer ensemble des activités d'intérêt général à but non lucratif, en mettant en commun les moyens nécessaires à leur exercice et, d'autre part, permet que le GIP ait une durée indéterminée si c'est le choix de ses membres.

[27] Contrairement à cet égard au modèle retenu pour l'INDS qui est aussi un GIP, associant 15 partenaires différents.

[28] Les GIP peuvent être constitués avec capital (article 104 de la loi n°2011-525 du 17 mai 2011).Les personnes morales de droit public et les personnes morales de droit privé chargées d'une mission de service public doivent alors détenir ensemble plus de la moitié du capital (article 103 de la loi n°2011-525 du 17 mai 2011). La contribution des membres aux dettes du groupement est déterminée à proportion de leur part dans le capital (article 108 de la loi n°2011-525 du 17 mai 2011).

d'Etat impose un certain nombre de contraintes supplémentaires. Un enjeu clé à cet égard est de déterminer les conditions d'articulation entre ses activités d'intérêt général et ses activités marchandes. Pour les activités non marchandes, une assez grande liberté d'organisation est laissée aux Etats membres, les activités marchandes sont en revanche strictement encadrées par le principe de concurrence. Des exemples d'entreprises potentiellement concurrentes du Hub sur ces dernières fonctions ont été présentés pendant les auditions. La question du positionnement du Hub se pose ainsi au regard de la réalité d'un marché ouvert à ces entreprises. Le droit communautaire de la concurrence s'attache à cet égard à définir un certain nombre de règles visant à assurer une étanchéité économique entre activités marchandes et non marchandes. Cela implique principalement d'éviter que les bénéfices tirés de l'exercice de prérogative de puissance publique ou de financements publics ne placent le Hub dans une situation qui viendrait fausser la concurrence dans le cadre de ses activités marchandes[29].

Au-delà de cet effort de séparation, la mission estime qu'il faudrait privilégier les modèles/règles offrant le plus de latitudes/flexibilités possibles au Hub, si la formule du GIP était retenue. Les contraintes qui s'exercent sur un certain nombre de GIP sont incompatibles avec les ambitions que l'on souhaite atteindre dans ce projet, et qui supposent de :

- Pouvoir recruter des compétences qui sont aujourd'hui rares, chères et convoitées, ce qui implique des contrats de droit privé et des rémunérations cohérentes avec le niveau du marché ;
- Avoir une souplesse de gestion que les procédures habituellement appliquées (règles de la comptabilité publique, contrôle économique et financier) ne favorisent pas.

Sur ces différents aspects, les textes laissent un espace de choix et les décisions reviendront à l'Etat, mais il est clair qu'une organisation enserrée dans des règles trop bureaucratiques ne sera pas en mesure de porter l'innovation et le développement d'une filière industrielle dont notre pays a besoin.

MODELE ECONOMIQUE

Compte tenu de l'importance stratégique du Hub, de sa contribution à la transformation du système de santé et des externalités multiples qu'il va générer, la mission considère que le Hub doit bénéficier, au-delà de cette période de démarrage, d'un soutien public pérenne. En effet, le Hub va vendre des services mais doit bénéficier d'une période d'incubation, avant de trouver son équilibre économique, en particulier les coûts fixes seront importants au départ : cela implique de mettre à profit un soutien public spécifique au démarrage avant la définition d'un modèle de partage de la valeur plus pérenne. De plus, le Hub est un intermédiaire de valorisation : il a vocation à restituer aux apporteurs de données une part prépondérante de la valeur créée à partir de leurs données. Enfin, dans la suite du plan IA, le Hub devrait jouer un rôle actif dans le soutien aux

[29] La jurisprudence communautaire a identifié des hypothèses dans lesquelles peuvent échapper à la qualification d'aide d'État les compensations de charges de service public, sous certaines conditions (CJCE, 24 juillet 2003, Altmark Trans GmbH, aff. C-280/00, Rec. p. I-7747), la principale étant d'établir de façon préalable, objective et transparente, les paramètres de compensation des obligations de service public endossées par l'entité concernées et d'éviter la surcompensation de ces activités.

jeunes entreprises innovantes et doit donc être en mesure d'assumer une part du risque que ces projets comportent.

Marché

Le potentiel de valorisation des activités du Hub se concentre principalement autour des industriels de santé, laboratoires pharmaceutiques et *medtech*. Ils sont notamment intéressés par l'exploitation des données qui seraient partagées via le Hub, à l'instar de ce qui se fait actuellement sur le champ du SNDS, pour mener par exemple – au travers de bureaux d'études – des analyses sur données en vie réelle, qu'elles soient ou non demandées par le régulateur (commission de la transparence ou CEPS) ou des recherches sur la personnalisation des traitements. Via une cotisation annuelle, et/ou une contribution au jeu de données, ce segment de marché est le plus à même de fournir des revenus concrets et stables au Hub, au-delà des financements publics.

Les startups innovantes du domaine de la santé forment aussi un potentiel de revenu important pour le Hub. Bien qu'elles soient moins bien financées que les laboratoire pharmaceutiques ou *medtech*, elles présentent un fort potentiel d'innovation et de rupture auquel le Hub pourrait directement contribuer. Les startups ont des profils bien plus risqués que les plus gros industriels. Il s'agira donc d'offrir au plus tôt des services à un nombre substantiel de startups afin de ne pas dépendre trop fortement de leur niveau de risque individuel.

De manière à tirer le meilleur de ces deux mondes, les consortiums rassemblant industriels de grande taille (e.g. laboratoire pharmaceutique) et startups sont probablement les plus naturellement indiqués pour l'utilisation du hub. Ils fournissent à la fois les opportunités d'innovation rapide et la taille critique qui permettrait de pérenniser les différents acteurs.

Offre de valeur

Parmi les services proposés par le Hub à l'ensemble des utilisateurs, certains sont particulièrement susceptibles d'intéresser les divers industriels et de générer des revenus pour le Hub :

- **L'accès** à des bases de données de grande taille et de qualité. Le Hub devra bénéficier d'un catalogue significatif de données pertinentes afin d'attirer systématiquement les industriels français et étrangers qui souhaitent manipuler des données en masse. La rétribution des producteurs de données doit les inciter à faire évoluer leurs données vers des standards de qualité promus par le Hub et reconnus à l'international.
- **L'appariement** des bases de données partagées via le Hub entre elles, ou avec des données industrielles dans le but d'enrichir ces dernières et d'ouvrir de nouvelles applications (recherche clinique, évaluation en vie réelle) pour les industriels. Cette opération technique d'alignement du patrimoine public avec des données privées est génératrice d'une grande valeur ajoutée.
- La **protection de la propriété industrielle** des clients demandera un effort particulier au Hub et est un prérequis indispensable pour certaines applications industrielles. Que ce soit pour des jeux de données de grande valeur ou des modèles prédictifs pré-entrainés, le Hub pourra fournir un mode de fonctionnement qui garantit la confidentialité complète des échanges sur la plateforme.
- Le **passage à l'échelle** et **la mise en production** de certains traitements. Le Hub doit être un moyen est non une finalité, il doit permettre à des acteurs industriels innovants de proposer des outils à

destination des autres acteurs, à commencer par les professionnels de santé qui contribuent à la collecte de la donnée et qui pourraient être intéressés de bénéficier d'outils de benchmark par exemple.

Chacune de ces activités est susceptible d'apporter une valeur ajoutée significative au Hub du point de vue industriel. L'offre de service du Hub s'adaptera à ses divers cas d'utilisation et devra fournir des tarifs échelonnés afin de correspondre au mieux aux demandes d'industriels. Dans cette perspective, le modèle du Hub peut donc s'inspirer à la fois des travaux en cours pour la tarification de l'accès au SNDS et des modalités d'accès aux cohortes qui cumulent souvent une adhésion sur le long terme donnant lieu à un droit de tirage associée à une tarification au projet. Le modèle économique devra prendre en considération les réflexions menées au niveau national sur la place du marché français et européen dans un contexte numérique dominé par les grands acteurs américains et chinois.

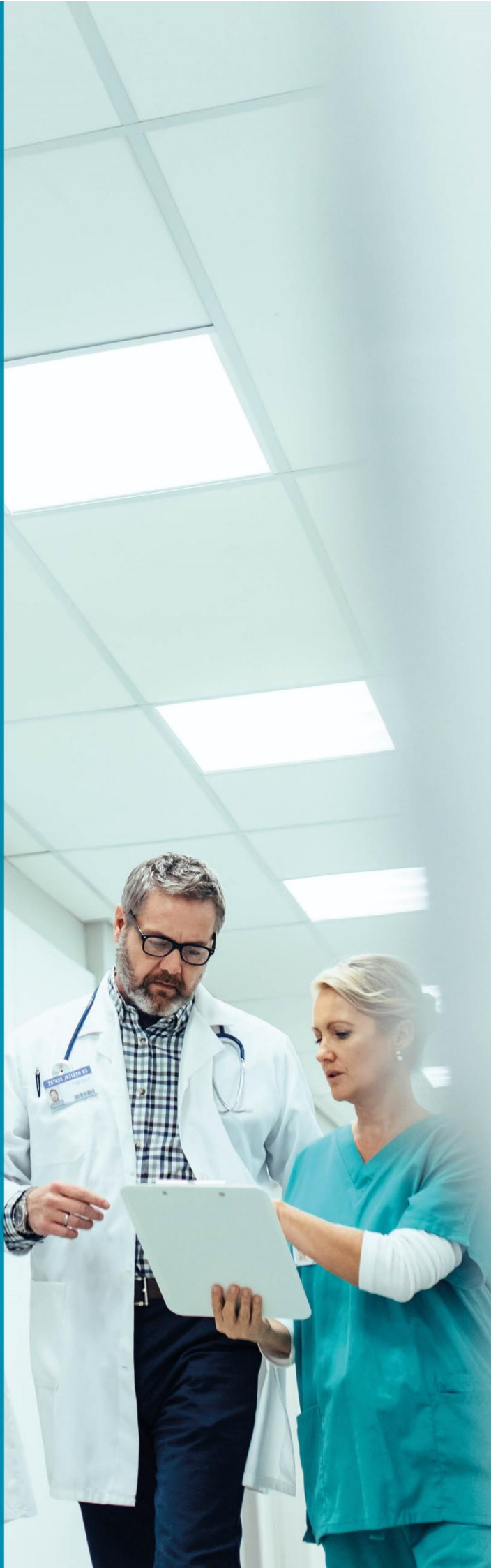
A ces modalités peuvent s'ajouter des réflexions autour du partage de la valeur (partage de la propriété intellectuelle, mise à disposition d'un droit d'usage de longue durée sur les technologies produites à partir des données, royalties…), en particulier dans le domaine de l'intelligence artificielle. Les projets sélectionnés via des appels à manifestation d'intérêt permettront d'explorer différentes pistes dans un premier temps, mais la mission insiste sur le fait que le Hub devrait disposer des équipes nécessaires pour instruire et construire, en collaboration avec l'écosystème, des contrats types afin de réduire les situations de blocage entre producteurs et utilisateurs, contribuer à améliorer la lisibilité du système et faciliter l'accès à la donnée[30]. A court terme, les efforts à consentir sont du côté du partage et du soutien aux usages, et il semble indispensable de ne pas freiner l'innovation par crainte de ne pas savoir estimer à sa juste valeur le potentiel de réutilisation. C'est en favorisant les usages que l'on acquerra collectivement une meilleure vision du véritable potentiel économique de la donnée.

Enfin, il peut être utile de rappeler qu'à court-terme, les principales pistes de valorisation économiques du Hub sont à trouver dans la réutilisation des données dites « rétrospectives », même si certaines données du catalogue seront actualisées de manière régulière. Les études sur les données du Hub, ainsi que l'apprentissage d'algorithmes d'intelligence artificielle, constituent donc les usages privilégiés dans un premier temps. On peut également imaginer des outils fondés sur les données et valorisables par les utilisateurs du Hub (tableaux de bord par exemple). Des outils de même type pourraient être exposés à des utilisateurs extérieurs au Hub à condition que les données mobilisées aient été anonymisées ou que les accès soient correctement sécurisés (outils à destination des professionnels de santé). L'absence de possibilité de remonter à l'individu pour des raisons de sécurité interdit la mise en « production » de certaines applications à l'intérieur de l'environnement du Hub. A titre d'exemple, une fois la preuve faite qu'un outil d'aide à la décision peut être efficacement construit sur les données du Hub, c'est au niveau des systèmes d'information des établissements hospitaliers, des constructeurs et logiciels des professionnels de santé ou encore du DMP que les applications pourront s'intégrer pour un usage opérationnel. Le Hub n'ayant pas de rôle de régulateur en termes de certification des DM intégrant des technologies d'intelligence artificielle.

[30] Le dispositif pourra s'inspirer de ce point de vue du *Data Trusts Support Organisation (DTSO)* mis en place par le gouvernement britannique suite au rapport de J Pesenti et W Hall. Le DTSO joue le rôle de « *trustee* » pour la conclusion d'alliance entre les opérateurs de données et les entreprises qui souhaitent développer des technologies à partir de ces données, particulièrement dans le champ de l'IA.

5

PATRIMOINE DE DONNÉES



5 PATRIMOINE DE DONNEES

Vision d'ensemble du patrimoine de données de santé

La liste qui suit présente des bases de données ou catégories de données régulièrement évoquées par les acteurs consultés et susceptibles d'intégrer à terme le catalogue de données partagées par le Hub. Pour chacune d'entre elles, la mission propose une synthèse des attentes exprimées par les parties prenantes lors des auditions ainsi qu'une première appréciation de leur niveau d'accessibilité et de maturité. Cette liste n'est toutefois pas exhaustive, les travaux d'identification des ressources pertinentes pour le Hub devront être poursuivis, certaines données non décrites ici pourraient en effet gagner à être inscrites au catalogue rapidement (certaines enquêtes, données de surveillance et de pharmaco-vigilance…).

Système national des données de santé (SNDS)

Données médico-administratives constituées des données du SNIIRAM, du PMSI, et du CépicDc (causes médicales de décès). A horizon 2020, le SIMDPH (données relatives au handicap) et un échantillon de données de l'assurance maladie complémentaire devraient venir compléter ces données.
Les utilisateurs de ces données sont déjà nombreux aujourd'hui (plusieurs centaines), notamment à des fins de recherche publique ou dans le cadre des missions de service public de l'Etat, des agences sanitaires ou opérateurs ; ou plus récemment les industriels par le biais de bureaux d'étude. Mais des usages nouveaux se développent, visant notamment à exploiter la dimension « big data » de cette base de plusieurs centaines de téraoctets. Une expérimentation est actuellement en cours via un partenariat entre la CNAM et l'Ecole Polytechnique pour construire sur d'importantes cohortes des vues « parcours de soin » plus facilement grâce à une plateforme technologique expérimentale reposant sur une architecture distribuée des fichiers. Ces travaux permettront d'investiguer des questions de détection de signaux faibles, de classification de parcours de soin…
Des start-ups commencent également à s'intéresser à ces données. Une application proposée par la start-up Sêmeia vise à développer des programmes d'accompagnement des patients incluant plusieurs modalités (outils numériques, accompagnement par des professionnels de santé), avec des interventions personnalisées en fonction des besoins et des risques de non adhésion au traitement. La start-up CanopyHealth développe des outils d'aide à la décision en mobilisant des données du SNDS chaînées avec des données socio-démographiques pour prédire au niveau d'une unité géographique le nombre de patients en risque de ré-hospitalisations potentiellement évitables, ceci dans le but de permettre à l'ARS (Agence Régionale de Santé) de dimensionner au mieux l'offre de soin.

Attentes de l'écosystème	Maturité de la donnée
Très fortes attentes de l'écosystème pour avoir la possibilité de valoriser le SNDS avec des outils à l'état de l'art.	Très forte : la base de données et les technologies existent, la gouvernance est inscrite dans la loi.

Systèmes-fils

Il existe environ 200 systèmes-fils du SNDS (bases intégrant un échantillon du SNDS). Ces bases sont soumises à un contexte législatif imposant une mise en conformité des systèmes d'information les hébergeant au référentiel de sécurité du SNDS pour mars 2019. Au-delà du poids en termes d'investissement matériel et financier que représenterait une mise à niveau de l'ensemble de ces systèmes d'information, la CNIL pourrait être sensible à éviter une démultiplication de ces systèmes ou « bulles ». En effet, cette démultiplication rend la traçabilité de la circulation et du traitement de la donnée plus complexe.

Attentes de l'écosystème	Maturité de la donnée
Très fortes attentes des producteurs/responsables de traitement pour la mutualisation d'une plateforme sécurisée de mise à disposition des données.	Forte : les bases de données et les technologies existent.

Données issues de la recherche et des grandes cohortes nationales

Les grandes cohortes nationales rassemblent des données longitudinales sur des échantillons de patients ou citoyens (certaines d'entre elles sont des systèmes-fils).

Attentes de l'écosystème	Maturité de la donnée
Fortes attentes des promoteurs de cohortes pour gagner en visibilité et valoriser économiquement des données dont la collecte et la mise en qualité sont coûteuses. Fortes attentes autour de la mutualisation de plateformes technologiques et de ressources juridiques et techniques pour faciliter l'appariement avec des nouveaux flux. Fortes attentes des utilisateurs pour obtenir davantage d'information en amont de l'accès, pour un accès simplifié et des outils à l'état de l'art pour la valorisation des données. Fortes attentes des industriels pour renforcer le dialogue précédant la conception des cohortes afin, lorsque cela est souhaité, de penser ensemble la finalité industrielle à venir.	Forte : les bases de données et les technologies existent. On peut citer notamment les cohortes financées par les Investissements d'avenir (Constances, Hope-Epi, i-Share, Coblance, CKD-REIN, OFSEP, E4N, ELFE, RADICO, CANTO...), qui doivent en outre avoir mis en place une procédure de partage de données mais dont les infrastructures sont peu mutualisées. On peut aussi penser à celles qui s'inscrivent dans le plan Alzheimer par exemple (MEMENTO) ou sur des thématiques comme la nutrition (Nutrinet-Santé).

Données d'hospitalisation et dossier patient informatisé (DPI) hospitalier et de GHT

Les données d'hospitalisation sont extrêmement riches car elles concentrent plusieurs sources avec les DPI : suivis de diagnostic, traitements, échanges écrits entre les professionnels de santé, données de biologie, de médicaments, d'imagerie, des urgences.

Attentes de l'écosystème

Les données hospitalières sont extrêmement importantes pour la compréhension des parcours de soin, notamment des patients atteints de pathologies lourdes. L'imagerie médicale et les signaux (ex. ECG) ont par ailleurs un très fort potentiel compte tenu des techniques d'intelligence artificielle qui existent aujourd'hui. Les comptes-rendus médicaux contiennent beaucoup d'informations susceptibles de redresser des données structurées comme le PMSI et de compléter les informations, entre autres sur le contexte clinique, les facteurs de risque (alcool, tabac…), les antécédents… Les données sont également mobilisables pour améliorer le pilotage de l'offre de soin et développer des outils d'aide à la décision.

Maturité de la donnée

Très faible à forte : la difficulté provient de la multiplicité et la dissémination des systèmes d'information hospitaliers. Certaines sources peuvent facilement être remontées comme les RPU qui sont déjà rassemblées dans le cadre de la fédération FEDORU et ont vocation à intégrer le SNDS à l'horizon 2020, d'autres sources sont relativement structurées comme la biologie ou la microbiologie ou encore les données du circuit du médicament. Les difficultés rencontrées au niveau des données hospitalières s'expliquent par des systèmes d'information hétérogènes et ayant évolué. De plus, certaines données ne sont pas numérisées, des ressources sont nécessaires pour collecter puis rassembler ces données. L'imagerie médicale (radiologie, ophtalmologie, dermatologie, anatomopathologie) et les signaux (encéphalogramme, électrocardiogramme, magnétoencéphalographie) sont moins hétérogènes et pourraient être plus rapidement valorisables si les données étaient rassemblées. Les historiques cliniques et comptes-rendus médicaux sont peu voire non structurés, l'utilisation d'ontologie ou de techniques d'intelligence artificielle sera essentielle pour les valoriser à leur plein potentiel.

Registres épidémiologiques et de pratiques

De nombreux registres épidémiologiques (cancers, maladies rares, malformations congénitales, …) ou registres de pratiques (interventions chirurgicales par exemple) existent en France. Ils sont alimentés et financés par divers acteurs institutionnels ou privés. Santé Publique France, l'INCa et l'Inserm cofinancent notamment une quarantaine de registres pour un montant de 7 millions d'euros de fonctionnement.

Attentes de l'écosystème

Les registres représentent un fort intérêt en tant que « gold standard » compte tenu de leur exhaustivité sur un territoire donné.
Fortes attentes des utilisateurs pour davantage d'information en amont de l'accès, d'un accès simplifié et d'outils à l'état de l'art pour la valorisation des données.
Fortes attentes de certains promoteurs de registres à être enrichis par des extractions du SNDS dans une optique d'amélioration du suivi des patients (parfois près de 80% de perdus de vue à un an) et de réduction des coûts de collecte (à l'instar du registre cancer du Poitou Charentes). Des travaux en ce sens ont été lancés par Santé Publique France.
Le rassemblement des registres pourrait faciliter l'harmonisation et le partage de bonnes pratiques de collecte comme le fait, par exemple, le réseau FRANCIM pour les registres sur le cancer. Il pourrait également contribuer à sécuriser ce patrimoine en en assurant la pérennisation.

Maturité de la donnée

Maturité variable. Certains registres épidémiologiques sont en place depuis longtemps (par exemple sur le cancer). En revanche le développement de registres par des sociétés savantes ou des conseils nationaux professionnels portant sur des activités spécifiques (par exemple le registre sur les bioprothèses valvulaires aortiques implantées par cathéter - TAVI) sont encore peu développés en France.

Données du Dossier pharmaceutique (DP)

Le Dossier Pharmaceutique recense, pour chaque bénéficiaire de l'assurance maladie qui le souhaite, tous les médicaments délivrés en ville au cours des quatre derniers mois, qu'ils soient prescrits par le médecin ou conseillés par le pharmacien (21 ans pour les vaccins, 3 ans pour les médicaments biologiques). Le DP est également accessible aux pharmaciens et médecins exerçant en établissement de santé.

Attentes de l'écosystème

Ces données sont nécessaires pour avoir des informations sur les prescriptions de médicaments non remboursés (puisque les autres sont déjà dans le SNDS).

Maturité de la donnée

Forte : les données existent et sont structurées, elles sont gérées par l'ordre des pharmaciens.

Données des laboratoires de biologie médicale

Les laboratoires de biologie médicale « de ville » transmettent de manière quasiment systématique les résultats de biologie aux médecins généralistes équipés de logiciels de dossier patient. Les analyses les plus communément réalisées en laboratoires de biologie médicale incluent l'HbA1c pour suivre l'équilibre d'un diabète, l'INR (du taux de prothrombine) pour suivre un traitement anticoagulant par antivitamine K, etc.

Attentes de l'écosystème	Maturité de la donnée
Ces données sont indispensables car elles permettront d'enrichir le SNDS des effets des traitements (entre autres) sur un périmètre restreint (médecine de ville).	Moyenne à forte : les données sont relativement structurées et les éditeurs de logiciels sont relativement concentrés (4 éditeurs pour les laboratoires de biologie médicale pour 78% des professionnels de santé). Un alignement terminologique des résultats de biologie est toutefois nécessaire, LOINC n'étant pas encore systématiquement utilisé.

Données des cabinets de médecine de ville

Les données de médecine de ville regroupent les données collectées par les logiciels de médecine de ville, généralistes ou spécialistes. On y trouve par exemple des diagnostics codés selon des terminologies propres (une cinquantaine de codes souvent), qu'il s'agisse de diagnostics, motifs de consultation ou motifs de prescription, des médicaments prescrits, généralement assortis de leurs codes ATC et UCD, des résultats de biologie, directement reçus des laboratoires, des vaccins injectés dans le cabinet…

Attentes de l'écosystème	Maturité de la donnée
Ces données sont nécessaires pour disposer des diagnostics et des données de prise en charge en ville, et suivre le parcours d'un patient. Elles peuvent également être exploitées pour améliorer les outils des professionnels de santé. Ces derniers sont intéressés d'avoir des informations leur permettant de comparer leur patientèle à celle de certains confrères, de visualiser simplement et rapidement des éléments les plus importants du dossier du patient qu'ils reçoivent en consultation, d'obtenir de manière simple et ergonomique des indications issues des recommandations de la HAS (Cegedim travaille à l'implémentation de ce type de fonctionnalités par exemple).	Faible : Malgré une forte informatisation de la médecine de ville, ces données ne sont pas complètement structurées, pas ou peu collectables à l'heure actuelle, les éditeurs de logiciels sont nombreux (15 éditeurs de logiciels pour les médecins généralistes et spécialistes représentent 80% des parts de marché).

Données d'imagerie de ville

Les données d'imagerie de ville sont produites par des centres de radiologie ou autres cabinets de ville et peuvent être disponibles rapidement et facilement. Les examens réalisés incluent des échographies, des mammographies, des scanners, des IRM, etc.

Attentes de l'écosystème	Maturité de la donnée
Ces données sont très intéressantes pour développer des outils d'aide au diagnostic avec des méthodologies d'intelligence artificielle. L'enrichissement du SNDS avec ces données permet d'enrichir la description des effets des traitements, entre autres…	Moyenne : ces données existent mais sont décentralisées, elles sont non structurées mais les technologies de valorisation sont mûres.

Données génomiques

Les données génomiques proviennent des plateformes de séquençage à très haut débit du génome humain et ont une visée diagnostique et de suivi thérapeutique et permettent d'aller vers une médecine de plus en plus personnalisée.

Attentes de l'écosystème	Maturité de la donnée
Ces données sont primordiales pour développer des prises en charge personnalisées. L'enrichissement de ces données avec d'autres données du parcours de soin au catalogue du Hub sera déterminant pour identifier les profils génomiques répondant mieux à un traitement ou pour mieux cibler la prévention.	Moyenne : ces données existent en faibles quantités, le plan France Médecine Génomique vise à industrialiser les séquençages et à rassembler les données pertinentes au niveau d'un "Hub" génomique, le CAD (collecteur analyseur de données) qui devra être interopérable avec le Hub.

Données d'évaluation par les patients

Les patients sont de plus en plus amenés à remplir des questionnaires de satisfaction (exigences réglementaires dans certains cas). On peut également noter l'usage des mesures appelées PROM (Patient reported outcome measures - pour la manière dont il évalue son état de santé) et PREM (Patient reported experience measures - pour la manière dont il a vécu ses soins) qui seront de plus en plus utilisées pour évaluer la qualité des soins.

Attentes de l'écosystème	Maturité de la donnée
Il y a un véritable intérêt pour intégrer les indicateurs issus du patient dans l'évaluation d'un traitement ou d'une prise en charge, par exemple à la demande d'agences sanitaires de régulation. Les associations de patients récoltent elles aussi parfois des données par le biais d'enquêtes auprès des patients pouvant renseigner sur leur qualité de vie.	Faible : la collecte de ces données est assez récente et non généralisée.

Données contextuelles

Les données contextuelles peuvent inclure des données sociodémographiques, des données environnementales, des facteurs de risques, des données scolaires, etc.

Attentes de l'écosystème	Maturité de la donnée
Un certain nombre de données peuvent être extrêmement enrichissantes pour l'analyse des données de santé notamment dans une optique de réduction des inégalités sociales de santé.	Variable : les sources sont à identifier, ainsi que leurs circuits d'accès, finalités et contraintes d'utilisation. Certains projets d'appariement sont en cours (SNDS et échantillon démographique permanent de l'Insee).

Données issues de la télémédecine et des dispositifs médicaux connectés

Les données de santé issues de projets numériques (ex. télémédecine, dispositifs médicaux connectés, etc.) peuvent également constituer des sources d'informations très riches.

Attentes de l'écosystème	Maturité de la donnée
Les données de télésurveillance, télémédecine et plus généralement d'objets connectés sont d'un grand intérêt pour décrire la prise en charge en ambulatoire, les modes de vie, permettre un suivi à domicile… Elles permettent aussi de réaliser des essais cliniques décentralisés dans des conditions plus réelles mais également pour des coûts moindres.	Faible : la difficulté pourra être dans le chaînage de ces informations avec les autres sources du catalogue. A noter que les dispositifs médicaux connectés ne sont pas tous remboursés par l'Assurance Maladie.

Données régionales sur le parcours des patients

Les ARS sont amenées à collecter des données très riches de parcours, par exemple avec ViaTrajectoire, ou les projets pilotes de territoires santé numériques (Auvergne Rhône-Alpes, Bourgogne Franche-Comté, Ile-de-France, Nouvelle Aquitaine, Océan Indien).

Attentes de l'écosystème	Maturité de la donnée
Il existe une véritable attente autour de l'exploitation des données de parcours de soin d'un patient pour favoriser la coordination de la prise en charge des pathologies chroniques par exemple.	Moyenne : les données de parcours ont un niveau de structuration avancée, en revanche, les usages restent encore limités.

Données du secteur médico-social

Une partie des données issues du secteur médico-social sera intégré au SNDS à horizon 2020, pour ce qui est des données du handicap. Des données issues des EHPAD nécessitent encore d'être structurées pour être exploitables.

Attentes de l'écosystème	Maturité de la donnée
Il y a un enjeu fort autour du vieillissement et de la population qui est une thématique prioritaire.	Faible : globalement, les structures médico-sociales sont à un niveau d'informatisation moindre par rapport aux structures sanitaires.

Données du Dossier Médical Partagé (DMP)

Le DMP est en cours de déploiement et servira de carnet de santé à l'ensemble des usagers du système de santé français. Le chantier « virage numérique » de la Stratégie de Transformation du Système de Santé piloté par Dominique Pon et Annelore Coury vise à élargir le DMP à un portail plus large pour le patient et le professionnel de santé

Attentes de l'écosystème	Maturité de la donnée
Le DMP pourra à terme se substituer en partie à l'alimentation d'un SNDS élargi dans la mesure où il rassemblera par construction un certain nombre d'informations chaînées p pour un individu donné (remboursement de l'assurance maladie, biologie de ville, comptes-rendus de consultation, etc...).	Faible : le DMP est en cours de déploiement.

Dossier médical en santé au travail (DMST)

Le Dossier Médical en Santé au Travail (DMST) est constitué d'éléments objectifs communicables (tels que les antécédents médicaux, les résultats d'examens médicaux, l'historique des postes occupés et des entreprises, etc.) et des éléments subjectifs non communicables (tels que les confidences du salarié et les appréciations personnelles du médecin du travail). Informatisé depuis 2010, il couvre aujourd'hui une population de 15 millions de salariés.

Attentes de l'écosystème	Maturité de la donnée
Les données du DMST pourraient être utilisées à des fins de recherche, notamment pour évaluer l'impact de l'environnement de travail sur la santé et la sécurité (stress, travail posté, risques chimiques…)	Forte : Utilisation d'un outil unique enrichi par les médecins du travail. Les données donc structurées et codées et pourraient facilement être liées au DMP.

Données en Sciences de la Vie (niveau européen)

Le projet ELIXIR, infrastructure de recherche inscrite sur la feuille de route européenne (ESFRI) depuis 2006, réunit les principales organisations européennes du secteur des sciences de la vie pour la gestion et la sauvegarde du volume croissant de données générées par la recherche financée par des fonds publics. Elle coordonne, intègre et maintient les ressources en bioinformatique dans ses États membres et permet aux utilisateurs universitaires et industriels d'avoir accès à de nombreux services dans le domaine de la donnée et de son exploitation. ELIXIR coordonne le projet EOSC-Life, qui regroupe les 13 infrastructures de recherche biologiques et médicales (BBMRI ERIC : biobanques, EATRIS ERIC : recherche translationnelle en médecine et ECRIN ERIC : réseau de recherche clinique) de l'ESFRI afin de créer un espace collaboratif ouvert pour biologie numérique.

Attentes de l'écosystème	Maturité de la donnée
La gestion et la sauvegarde au niveau européen des données générées par l'ensemble des 21 états membres de l'infrastructure couvre tous les domaines des sciences de la vie (y compris en biobanking, en recherche translationnelle et en recherche clinique).	Forte : ELIXIR est sur la feuille de route ESFRI depuis 2006 et en est l'un des projets emblématiques totalement opérationnels (Landmark).

Version initiale du catalogue de données partagées via le Hub

Certaines sources, pour lesquelles les acteurs de l'écosystème ont exprimé de fortes attentes, apparaissent comme suffisamment matures et accessibles pour être référencées au catalogue du Hub dès son lancement.

C'est notamment le cas du SNDS, dont les perspectives d'enrichissement avec d'autres sources ouvrent un large spectre de cas d'usages potentiels. A noter toutefois, que l'enrichissement pérenne de ces sources avec le SNDS implique une évolution législative. En effet, les accès aux données du SNDS visant à enrichir de manière permanente une base de données telle qu'un registre ou une cohorte, ou plus généralement, tout type d'appariement pérenne des données du SNDS avec d'autres sources de données (par exemple sociales, comportementales, cliniques etc..) sont aujourd'hui interdits. Ce point limite le potentiel de réutilisation de ces données et, même si l'accès sur projet spécifique est autorisé, il n'est souvent pas considéré comme rentable de conduire un projet d'appariement pour une utilisation ponctuelle.

Les « systèmes fils » pourraient également être intégrés - d'abord à travers une sélection de quelques bases à des fins de test, progressivement étendue à l'ensemble des 200 systèmes fils. Cela permettrait de mutualiser les investissements requis pour mettre l'hébergement de ces bases en conformité avec les exigences du référentiel de sécurité du SNDS.

Une sélection de grandes cohortes nationales pourrait également être référencée au catalogue dès le lancement. Ce pourrait par exemple être le cas d'une partie des grandes cohortes thématiques et en population générale, dont certaines sont financées par les investissements d'avenir et sous responsabilité de l'Inserm, de la cohorte Cancer de l'INCa, ou encore de la cohorte Elfe de l'INED. En complément, un travail d'identification et de sélection des registres stratégiques pourrait également alimenter la première version du catalogue.

Les données hospitalières présentent enfin un potentiel très riche. Si le patrimoine reste fragmenté, certains gisements consolidés aux niveaux des « entrepôts » devront intégrer les premiers projets « pilotes » du Hub. La visibilité de ces bases et leur valorisation économique permettront de soutenir les efforts des producteurs dans la collecte et la mise en qualité nécessaires à leur constitution et leur exploitation.

D'une manière générale, les données d'imagerie médicale et les signaux (ex. ECG) présentent un très fort intérêt compte tenu des techniques d'intelligence artificielle qui existent aujourd'hui : en particulier, les données d'imagerie médicale (radiologie, ophtalmologie, dermatologie, anatomopathologie) et les signaux (encéphalogramme, électrocardiogramme, magnétoencéphalographie). Le projet annoncé par le Conseil national professionnel de la radiologie française (G4) pourrait s'inscrire tout naturellement dans la démarche de rassemblement des données qu'entreprend le Hub. A noter toutefois que la valorisation de ce type de données est déjà très avancée sur le plan international et se distinguer sur ce segment sera difficile. La capacité à traiter ces données, appariées à d'autres sources, pourrait être un élément différenciant.

Potentiel de valeur de l'appariement de sources de données partagées via le Hub

Au-delà de la mise à disposition d'un catalogue de données accessibles de façon permanente, sous réserve d'une autorisation, le Hub offrirait aussi la possibilité d'apparier ces bases entre elles ou avec d'autres sources, dans le cadre de projets ponctuels. En effet, le cloisonnement des sources et les difficultés rencontrées par les acteurs pour les rapprocher, tant sur le plan technique que juridique, confirment que cette offre de service du Hub est essentielle. Certains acteurs ont développé une réelle expertise en la matière, comme la CNAM ou la plateforme de pharmaco épidémiologie (PEPS), mais cela reste très en-deçà des possibilités et des besoins.

Au cours des auditions menées par la mission, un grand nombre de cas d'usage résultant de la capacité à rapprocher des sources ont été présentés. A titre d'exemple, nous en dressons ci-dessous une liste non exhaustive. Bien souvent, les exemples proposés impliquent les données du SNDS, données pour lesquelles les acteurs ont témoigné le plus grand intérêt, mais le rapprochement d'autres sources représente également un important potentiel.

En recherche médicale par exemple, le rapprochement du SNDS avec diverses sources permettrait certainement d'importantes avancées. Leur chaînage avec des données génomiques et cliniques permettrait de mettre en évidence des biomarqueurs et prédispositions génétiques à certains cancers, voire d'identifier des états précancéreux (la start-up VitaDx investit notamment le domaine du cancer de la vessie dans l'idée de permettre aux urologues de prendre en charge leurs patients plus précocement pour avoir davantage d'options thérapeutiques). En effet, les données du SNDS, bien que non médicalisées, permettent d'avoir une vision plus complète du parcours de soin d'un patient en ville et à l'hôpital. L'appariement avec des données produites pour l'observation des maladies rares (cohortes, Banque Nationale des Données Maladies Rares) permettrait également d'avoir une vision plus détaillée de ces maladies qui sont nombreuses (plusieurs milliers) et touchent une personne sur vingt dans le monde. Les connaissances ainsi produites permettraient aussi de faciliter le pilotage des politiques publiques issues des trois plans nationaux. Ces données de consommation de soin, associées à des données cliniques (biologiques, radiologiques), peuvent également favoriser la découverte de « nouvelles molécules » et contribuer à l'exercice d'une médecine de plus en plus personnalisée. En brassant ces gros jeux de données, on pourrait mettre en évidence l'impact de co-médications sur le taux de réponse à un traitement (par exemple la chimiothérapie). Dans un tel contexte, les données du SNDS permettent de compléter l'information parcellaire contenue dans les dossiers des patients sur les médicaments qu'ils déclarent prendre par ailleurs. Des données directement issues des patients permettraient également de mieux circonscrire toutes les interactions (et de mieux mesurer l'observance). Des résultats très prometteurs ont été évoqués dans ce sens par l'Institut Curie. De même, certaines recommandations en termes de posologie et dosage de médicaments, pour des typologies de patients très particulières, peuvent ne pas être respectées et produire de bien meilleurs taux de réponse. L'appariement systématique des registres et du SNDS autoriserait une meilleure caractérisation des profils de patients face à telle ou telle option thérapeutique, notamment via le développement de modèles de prédiction du risque de complication (suite à une chirurgie baryatrique dans la prise en charge d'une forte obésité par exemple). L'appariement systématique permet aussi d'améliorer le suivi des patients, qui est difficile et coûteux. L'appariement avec des données de prises en charge de nouveaux nés prématurés fournit à ce titre des indications utiles sur les premières années de leur vie.

Les données d'imagerie ou de signaux sont des données au potentiel bien identifié pour le développement d'outils d'aide au diagnostic ou à la décision thérapeutique. S'il est aujourd'hui possible de détecter automatiquement, avec un petit taux d'erreur, certaines tumeurs à partir de scanner, comme le propose par exemple Therapixel pour le cancer du sein, cela implique d'annoter au préalable d'importants jeux de données pour calibrer (réaliser « l'apprentissage ») de l'algorithme. Le rapprochement de ces données d'images ou de signaux avec des données cliniques ou du SNDS pourrait automatiser en partie le travail d'annotation, voire réduire encore les taux d'erreurs en utilisant la suite des épisodes de soin renseignés dans le SNDS pour identifier sur le scanner quelque chose qu'un expert n'aurait pas forcément su détecter avec certitude. Mobilisant les données d'imagerie dans un contexte différent (pour mesurer et prédire l'efficacité d'un traitement par exemple), la start-up Owkin pourrait également améliorer les performances de ses outils en exploitant le SNDS. Le rapprochement des données présente aussi un grand intérêt pour automatiser certaines tâches comme la codification des actes. Ainsi pour la facturation hospitalière, des médecins renseignent des codes de diagnostics et d'actes pour tous les séjours, ces données constituent le PMSI et sont incluses dans le SNDS. Elles sont réputées être en partie biaisées[31]. La mobilisation des comptes-rendus médicaux textuels pour automatiser en partie cette tâche constituerait un moyen de faciliter le travail de ces médecins mais également de produire des données potentiellement moins biaisées et donc de meilleure qualité pour des réutilisations à des fins de recherche (comme le propose la start-up Sancare par exemple). D'une manière générale, les comptes-rendus médicaux contiennent énormément d'informations mais souffrent de leur manque de structure. Il faut donc développer des outils pour extraire et classer l'information pertinente. C'est essentiel pour traiter des données, par exemple, de RCP (réunion de concertation pluridisciplinaire) dans la prise en charge du cancer. L'élaboration de ces algorithmes sont à la frontière de la recherche mais pourrait bénéficier du chaînage des textes avec les données plus longues du SNDS, via la mise en relation du contenu de ces textes et les épisodes de soin passés ou futurs. A moyen terme, lorsque ces outils seront mûrs, ils pourront permettre de réduire le « retour au dossier » manuel aujourd'hui très utilisé pour la collecte d'informations précises sur un patient donné. Ce retour au dossier, et d'une manière générale la collecte manuelle, expliquent en grande partie pourquoi la production de données de qualité est extrêmement onéreuse. Développer des outils capables d'automatiser, même en partie, ces processus et réutiliser au maximum des données existantes comme celles du SNDS dans une logique de réduction des multiples saisies, permettrait d'obtenir à terme des données de meilleure qualité (davantage d'informations moins biaisées grâce au recoupement d'informations) pour un coût moindre.

Pour l'amélioration du système de soin dans son ensemble, le rapprochement de données est tout aussi essentiel. Il est, par exemple, indispensable de recueillir de plus en plus de données auprès du patient pour évaluer sa qualité de vie et la mettre en regard de la palette de traitements ou prises en charge disponibles. Certaines associations de patients s'investissent déjà particulièrement en ce sens et seraient très intéressées de pouvoir aller plus loin en analysant l'appariement des données du SNDS avec les données qu'elles collectent (Renaloo, le DiabeteLab de la Fédération Française du Diabète par exemple). L'appariement de registres épidémiologiques ou de pratiques avec d'autres sources de données, qu'elles soient de santé ou socio démographiques, rendrait possible l'étude de l'accessibilité à certaines options thérapeutiques (par

[31] Non seulement les données sont renseignées à des fins de remboursement et non à des fins médicales mais par ailleurs, l'évolution de la classification des actes (CCAM) n'est pas assez dynamique pour restituer la réalité des actes. En particulier les nouveaux actes chirurgicaux ne disposent pas de code dédié, donc bien que 70 000 patients aient été opérés en France par chirurgie robot assistée, il n'est pas possible de suivre cette prise en charge dans le PMSI à ce jour.

exemple la FIV) sur le territoire et fournirait des clés pour réduire les inégalités sociales de santé. L'appariement avec les données du SNDS permet aussi de prédire le risque de rejet de greffon et donc d'optimiser l'attribution de ces derniers en fonction des chances de succès. Il est évident que pour l'Etat, les agences sanitaires et opérateurs, le rapprochement du SNDS avec des données cliniques permettrait de faciliter l'exercice de leurs missions. A titre d'exemple, l'ANSM (Agence Nationale de Sécurité du Médicament) pourrait plus aisément identifier des effets indésirables si elle disposait également de résultats d'examens cliniques et biologiques, et non seulement des consommations de soin des citoyens. Si elle a pu récemment évaluer des effets indésirables des traitements d'anti-TNF alpha prescrits dans les maladies inflammatoires du côlon et de l'intestin, le même travail n'a pu être mené pour la polyarthrite rhumatoïde, car il faut prendre en considération le degré de gravité qui n'est disponible que dans les dossiers médicaux. La HAS (Haute Autorité de Santé) pourrait plus facilement produire des recommandations de bonnes pratiques sur la base des prises en charge ayant démontré des bonnes performances. A l'heure actuelle, ces institutions s'appuient sur des algorithmes pour inférer les diagnostics posés en ville, comme cette information n'est pas directement disponible dans le SNDS. Du côté des industriels des produits de santé, le chaînage du SNDS avec d'autres sources pourrait permettre d'atteindre un niveau de qualité suffisant pour que ces données puissent être considérées comme fiables pour réaliser des études médico économiques après une mise sur le marché d'un produit. S'agissant des dispositifs médicaux connectés, de plus en plus nombreux, pouvoir rapprocher les données du SNDS avec celles qui sont collectées par le dispositif est essentiel pour avoir une vision plus fidèle de son impact. Le service rendu par certains de ces dispositifs (par exemple les gestes médicaux et chirurgicaux assistés par ordinateur) peut être en effet très dépendant de ses conditions d'utilisation, et il est important de les intégrer à l'évaluation de leur efficacité.

A noter que les exemples donnés se concentrent souvent sur la réutilisation de données financées par la solidarité nationale mais qu'il pourrait être très intéressant, selon les cas de croiser des données privées, par exemple des données d'essais cliniques plus fines, avec d'autres sources (cliniques ou SNDS). A titre d'illustration, cela permettrait d'étudier les conditions de recrutement dans les essais cliniques en oncopédiatrie où les enfants en échec thérapeutique ne semblent pas toujours avoir accès de manière équivalente aux solutions proposées. De même, le croisement de données cliniques avec des données d'objets connectés objectiverait leur impact sur l'état de santé ou permettrait d'identifier des biomarqueurs numériques simples (tels que les données de sommeil) pour comprendre, aiguiller le patient et prédire l'évolution de certaines maladies (diabète, Alzheimer, d'après la start-up Rythm).

Interopérabilité des données

La France présente une forte fragmentation de ses systèmes d'informations dans l'ensemble des secteurs sanitaires. Si le SNDS actuel est une base de données homogène de grande couverture et enviée à l'étranger, elle représente une toute petite partie du patrimoine national des données de santé.

Ainsi, la plupart des données sont encore largement cloisonnées et captives dans les logiciels des professionnels de santé, ou dans des systèmes d'information hospitaliers presque tous différents d'un établissement à l'autre. A titre d'exemple, les entrepôts de données de santé d'établissement hospitaliers ou leurs déclinaisons à l'échelle des territoires (GHT) visent à décloisonner ces données de manière à les rendre partageables et exploitables. Pour autant ce processus d'intégration est à la fois long et coûteux. De plus, les données accessibles sont encore faiblement normalisées et donc peu partageables. La mise en cohérence de ces gisements représente donc un défi national.

Dans ce contexte, on peut noter les initiatives de l'administration, en particulier les travaux du Ministère des Solidarités et de la Santé et de l'ASIP qui a une mission de labélisation des logiciels, développe des modèles dans le cadre d'un volet « contenu des données » [32] (modèles de la lettre de liaison par exemple) et va mettre en place le centre de gestion des terminologies. Ce dispositif permettra de mieux diffuser les terminologies de référence et de promouvoir leur utilisation par les différents acteurs qu'ils soient professionnels de santé, institutionnels ou, en premier lieu, industriels du logiciel de santé.

Il est par ailleurs nécessaire de mettre en place une véritable feuille de route de l'intéropérabilité, à l'instar de ce que font d'autres pays tels que les Etats-Unis ou la Suisse. Les Etats-Unis ont, en effet, défini une feuille de route pluriannuelle (à deux et cinq ans) permettant aux différents acteurs d'acquérir une certaine visibilité, et ils organisent des bilans annuels pour faire l'état des lieux de l'adoption des standards par l'écosystème. D'importants moyens sont également mobilisés pour faire monter en compétence les éditeurs pour qui l'implémentation de standards d'interopérabilité toujours en évolution engendre d'importants coûts. Sans cela, certaines décisions, telles que l'adoption de la SNOMED[33] comme terminologie clinique de référence, entraîneraient une barrière à l'entrée de certains acteurs français. Quoiqu'il en soit les orientations en la matière, notamment le choix de terminologies pivot qui permettront de mieux partager l'information médicale, doivent être claires.

[32] Dans le cadre de sa mission de régulation, l'ASIP Santé publie deux référentiels : le CI-SIS (Cadre d'Interopérabilité des Systèmes d'Information de Santé) et la PGSSI-S (Politique Générale de Sécurité des Systèmes d'Information de Santé). Le CI-SIS est un recueil de spécifications d'interopérabilité, appelées volets, qui s'appuient sur des profils (IHE) et des standards (HL7, DICOM, …) internationaux. Le CI-SIS est découpé en couches : transport, service et sémantique. Au sein de la couche sémantique, chaque « volet de contenu » spécifie un format d'échange/partage de données (comme le CDA – Clinical Document Architecture) et des terminologies pour coder ces données.

[33] La dernière version de la SNOMED CT reprend maintenant la dénomination générique de la SNOMED.

La mission recommande de mettre en place d'importants moyens pour mener une véritable politique de l'intéropérabilité et en particulier de :

- Renforcer les moyens de l'ASIP pour la mise en œuvre du centre de gestion des terminologies et la valorisation de celui-ci auprès de l'écosystème ;

- Rendre obligatoire et opposable l'adoption de profils d'interopérabilité du CI-SIS dans les logiciels et les systèmes d'information en santé. Ceux-ci s'appuient le plus souvent sur des profils d'interopérabilité internationaux tels que définis par IHE (Integrating the Healthcare Enterprise). Pour cela, l'Etat doit s'appuyer sur les agences d'état (ASIP, ANAP, ATIH) et les collectifs d'industriels de promotions de standards (tels qu'InteropSanté, ou d'autres collectifs représentatifs). Ces profils doivent s'inscrire dans une feuille de route de l'intéropérabilité pluriannuelle offrant de la visibilité aux acteurs notamment industriels français et cette dernière doit s'accompagner de moyens (financiers, formations) pour accompagner la transition vers ces standards. Le calendrier doit être réaliste mais prévoir l'atteinte des objectifs à court (2 ans) et moyen terme (4 ans) ainsi que des bilans d'adoption.

- En particulier, des terminologies de référence internationales doivent être identifiées à l'instar des travaux très largement engagés dans le domaine de la biologie où l'utilisation du référentiel LOINC fait consensus. L'Etat doit également rendre opposable l'utilisation de terminologies cliniques dans l'ensemble des systèmes d'informations. Ces terminologies sont par exemple (car très utilisées) HPO et le thésaurus Orphanet pour les maladies rares, ou encore la CISP-2 pour le codage des données en soin primaire ou la CIM-O en cancérologie. Elles devront être accessibles dans le serveur de terminologies et alignées à la terminologie pivot choisie. Une politique incitative et des moyens d'accompagnement devront être prévus auprès des industriels et professionnels de santé. Les sociétés savantes et les collèges de professionnels devront être associés pour définir les cadres de recueil normalisés et standardisés et s'engager à leur promotion. Un label doit être défini également pour qualifier l'utilisabilité et l'ergonomie des progiciels médicaux.

- Intensifier l'utilisation de l'INS – Identifiant National de Santé - au sein des systèmes d'informations des établissements et de la médecine de ville pour que le chainage entre les différentes données du patient, quel que soit son contexte de prise en charge, puisse être possible et fiable.

- Rendre obligatoire et opposable sous deux ans la mise à disposition par les éditeurs de logiciels de santé de demi-connecteurs applicatifs d'interopérabilité, capables de mettre à disposition des clients (établissements et professionnels de santé) l'ensemble des données produites. Dans le cas contraire, rendre obligatoire le plein accès documenté.

L'ensemble de ces mesures nécessitera des moyens adaptés et un alignement des principales parties prenantes. Elles pourraient à ce titre faire partie des chantiers à inscrire dans la feuille de route de la future direction numérique du Ministère des Solidarités et de la Santé et de la Santé. Elles permettront d'améliorer la qualité des données et seront à ce titre bénéfiques à toutes les étapes de leur utilisation (pour le soin, la partage à l'échelle d'un territoire ou pour le DMP, pour la recherche, l'innovation, etc.).

Enfin l'alignement des gisements de données (entrepôts hospitaliers par exemple) et la promotion des terminologies dans le domaine de la santé publique (pour les registres par exemple) est un vaste sujet encore peu exploré. L'Etat peut contribuer à l'adoption de standards internationaux (tels que le format OMOP

largement adopté dans plusieurs pays pour l'utilisation des données rétrospectives). Les initiatives dans les domaines où cet effort est en cours doivent être en outre être soutenues et promues à l'échelle nationale et européenne, à l'instar de l'initiative OSIRIS qui vise à structurer et normaliser les données de vie réelle en cancérologie par l'adoption d'un format et de terminologies communes et à l'état de l'art.

6

STRUCTURE DU PROGRAMME ET FEUILLE DE ROUTE

6 STRUCTURE DU PROGRAMME ET FEUILLE DE ROUTE

Structure du programme

Un comité de pilotage rassemblant les différentes parties prenantes sera chargé de mettre en œuvre la feuille de route en vue de lancer publiquement l'offre de service du Health Data Hub d'ici fin 2019. Avec l'appui du Ministère des Solidarités et de la Santé, il veillera à la cohérence, et à l'identification de synergies, avec les autres grands programmes nationaux. On peut citer à titre d'exemple :

- Le Plan France Médecine Génomique 2025 ; le Dossier Médical Partagé (DMP) ; le 3e plan national maladies rares ; le plan santé ;
- Le Comité stratégique des Filières (CSF) et les grands défis notamment sur le champ de la santé ;
- Le Plan National pour la Science Ouverte ; les Instituts Interdisciplinaires d'Intelligence Artificielle (3IA) ; et les éventuels programmes sur le sujet de la formation en lien avec la stratégie IA ;
- Le programme Hop'EN pour le soutien à la transition numérique du Système de Santé ; et la stratégie de transformation du système de santé (STSS) ;
- La réflexion stratégique sur les "Registres et politiques de santé publique" menée par Aviesan.

Feuille de route

La mise en œuvre du Health Data Hub devra s'inscrire dans une logique de construction progressive, itérative, collaborative et agile. La mission propose en première approche le calendrier suivant :

- A fin 2018 : cadrage détaillé du modèle opérationnel du Hub (services, principes de collaboration, conditions générales d'utilisations, sécurité), identification des partenaires institutionnels, cadrage des projets expérimentaux à lancer, identification des bases de données clés à ingérer, cahier des charges de la plateforme ;
- A mi-2019 : lancement d'un produit minimum viable (MVP) de la plateforme avec des premiers utilisateurs « tests » issus des projets expérimentaux, mise en œuvre de la gouvernance définie, et ingestion des premières bases de données ;
- A fin 2019 : lancement de la première version de la plateforme et ouverture à tous de l'offre de services du Hub ;
- A fin 2020 : amélioration de l'organisation, des processus et des outils, enrichissement du catalogue de données et création des premiers Hub Locaux ;
- A fin 2021 : déploiement d'un réseau de Hub locaux sur l'ensemble du territoire.

Focus sur le quatrième trimestre 2018

Pour initier la mise en œuvre opérationnelle du Health Data Hub, une équipe de préfiguration sera montée, et devra notamment se concentrer sur un certain nombre de travaux préliminaires.

La mise en place de la structure juridique du Hub implique de rédiger une convention constitutive dans laquelle seront précisées les différentes parties prenantes. Les signataires de cette convention pourront notamment être : le Ministère des Solidarités et de la Santé et de la Santé, le Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, l'Assurance Maladie, un ou plusieurs organismes/opérateurs de recherche, des représentants des associations de patients, d'établissements et des professionnels de santé et des industriels. Cette convention pourrait être rédigée au mois de novembre pour être signée au mois de décembre. Le directeur de la structure devra également être choisi d'ici la fin de l'année. Le GIP devra pouvoir être opérationnel au début de l'année 2019. Parallèlement, les instances de gouvernance, notamment les divers comités (cf. Organisation, compétences et gouvernance du réseau Hub) pourront être préfigurés.

L'équipe de préfiguration devra, en avance de phase, veiller à ce que le délai nécessaire à la mise en place de la structure juridique ne se répercute pas sur l'avancée du projet. En particulier, des travaux de différentes natures pourront être enclenchés dès la fin des arbitrages faisant suite à la remise du rapport :

- L'évolution du Hub après sa création, plusieurs options sont possibles : le GIP pourrait être amené à intégrer l'Institut National des Données de Santé sous réserve d'une évolution législative ; ou évoluer vers un établissement public ou toutes sortes d'organisations permettant d'associer l'ensemble des partenaires mais offrant une certaine robustesse ;

- Toujours sur le plan juridique, les travaux en cours avec le Ministère des Solidarités et de la Santé pour élargir le SNDS aux autres sources de données de santé devront être suivis par cette équipe de préfiguration pour s'assurer que les appariements, notamment, entre sources de données pourront bien être réalisables sur un plan juridique lorsque le Hub sera opérationnel ;

- L'équipe de préfiguration devra boucler le financement du projet, notamment pour l'année 2019 et au-delà ;

- Afin que le Hub puisse être directement opérationnel à sa mise en place, l'équipe de préfiguration devra commencer à identifier l'équipe cible et les profils recherchés, préparer les fiches de poste et commencer à identifier les candidats potentiels pour les différentes fonctions envisagées ;

- L'équipe de préfiguration devra préparer le marché que passera le Hub à sa création afin de construire la première version de sa plateforme technologique pour mi 2019 ;

- Afin de pouvoir répondre courant 2019 aux besoins d'hébergement et de mise à disposition des « systèmes fils » du SNDS (extractions du SNDS soumises au référentiel de sécurité décrit dans l'arrêté du 22 mars 2017 et devant se mettre en conformité d'ici mars 2019), des travaux pourront

être menés avec la CNIL pour assurer une communication et une visibilité auprès de ces acteurs et leur assurer une possibilité d'hébergement avant la fin 2019 ;

- L'équipe de préfiguration devra également identifier les locaux du Health Data Hub, susceptibles de pouvoir accueillir les équipes mais également de réaliser ses activités d'animation de réseau ;

- Les attentes de l'écosystème étant élevées, s'agissant du Hub, un comité de pilotage devra être mis en place par la mission de préfiguration pour le suivi des travaux. Ce comité de pilotage pourra être composé des institutions qui ont fait connaître leur volonté d'être partenaire du Hub, ainsi que les autres partenaires potentiels qui seront identifiés dans les prochains mois ;

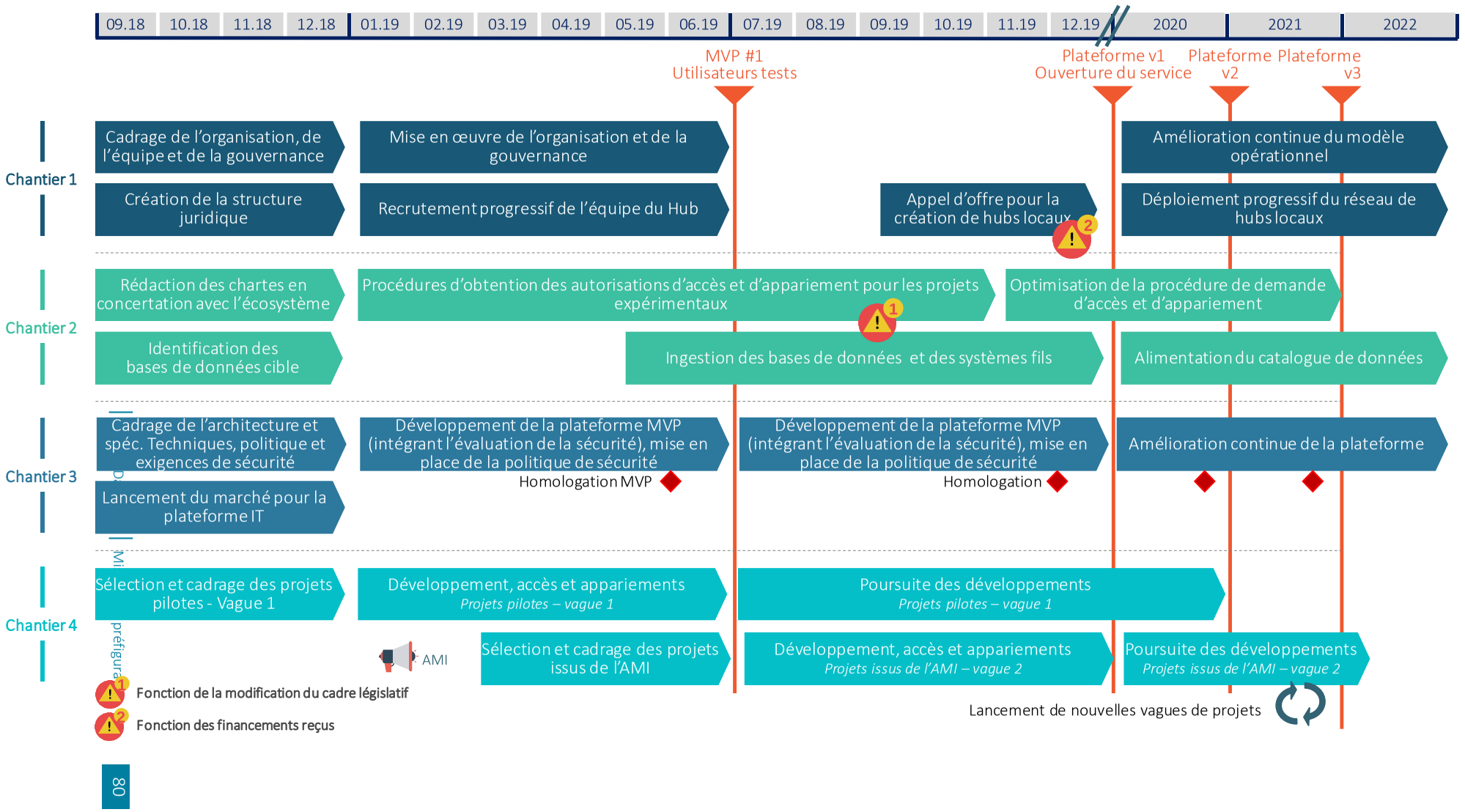
- Ce comité de pilotage pourra contribuer au choix des premiers projets « pilote » qui permettront de co-construire sur des cas concrets les services techniques et juridiques du Hub ; de poser les questions de la valorisation économique de ces services et de préciser les modalités d'utilisation du Hub ;

- Dès l'identification de ces projets pilote, des travaux devront être lancés avec la CNIL pour anticiper les appariements des sources de données qu'ils impliqueront ;

- Le comité de pilotage et les producteurs de données partenaires pourront poser les premières briques de la convergence des gouvernances d'accès aux sources de données et travailler de concert à l'élaboration de premières versions des chartes « producteurs » et « utilisateurs », ainsi que travailler à l'identification des ressources financières nécessaires pour la mise à disposition et la mise en qualité de certaines sources jugées de grand intérêt pour la communauté ;

- L'équipe de préfiguration devra également faire le lien avec les initiatives connexes : notamment la mise en place du programme du grand défi « IA et santé » dont le *program manager* devrait être identifié en novembre pour le lancement d'appels à projets en début d'année 2019. Ces appels à projets, qui ont vocation à faire émerger des champions nationaux dans le domaine, pourront contribuer au déploiement du Hub en encourageant le partage des données via le Hub et en alimentant la réflexion autour de modèles « type » de partage de la valeur. Ces défis peuvent être également l'occasion de favoriser des collaborations avec l'international. De manière similaire, des instituts 3IA spécialisés, entre autres, en santé et intelligence artificielle, seront sélectionnés par un jury international d'ici la fin de l'année. Ces instituts doivent travailler au plus près du Hub pour valoriser les données qui y seront mises à disposition et participer à la diffusion de travaux de recherche mobilisant ces données, ainsi qu'à l'effort de formation. Enfin, le plan France Médecine Génomique va lancer un marché en 2019 pour le CAD (collecteur analyseur de données), l'interopérabilité avec les deux infrastructures devrait être assurée ;

- L'équipe de préfiguration définira et mettre en œuvre un plan de communication auprès des acteurs pour garder informés les producteurs de données, utilisateurs et partenaires potentiels ; ainsi que le citoyen.


09.18
10.18
11.18
12.18
01.19
02.19
03.19
04.19
05.19
06.19
07.19
08.19
09.19
10.19
11.19
12.19
2020
2021
2022
MVP #1
Utilisateurs tests
Plateforme v1
Ouverture du service
Plateforme v2
Plateforme v3
Chantier 1
Cadrage de l'organisation, de l'équipe et de la gouvernance
Mise en œuvre de l'organisation et de la gouvernance
Amélioration continue du modèle opérationnel
Création de la structure juridique
Recrutement progressif de l'équipe du Hub
Appel d'offre pour la création de hubs locaux
Déploiement progressif du réseau de hubs locaux
Chantier 2
Rédaction des chartes en concertation avec l'écosystème
Procédures d'obtention des autorisations d'accès et d'appariement pour les projets expérimentaux
Optimisation de la procédure de demande d'accès et d'appariement
Identification des bases de données cible
Ingestion des bases de données et des systèmes fils
Alimentation du catalogue de données
Chantier 3
Cadrage de l'architecture et spéc. Techniques, politique et exigences de sécurité
Développement de la plateforme MVP (intégrant l'évaluation de la sécurité), mise en place de la politique de sécurité
Homologation MVP
Développement de la plateforme MVP (intégrant l'évaluation de la sécurité), mise en place de la politique de sécurité
Homologation
Amélioration continue de la plateforme
Lancement du marché pour la plateforme IT
Chantier 4
Sélection et cadrage des projets pilotes - Vague 1
Développement, accès et appariements
Projets pilotes – vague 1
Poursuite des développements
Projets pilotes – vague 1
AMI
Sélection et cadrage des projets issus de l'AMI
Développement, accès et appariements
Projets issus de l'AMI – vague 2
Poursuite des développements
Projets issus de l'AMI – vague 2
Lancement de nouvelles vagues de projets
1
Fonction de la modification du cadre législatif
2
Fonction des financements reçus

Chantier 1 - Offre de service et modèle opérationnel du Hub

Au cours du second semestre 2018, le modèle opérationnel et les principes de collaboration avec l'écosystème seront stabilisés (chartes, etc.). En parallèle, la création de la structure juridique devra être sécurisée. L'organisation du Hub, cadrée dès 2018, sera implémentée en 2019, avec la création des instances de gouvernance, la mise en place des processus, et le recrutement progressif des compétences clés. Cette organisation pourra ensuite évoluer, dans une logique d'amélioration continue. Dans une perspective d'ouverture vers l'Europe, les liens éventuels avec des initiatives étrangères similaires seront également étudiés dans le cadre de ce premier chantier.

Chantier 2 – Collecte et gouvernance de la donnée

Dans un premier temps, les procédures existantes d'accès aux données et d'appariement seront étudiées et lancées en vue d'obtenir les habilitations nécessaires pour la réalisation des premiers projets expérimentaux. Des pistes de simplification devront être rapidement identifiées en co-construction avec la CNIL, par exemple au travers de la création d'une méthodologie de référence ou par la modification du cadre législatif.

Les chartes (producteurs, utilisateurs, et citoyens) seront rédigées au cours du dernier trimestre 2018, en concertation avec l'écosystème, pour poser une gouvernance, des principes et des modalités juridiques pour régir le partage de données. Des travaux devront être prévus pour définir les indicateurs et standards à respecter en termes de qualité des données.

Enfin, après une phase de sélection et de qualification, les bases de données stratégiques pour le Hub, les bases de données requises pour les projets pilotes, ainsi que les système fils à héberger seront progressivement ingérés. Le catalogue de données sera ensuite alimenté en continu.

Chantier 3 - Architecture technologique et sécurité

D'ici la fin de l'année 2018, les fournisseurs de solutions technologiques devront être consultés. En parallèle, un travail de description de l'architecture cible et de spécifications techniques sera engagé pour lancer le marché et contractualiser avec le(s) partenaire(s) en charge du développement de la plateforme. Ce chantier doit intégrer dès le départ d'instruire les démarches à conduire pour atteindre le haut degré de sécurité visé et, au-delà des spécificités techniques, il doit être l'occasion de prévoir les différents jalons relatifs à l'évaluation de la sécurité de la plateforme (notamment l'analyse de risque, l'homologation mais également l'audit), ainsi que les mesures organisationnelles (sensibilisation, habilitation, etc.). L'ANSSI devrait être associée à ces travaux dès leur démarrage.

Pour accueillir les premiers utilisateurs dès juin 2019, un premier socle technologique offrant un nombre limité de fonctionnalités sera développé. Ce produit minimum viable (MVP) sera soumis aux tests d'aptitude et homologué. La première version de la plateforme intégrant l'ensemble des fonctionnalités prévues sera ensuite développée, sur la base du MVP et des retours utilisateurs, en vue d'un lancement du service fin 2019. De façon itérative, de nouvelles versions de la plateforme seront livrées entre 2020 et 2021.

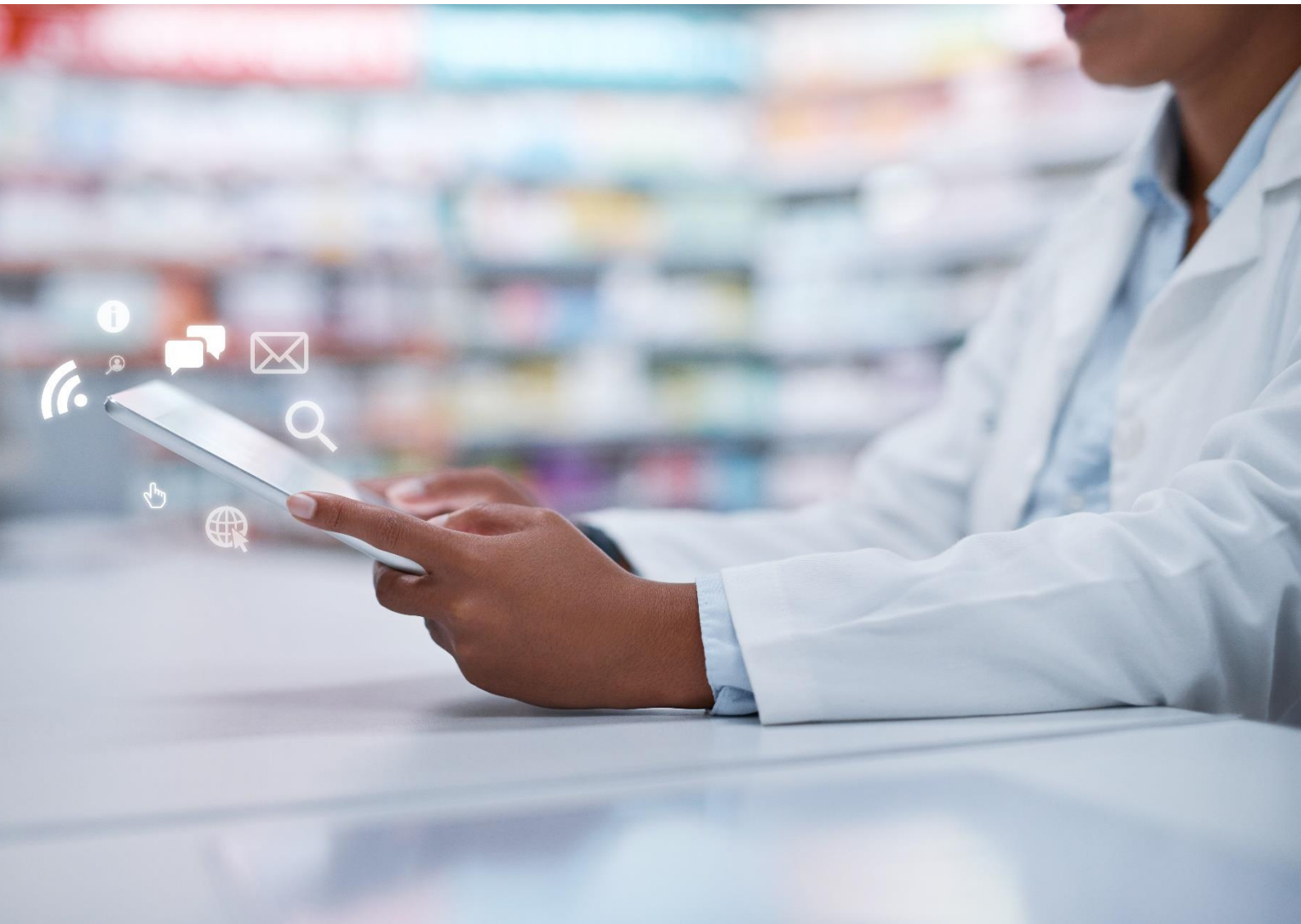
Chantier 4 - Projets expérimentaux

Après chacune des phases de sélection et de qualification, les projets seront cadrés de façon détaillée durant trois mois sur l'ensemble des dimensions suivantes :

- Usages : description détaillée des fonctionnalités et définition du *backlog*
- Sources de données : identification, qualification et appariement des données requises
- *Data science* : anticipation des analyses et des algorithmes à développer
- Ressources humaines et financières pour mener le projet
- Feuille de route détaillée

A la suite de ce travail et après avoir obtenu les habilitations nécessaires, les jeux de données ciblés seront collectés et préparés, et les équipes projets pourront bénéficier de l'accompagnement du Hub pour la réalisation des projets.

Dès la vague 2, des projets seront sélectionnés tous les six mois au travers d'Appels à Manifestation d'Intérêt au regard des critères détaillés dans la partie 4.4 Offre de service.



Moyens requis

Une première estimation du coût de développement et de fonctionnement du Health Data Hub a été réalisée sous l'hypothèse d'une montée en puissance progressive de la plateforme technologique, des équipes, et de la mise en place de hubs locaux sur le territoire.

Après le développement au premier semestre 2019 d'un premier MVP permettant de mettre à disposition les premières sources du catalogue auprès de 50 utilisateurs tests, la plateforme sera enrichie chaque année pour permettre aux 500 utilisateurs cibles d'accéder à une cinquantaine de bases de données et de disposer de capacités et d'outils de traitement à l'état de l'art. Les capacités de stockage et de calcul (CPU, GPU) seront scalables et pourront être augmentées à la demande pour répondre aux besoins des projets. Dans une logique de « service à la demande », cet incrément ferait l'objet d'une refacturation pour les acteurs du secteurs privés.

Le Health Data Hub s'appuiera fin 2018 et début 2019 sur une équipe de préfiguration afin de concevoir, de déployer, d'éprouver et d'améliorer l'organisation, les processus et outils mis en place. La mise en place de la première version de l'infrastructure sécurisée sera un point central de la feuille de route avec les contraintes juridiques entourant la mise en conformité au référentiel de sécurité des 200 systèmes fils. La sélection et le cadrage des projets pilotes seront également en tête de la liste des priorités. En cible, cette équipe sera constituée d'une trentaine de personnes, rassemblant des architectes techniques, des data engineers, des ressources en charge du pilotage du programme et de la conduite du changement, ainsi que des experts sécurité, juridiques et médicaux. Dès l'ouverture du Hub au second semestre 2019 et tous les six mois, trois à cinq projets seront accompagnés par une équipe dédiée ou au moyen d'un financement, afin de dynamiser les usages autour des évolutions du catalogue, et de contribuer à des enjeux de santé publique, de recherche scientifique et d'innovation industrielle. Après un temps d'apprentissage et de consolidation, le réseau des hubs locaux se développera petit à petit.

Le Health Data Hub devra donc se doter d'un budget annuel de près de 40 millions d'euros pour atteindre le niveau d'ambition visé. Ces chiffrages reposent notamment sur l'hypothèse d'une équipe d'une trentaine de personnes dans le hub central et une quinzaine dans les hub locaux[34], correspondant au scénario budgétaire médian. Ces chiffres peuvent être mis en regard des 29 et 23 personnes respectivement prévues pour l'intégration et la réutilisation des données d'entrepôts à l'université d'Erlangen (Allemagne) et Dundee (Royaume-Uni). Le chiffrage proposé repose sur l'hypothèse de cinq hubs locaux mis en place d'ici 2022.

Les hypothèses de chiffrage relatives à la cybersécurité s'appuient sur les textes relatifs à l'hébergement des données de santé et sur le référentiel de sécurité relatif au SNDS. Si le Hub était désigné opérateur de service

[34] Par exemple dans le Hub central : 6 architectes, 10 data managers, 3 experts sécurité, 3 experts médicaux, 3 experts juridiques, 2 experts valorisation, 5 personnes pour le pilotage.

essentiel, il devrait également respecter les règles de sécurité associées dont l'écart avec celles prévues par les textes précédents doit être évalué et chiffré, pouvant entraîner une modification des montants estimés.

Poste de coûts (en M€)	T4 2018	2019	2020	2021	2022
Nombre de hubs locaux déployés	0	0	2	4	5
1. Plateforme technologique	0,0	2,8	5,0	6,2	6,4
2. Compétences et locaux - Hub central	0,3	5,0	3,7	3,4	3,4
2.1 Equipe technique, sécurité et data management	0,1	2,9	2,0	1,7	1,7
2.2 Equipe Expertise Juridique	0,0	0,2	0,2	0,2	0,2
2.3 Equipe Expertise Médicale	0,0	0,1	0,3	0,3	0,3
2.4 Equipe Pilotage Programme, Valorisation et Change	0,2	1,6	1,0	1,0	1,0
2.5 Locaux et hébergement	0,0	0,1	0,2	0,2	0,2
3. Compétences et locaux - Hubs locaux	0,0	0,0	2,0	4,6	6,5
3.1 Equipe technique, sécurité et data management	0,0	0,0	1,1	2,5	3,5
3.2 Equipe Expertise Juridique	0,0	0,0	0,2	0,4	0,5
3.3 Equipe Expertise Médicale	0,0	0,0	0,2	0,4	0,6
3.4 Equipe Pilotage Programme, Valorisation et Change	0,0	0,0	0,5	1,1	1,5
3.5 Locaux et hébergement	0,0	0,0	0,1	0,2	0,3
4. Accompagnement des projets	0,0	2,4	6,6	11,4	11,4
4.1 Hub central	0,0	2,4	4,8	5,4	5,4
4.2 Hubs locaux	0,0	0,0	1,8	6,0	6,0
5 Dotations aux producteurs de données	0,0	8,0	10,0	12,0	12,0
Total	0,3	18,2	27,3	37,6	39,7

ANNEXES

1 LE SYSTEME NATIONAL DES DONNEES DE SANTE (SNDS)

La France a mis en œuvre, dans les années quatre-vingt-dix, un projet ambitieux de constitution d'un entrepôt de données « médico-administratives », c'est-à-dire collectées à des fins de gestion : ce système national d'information inter-régimes d'assurance maladie (SNIIRAM), qui contient le détail de toutes les feuilles de soins, a été très vite chaîné avec les données de facturation hospitalière (le PMSI) et permet ainsi de retracer, de manière détaillée et exhaustive, les parcours de soins de la totalité de la population française, soit 67 millions de personnes, avec des données pseudonymisées.

Ces données, avec aujourd'hui un historique de plus de dix ans, constituent un patrimoine remarquable : elles couvrent une large population, offrent des possibilités de suivi sur une longue période, sans perdu de vue en cours de suivi, avec une assez bonne homogénéité de codage. En outre, si des investissements importants ont été consentis pour constituer l'entrepôt, elles ont aujourd'hui un faible coût de collecte, puisqu'étant obtenues en sous-produit d'un système de gestion.

Si ce système d'information a été progressivement de plus en plus utilisé, au-delà de l'assurance maladie, par des équipes de recherche et par des agences sanitaires (par exemple Santé publique France, l'Agence nationale de sécurité du médicament), le constat a été fait au début des années 2010 que les accès restaient trop limités pour exploiter tout le potentiel de ces données très riche.

À l'issue de quelques années de débats publics et de réflexions, la Loi de modernisation de notre système de santé du 26 janvier 2016 a réorganisé la gouvernance de ces bases de données, clarifié la doctrine et remis à plat le dispositif d'accès.

La base de données

Le Système national des données de santé (SNDS) a été instauré par l'article 193 de la LMSS. Il prévoit de chaîner dans ce système d'information cinq flux de données :

- les données de remboursement de l'Assurance Maladie (SNIIRAM),
- les données des établissements de santé (PMSI),
- les causes médicales de décès,
- les données relatives au handicap en provenance des Maisons départementales des personnes handicapées,
- un échantillon représentatif des données de remboursement des organismes d'assurance maladie complémentaire.

La première version du SNDS était constituée du SNIIRAM et du PMSI. Les causes médicales de décès ont commencé à être intégrées au dernier trimestre de 2017, et les deux derniers flux le seront à l'horizon 2019-2020. Toutes ces données ont vocation à être chaînées au niveau individuel, mais pour garantir et protéger la confidentialité de ces données, un pseudonyme, code non signifiant obtenu par un procédé cryptographique irréversible du NIR, est associé aux données se rapportant à chaque personne. Ce procédé permet de relier, pour une même personne, l'ensemble des données contenues dans le SNDS. Il permet également d'apparier ces données avec des informations figurant dans d'autres systèmes, avec l'autorisation de la CNIL.

Le ministère chargé de la santé pilote les orientations stratégiques de ce projet, qui est géré opérationnellement par la Caisse nationale d'assurance maladie (CNAM).

Les finalités

Les finalités prévues par la loi sont :

- L'information sur la santé ainsi que sur l'offre de soins, la prise en charge médico-sociale et leur qualité,
- la connaissance des dépenses de santé, des dépenses d'assurance maladie et des dépenses médico-sociales,
- l'information des professionnels, des structures et des établissements de santé ou médico-sociaux sur leur activité,
- la surveillance, la veille et la sécurité sanitaires,
- la recherche, les études, à l'évaluation et l'innovation dans les domaines de la santé et de la prise en charge médico-sociale.

Les principes et conditions d'accès aux données

Tous les acteurs publics et privés peuvent accéder aux données du SNDS.

Les jeux de données rendus totalement anonymes (au sens où toutes les possibilités de réidentification ont été éliminées, parce que ce sont des données agrégées ou appauvries) ont vocation à être en open data, c'est-à-dire ouverts à tous, avec possibilité de réutilisation, sans nécessité d'autorisation préalable.

Pour les données présentant un risque de ré-identification, la loi a prévu :

- pour des organismes publics ou chargés d'une mission de service public, des accès permanents dont le périmètre est fixé par un décret en conseil d'Etat[35],
- pour les autres utilisateurs des données ou pour des besoins excédant le périmètre de ces accès permanents, des accès autorisés par la CNIL pour des projets spécifiques.

Les accès aux données ne peuvent être autorisés que pour permettre des traitements à des fins de recherche, d'étude ou d'évaluation contribuant à l'une des finalités énumérées dans la loi (cf ci-dessus), et sous réserve de plusieurs conditions :

1. La recherche, étude ou évaluation doit présenter un caractère d'intérêt public. L'Institut national des données de santé est chargé par la loi de se prononcer sur ce caractère d'intérêt public des études, lorsqu'il est saisi, et pour ce faire il a constitué en son sein un comité d'expertise sur l'intérêt public.

2. Certaines finalités sont par ailleurs interdites : la loi précise que « les données du système national des données de santé ne peuvent être traitées pour l'une des finalités suivantes : « 1° La promotion des produits mentionnés au II de l'article L. 5311-1 en direction des professionnels de santé ou d'établissements de santé ; « 2° L'exclusion de garanties des contrats d'assurance et la modification de cotisations ou de primes d'assurance d'un individu ou d'un groupe d'individus présentant un même risque. ».

3. Des garanties doivent être apportées sur le respect de la vie privée des citoyens. Les données du SNDS sont en effet des données sensibles à caractère personnel qui exigent un haut niveau de sécurité. La loi précise : « L'accès aux données s'effectue dans des conditions assurant la confidentialité et l'intégrité des données et la traçabilité des accès et des autres traitements, conformément à un référentiel défini par arrêté des ministres chargés de la santé, de la sécurité sociale et du numérique, pris après avis de la Commission nationale de l'informatique et des libertés ». L'arrêté du 22 mars 2017 relatif au référentiel de sécurité applicable au SNDS précise les exigences de sécurité en vigueur. Elles s'appliquent à la CNAM, responsable du traitement SNDS et chargé à ce titre de protéger les données de la base, aux gestionnaires des bases composantes du

[35] Décret no 2016-1871 du 26 décembre 2016 relatif au traitement de données à caractère personnel dénommé « système national des données de santé »

SNDS, aux gestionnaires des bases hébergeant des données du SNDS du fait d'une autorisation CNIL et aux utilisateurs du SNDS. Les grands principes de sécurité portés par ce référentiel sont la pseudonymisation des données, l'authentification, la traçabilité des accès et des traitements, le contrôle et les sanctions assorties, notamment pénales (des audits sont prévus), la sensibilisation et la formation obligatoire des utilisateurs.

4. La contrepartie d'un accès large est une obligation de transparence, de façon à rendre compte de l'utilisation de ce patrimoine au citoyen et de partager des éléments de connaissance issus de l'exploitation de ces bases.

Les procédures d'accès aux données

Dans le cas d'un accès sur projet, les textes prévoient une procédure classique d'accès aux données constituée de trois ou quatre étapes.

L'Institut national des données de santé (INDS) est désormais la porte d'entrée pour toutes ces demandes d'accès sur projet à des bases de données déjà constituées, notamment celles du système national de données de santé (SNDS).

L'INDS transmet les demandes à un comité d'experts (le CEREES, comité d'expertise pour les recherches, les études et les évaluations dans le domaine de la santé) qui émet un avis sur la méthodologie retenue, sur la nécessité du recours à des données à caractère personnel, sur la pertinence de celles-ci par rapport à la finalité du traitement et, s'il y a lieu, sur la qualité scientifique du projet.

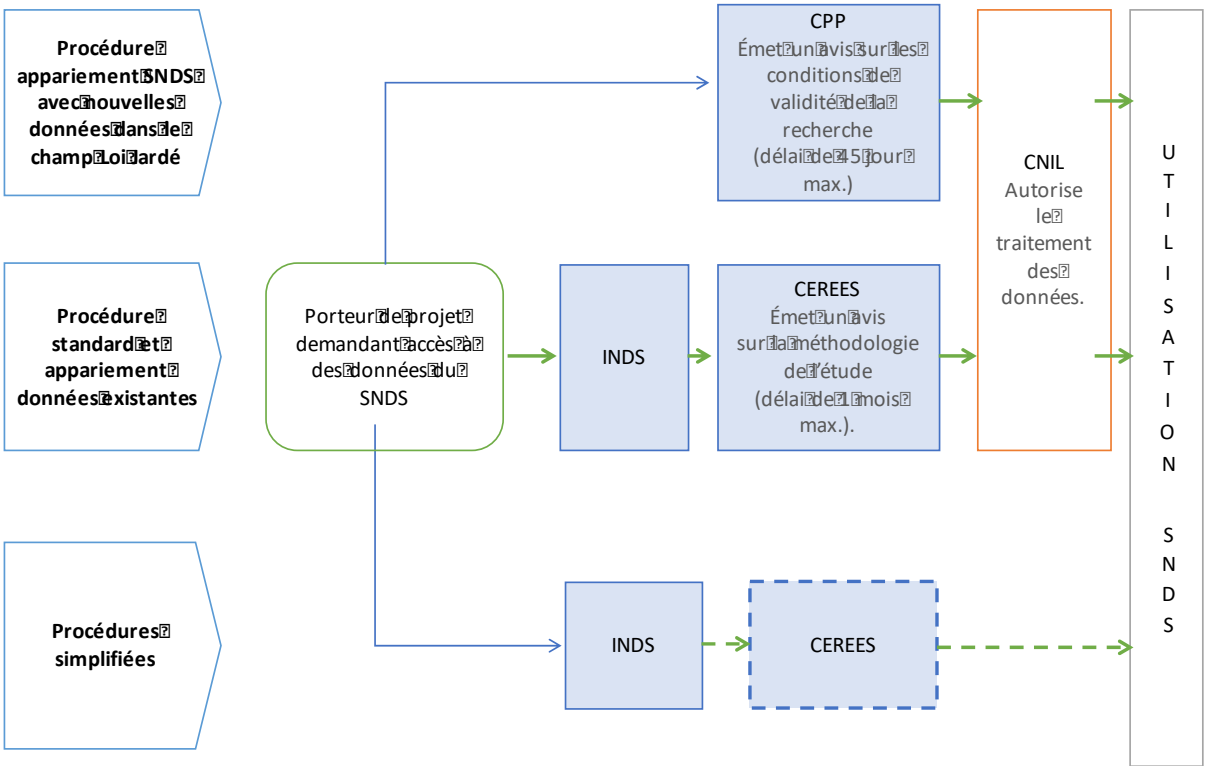
Le CEREES dispose d'un mois pour émettre un avis qui est transmis à la CNIL, laquelle autorise ou non le traitement.

La CNIL peut saisir l'INDS pour qu'il rende un avis sur le critère d'intérêt public. L'INDS peut être également saisi par la ministre chargée de la santé ou s'autosaisir, et doit rendre cet avis dans le délai d'un mois.

La CNIL dispose d'un délai de deux mois, renouvelable une fois, pour se prononcer. Au terme de ce délai , si la CNIL ne s'est pas prononcée, l'avis est réputé favorable.

Le circuit est différent si la demande fait intervenir un recueil de données auprès des personnes, avec leur consentement. La CNIL prend dans ce cas sa décision après avis d'un comité de protection des personnes (CPP), et ce y compris lorsque le projet prévoit un appariement de ces données nouvellement collectées avec les données du SNDS.

Le schéma ci-dessous récapitule ainsi les procédures d'accès aux données sur projet :



La loi prévoit également que des procédures d'accès simplifiées puissent être homologuées par la CNIL : décisions uniques permettant à un demandeur de réaliser un ensemble de traitements, méthodologies de référence, mises à disposition simplifiées de jeux de données agrégées ou d'échantillons.

Plusieurs décisions uniques ont été publiées, ainsi que deux méthodologies de référence pour l'accès au PMSI et une procédure simplifiée pour l'accès à l'échantillon généraliste de bénéficiaires (échantillon au 1/100ème extrait du SNIIRAM et du PMSI).

Le rôle de l'INDS

La loi indique que l'INDS a notamment pour rôle de « veiller à la qualité des données de santé et aux conditions générales de leur mise à disposition, garantissant leur sécurité et facilitant leur utilisation ».

L'objectif principal est bien en effet de faciliter les accès aux données : l'INDS est la porte d'entrée et l'interlocuteur pour toutes les demandes d'accès sur projet à des bases de données déjà constituées, notamment celles du système national de données de santé (SNDS), il a un rôle de conseil et d'accompagnement pour constituer les dossiers. Au-delà de ce rôle dans les procédures d'accès, il a pour mission de favoriser le dialogue entre les producteurs et les utilisateurs, qui sont tous membres du

groupement d'intérêt public[36], afin de faire progresser l'ensemble du système pour répondre aux besoins de tous, en s'appuyant sur les retours d'expérience. Ces progrès peuvent porter sur l'enrichissement des bases, la constitution de nouveaux jeux de données, la qualité des données, leur documentation, les conditions de mise à disposition, ... Il est force de proposition pour élaborer avec la CNIL les procédures simplifiées qui permettront des accès aux données plus rapides et dans le cadre de ces procédures, en fonction des conditions homologuées par la CNIL, il peut jouer un rôle dans la mise à disposition des jeux de données agrégées ou d'échantillons (c'est le cas aujourd'hui par exemple pour des accès rapides à l'échantillon généraliste de bénéficiaires, sans passage par le CEREES et la CNIL, pour certains types d'études). Il contribue aussi à l'expression des besoins en matière de données anonymes (open data).

Bilan à 1 an

Le nouveau dispositif d'accès a été mis en place le 28 août 2017. Les dossiers de demandes sont déposés de manière dématérialisée sur le site de l'INDS, qui ensuite se charge de l'adresser au CEREES puis à la CNIL.

Sur les 500 demandes d'accès à des bases de données[37], 168 demandes d'accès à des données du SNDS ont été déposées. 118 nécessitent une extraction ad hoc de la base complète, 20 utilisent l'échantillon généraliste de bénéficiaires (échantillon au 1/100ème de la base totale) ; certaines études ne requièrent que des données du PMSI seul (25) ou ne portent que sur les causes de décès (5).

Les dossiers déposés par les industriels et bureaux d'étude représentent près de 40% des demandes déposées, ce qui témoigne d'une réelle ouverture de ces données au secteur privé.

Sur les 168 demandes d'accès au SNDS, 105 ont été transmises à la CNIL, dont 68 ont fait l'objet d'une autorisation.

Le délai médian entre le dépôt du dossier et l'obtention de l'autorisation CNIL est de 70 jours ouvrés (minimum 34, maximum 177).

Pour les dossiers en attente d'une autorisation de la CNIL, le délai moyen depuis leur dépôt à la Commission est de 62 jours ouvrés (avec un maximum de 119 jours). Un facteur d'allongement des délais est notamment la méconnaissance des utilisateurs des nouvelles conditions de mise à disposition des données (qui ne peuvent plus être exploitées que dans des environnements sécurisés respectant les exigences du référentiel de sécurité).

[36] L'INDS est un GIP qui rassemble l'Etat, les autorités et agences sanitaires, l'assurance maladie, les professionnels et établissements de soins, les usagers, les organismes d'assurance complémentaire, les organismes de recherche, les industriels, les start-ups et les bureaux d'étude.

[37] La moitié des dossiers concerne des études sur des collections de dossiers médicaux nécessitant un accord CNIL pour des raisons de dérogation à l'information du patient (ceux qui n'ont pas besoin d'avoir cette dérogation réalisent les recherches dans le cadre des méthodologies de référence).

Dossiers de demandes d'accès reçus par l'INDS au 14/09/2018 – Répartition par catégorie d'utilisateurs et par source des données

	Extraction SNDS	Echantillon généraliste de bénéficiaires	PMSI seul	Causes de décès seules	TOTAL
INSERM-CNRS et autres EPST	6		2		**8**
CHU- centres anticancéreux	37	2	3	1	**43**
Universités/Ecoles et autres	8	1	1		**10**
Total recherche publique	**51**	**3**	**6**	**1**	**61**
Industriels de santé	29	12	0	1	**42**
Bureaux d'étude	11	2	7		**20**
Autres (Ministère, agences, collectivités locales, associations…)	27	3	12	3	**45**
Total hors recherche	**67**	**17**	**19**	**4**	**107**
TOTAL	**118**	**20**	**25**	**5**	**168**

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Ministère des solidarités et de la santé
Ministère du travail
Ministère de l'action et des comptes publics

Direction de la recherche, des études, de l'évaluation et des statistiques
Le directeur

Paris, le 4 mai 2018
DREES-DIR N° 15

Note à Madame la Ministre des Solidarités et de la Santé, sous couvert de Monsieur Nicolas Labrune, conseiller

Vous trouverez ci joint un projet de lettre de mission et une présentation du projet de mise en œuvre des annonces du président de la République sur l'intelligence artificielle dans le domaine de la Santé.

Suite à des discussions avec votre cabinet, celui la ministre de l'enseignement supérieur, de la recherche et de l'innovation, celui du Premier Ministre et celui du Président de la République, nous proposons la création d'un groupe de travail avec un double objectif.

Le premier objectif serait de proposer une feuille de route pour les 4 prochaines années portant sur le rapprochement des données cliniques et des données médico-administratives du SNDS. L'objectif serait de prioriser et de planifier les appariements mais aussi de définir les moyens nécessaires. La question de l'émergence d'une équipe capable d'industrialiser les appariements se pose notamment.

Le deuxième objectif serait la préfiguration du Hub. Le groupe de travail définirait les actions clefs qui devront être réalisés au cours des premières années de l'existence du Hub. Il proposerait aussi une gouvernance et identifierait les besoins en termes de moyens.

Ce groupe de travail serait piloté par trois experts, Madame Dominique Polton, présidente de l'Institut National de Données en Santé, le professeur Marc Cuggia, praticien hospitalier à Rennes, spécialiste des entrepôts hospitaliers et recommandé par le cabinet du Premier Ministre, et Monsieur Gilles Wainrib, président fondateur de la Start-up OWKIN, start up reconnue dans le domaine des données de santé. Ce groupe de travail pourrait comprendre une douzaine de membres : des membres des trois centres de recherche CNRS, INRIA et INSERM, conformément à l'annonce du Président de la République, ainsi que des représentants de la CNAM, du ministère de la Recherche, de l'APHP, un autre représentant de start-up, un représentant des industries de santé et deux représentants du ministère (DREES, DSSIS). La DREES mettrait par ailleurs à la disposition de ce groupe de travail une cheffe de bureau comme rapporteur.

Parallèlement, la DREES avec le soutien des directions du ministère et de la CNAMTS sécurisera le financement du HUB pour la période 2019-2022 et préparera les dispositions juridiques nécessaires, notamment la transcription législative de l'élargissement du SNDS à l'ensemble des données obtenues au cours de soins remboursées par les régimes obligatoires de sécurité sociale.

Si ce projet vous agrée, nous vous transmettrons les trois lettres de mission pour les copilotes et nous organiserons une première séance du groupe de travail autour du 20 mai.

Le directeur de la recherche, des études, de l'évaluation et des statistiques

Jean-Marc AUBERT

Adresse postale : 14 avenue Duquesne - 75350 PARIS 07 SP
Adresse visiteurs : 10 - 18, place des Cinq-Martyrs-du-Lycée-Buffon - 75014 PARIS
http://social-sante.gouv.fr/

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE DES SOLIDARITÉS ET DE LA SANTÉ

La Ministre

Paris, le 16.5.2018

Pég – D- 18-012185

Madame la Présidente, Chère Dominique,

Le Système national des données de santé (SNDS) constitue une des bases de données de santé les plus riches d'Europe.

Le potentiel de ces données est considérable. Elles peuvent être mobilisées pour la recherche clinique et en matière de santé publique, l'aide à la décision des professionnels de santé, l'allocation des financements aux offreurs de santé, le développement d'une médecine prédictive, préventive, personnalisée et participative, ou encore éclairer la transformation et le pilotage du système de santé. Cette liste n'est d'ailleurs pas exhaustive.

Malgré l'ouverture des données prévues par la loi de modernisation du système de santé de janvier 2016, les données du SNDS demeurent sous-exploitées pour deux raisons principales. Premièrement, le périmètre du SNDS est encore limité essentiellement à des données médico-administratives. Son appariement régulier avec des données cliniques apparaît indispensable. Par ailleurs, compte-tenu de la spécificité de ses données, notamment leur volumétrie et leur complexité, un simple droit d'accès ne peut qu'être une première étape pour inciter à leur usage. Il semble nécessaire de mettre à la disposition des chercheurs, des administrations et des acteurs privés, notamment des start-up, les moyens facilitant leur usage, tant en termes d'infrastructure sécurisée et de logiciels que de services de facilitation. Il semble aussi important d'animer la communauté des utilisateurs pour faciliter les partages d'expérience, de compétences et d'outils. Cette mutualisation permettra de révéler le potentiel du SNDS et d'industrialiser son usage.

C'est en ce sens que le Président de la République a annoncé, le 29 mars 2018, au sommet Intelligence artificielle « AI for Humanity », la création d'un Hub. Structure partenariale entre producteurs et utilisateurs des données, elle pilotera l'enrichissement continu mais aussi la valorisation du SNDS, pour y inclure, à terme, l'ensemble des données provenant des remboursements de l'Assurance-maladie, en ajoutant les données cliniques des hôpitaux, de la médecine de ville, des autres producteurs de soins, ainsi que les données de grande qualité, scientifique et médicale, créées dans le cadre de cohortes nationales. Ce « Health Data hub » permettra, dans un cadre parfaitement sécurisé, d'avoir, entre autres, un espace de travail pour l'intelligence artificielle, catalyseur d'innovations. Il associera les grands organismes de recherche français, l'INSERM, le CNRS et l'INRIA, ainsi que de nombreux partenaires publics et privés.

Madame Dominique POLTON
Présidente de l'Institut national des données de santé
19 rue Arthur Croquette
94220 Charenton-le-Pont

J'ai souhaité vous confier la mission de co-piloter un groupe de travail dédié à la préfiguration de ce Hub, compte tenu de votre expérience et de vos compétences. Vous serez associé à deux copilotes, M. Marc Cuggia et M. Gilles Wainrib.

L'objectif de ce groupe de travail est double : préparer la création du Hub et proposer, avec des partenaires, une feuille de route pour l'enrichissement du SNDS.

En ce qui concerne le Hub, le groupe de travail s'attachera à proposer une organisation pour cette alliance ainsi que la définition de la plateforme des données et de l'offre de service qu'elle fournirait et la manière dont elle faciliterait le développement de l'écosystème (acteurs publics, recherche, start-up et autres acteurs privés...). Le groupe de travail déterminera les besoins en nouvelles infrastructures en tenant compte des infrastructures existantes, de leur évolution et des demandes exprimées par les acteurs. Ces nouvelles infrastructures n'ont pas vocation à être exclusives, ni à concurrencer les portails existants ou systèmes d'information des producteurs de données, mais bien à venir enrichir l'offre à destination de l'utilisateur de la donnée. Un cahier des charges, un modèle économique ainsi qu'une réflexion sur l'administration de cette plateforme en régime permanent devront être proposés. Au-delà de la plateforme des données, le groupe de travail devra étudier les moyens à mettre en œuvre par le Hub pour développer les compétences, promouvoir leur partage, ainsi que permettre le développement de partenariats entre les acteurs pour accroître fortement l'usage des données de santé en France. Les modalités de l'action du Hub, l'éventail d'une offre de service initiale et les outils logiciels à faire développer ou à rassembler dans les premiers trimestres devront notamment être définis.

S'agissant de l'enrichissement du SNDS, le groupe de travail s'attachera à identifier les premières sources de données pouvant être fédérées. Des projets autour de leur exploitation pourront être par ailleurs prévus, par exemple sur des aires thérapeutiques à fort enjeu de santé publique. Il s'agira aussi de déterminer les moyens de toute nature nécessaires au développement de ces appariements au cours des prochaines années, notamment pour inciter les acteurs à se lancer dans de telles opérations et industrialiser le processus d'appariement.

A partir de ces éléments, le groupe de travail s'attachera à préciser les missions, le rôle, le périmètre, l'équipe nécessaire à la mise en œuvre, le budget sur les 3 prochaines années et la gouvernance du Hub.

Vous m'adresserez un rapport, d'ici le 21 septembre 2018, décrivant de manière précise les plans d'action proposés pour une mise en œuvre opérationnelle permettant la réalisation des premiers travaux et événements du Hub dès 2019.

Vous travaillerez en lien étroit avec des représentants d'instituts de recherche en santé et dans le domaine de l'intelligence artificielle, des membres de l'écosystème des start-up, des professionnels de santé, des professionnels hospitaliers, un membre du LEEM et un représentant de patients. Vous associerez l'assurance maladie et des représentants du ministère de la Santé et de la Recherche. La DREES et la DSIS vous apporteront leur soutien.

Je sais pouvoir compter sur votre mobilisation pour animer ce chantier en faveur d'une accélération de l'innovation, grâce aux données et dans un cadre parfaitement sécurisé, garantissant l'anonymat et le respect de la vie privée de chaque individu.

Je vous prie de croire, Madame, à l'expression de ma considération distinguée.

Bien amicalement

Agnès BUZYN

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
MINISTÈRE DES SOLIDARITÉS ET DE LA SANTÉ

La Ministre

Paris, le 16.5.2018

Pég. – D-18-012185

Monsieur le professeur,

Le Système national des données de santé (SNDS) constitue une des bases de données de santé les plus riches d'Europe.

Le potentiel de ces données est considérable. Elles peuvent être mobilisées pour la recherche clinique et en matière de santé publique, l'aide à la décision des professionnels de santé, l'allocation des financements aux offreurs de santé, le développement d'une médecine prédictive, préventive, personnalisée et participative, ou encore éclairer la transformation et le pilotage du système de santé. Cette liste n'est d'ailleurs pas exhaustive.

Malgré l'ouverture des données prévues par la loi de modernisation du système de santé de janvier 2016, les données du SNDS demeurent sous-exploitées pour deux raisons principales. Premièrement, le périmètre du SNDS est encore limité essentiellement à des données médico-administratives. Son appariement régulier avec des données cliniques apparaît indispensable. Par ailleurs, compte-tenu de la spécificité de ses données, notamment leur volumétrie et leur complexité, un simple droit d'accès ne peut qu'être une première étape pour inciter à leur usage. Il semble nécessaire de mettre à la disposition des chercheurs, des administrations et des acteurs privés, notamment des start-up, les moyens facilitant leur usage, tant en termes d'infrastructure sécurisée et de logiciels que de services de facilitation. Il semble aussi important d'animer la communauté des utilisateurs pour faciliter les partages d'expérience, de compétences et d'outils. Cette mutualisation permettra de révéler le potentiel du SNDS et d'industrialiser son usage.

C'est en ce sens que le Président de la République a annoncé, le 29 mars 2018, au sommet Intelligence artificielle « AI for Humanity », la création d'un Hub. Structure partenariale entre producteurs et utilisateurs des données, elle pilotera l'enrichissement continu mais aussi la valorisation du SNDS, pour y inclure, à terme, l'ensemble des données provenant des remboursements de l'Assurance-maladie, en ajoutant les données cliniques des hôpitaux, de la médecine de ville, des autres producteurs de soins, ainsi que les données de grande qualité, scientifique et médicale, créées dans le cadre de cohortes nationales. Ce « Health Data hub » permettra, dans un cadre parfaitement sécurisé, d'avoir, entre autres, un espace de travail pour l'intelligence artificielle, catalyseur d'innovations. Il associera les grands organismes de recherche français, l'INSERM, le CNRS et l'INRIA, ainsi que de nombreux partenaires publics et privés.

Professeur Marc CUGGIA
Faculté de Médecine
2 avenue du professeur Léon Bernard – CS 34317
35043 Rennes Cedex

J'ai souhaité vous confier la mission de co-piloter un groupe de travail dédié à la préfiguration de ce Hub, compte tenu de votre expérience et de vos compétences. Vous serez associé à deux copilotes, Mme Dominique Polton et M. Gilles Wainrib.

L'objectif de ce groupe de travail est double : préparer la création du Hub et proposer, avec des partenaires, une feuille de route pour l'enrichissement du SNDS.

En ce qui concerne le Hub, le groupe de travail s'attachera à proposer une organisation pour cette alliance ainsi que la définition de la plateforme des données et de l'offre de service qu'elle fournirait et la manière dont elle faciliterait le développement de l'écosystème (acteurs publics, recherche, start-up et autres acteurs privés...). Le groupe de travail déterminera les besoins en nouvelles infrastructures en tenant compte des infrastructures existantes, de leur évolution et des demandes exprimées par les acteurs. Ces nouvelles infrastructures n'ont pas vocation à être exclusives, ni à concurrencer les portails existants ou systèmes d'information des producteurs de données, mais bien à venir enrichir l'offre à destination de l'utilisateur de la donnée. Un cahier des charges, un modèle économique ainsi qu'une réflexion sur l'administration de cette plateforme en régime permanent devront être proposés. Au-delà de la plateforme des données, le groupe de travail devra étudier les moyens à mettre en œuvre par le Hub pour développer les compétences, promouvoir leur partage, ainsi que permettre le développement de partenariats entre les acteurs pour accroître fortement l'usage des données de santé en France. Les modalités de l'action du Hub, l'éventail d'une offre de service initiale et les outils logiciels à faire développer ou à rassembler dans les premiers trimestres devront notamment être définis.

S'agissant de l'enrichissement du SNDS, le groupe de travail s'attachera à identifier les premières sources de données pouvant être fédérées. Des projets autour de leur exploitation pourront être par ailleurs prévus, par exemple sur des aires thérapeutiques à fort enjeu de santé publique. Il s'agira aussi de déterminer les moyens de toute nature nécessaires au développement de ces appariements au cours des prochaines années, notamment pour inciter les acteurs à se lancer dans de telles opérations et industrialiser le processus d'appariement.

A partir de ces éléments, le groupe de travail s'attachera à préciser les missions, le rôle, le périmètre, l'équipe nécessaire à la mise en œuvre, le budget sur les 3 prochaines années et la gouvernance du Hub.

Vous m'adresserez un rapport, d'ici le 21 septembre 2018, décrivant de manière précise les plans d'action proposés pour une mise en œuvre opérationnelle permettant la réalisation des premiers travaux et événements du Hub dès 2019.

Vous travaillerez en lien étroit avec des représentants d'instituts de recherche en santé et dans le domaine de l'intelligence artificielle, des membres de l'écosystème des start-up, des professionnels de santé, des professionnels hospitaliers, un membre du LEEM et un représentant de patients. Vous associerez l'assurance maladie et des représentants du ministère de la Santé et de la Recherche. La DREES et la DSIS vous apporteront leur soutien.

Je sais pouvoir compter sur votre mobilisation pour animer ce chantier en faveur d'une accélération de l'innovation, grâce aux données et dans un cadre parfaitement sécurisé, garantissant l'anonymat et le respect de la vie privée de chaque individu.

Je vous prie de croire, Monsieur le professeur, à l'expression de ma considération distinguée.

Agnès BUZYN

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE DES SOLIDARITÉS ET DE LA SANTÉ

La Ministre

Paris, le 16.5.2018

Pég. – D – 18-012185

Monsieur le directeur,

Le Système national des données de santé (SNDS) constitue une des bases de données de santé les plus riches d'Europe.

Le potentiel de ces données est considérable. Elles peuvent être mobilisées pour la recherche clinique et en matière de santé publique, l'aide à la décision des professionnels de santé, l'allocation des financements aux offreurs de santé, le développement d'une médecine prédictive, préventive, personnalisée et participative, ou encore éclairer la transformation et le pilotage du système de santé. Cette liste n'est d'ailleurs pas exhaustive.

Malgré l'ouverture des données prévues par la loi de modernisation du système de santé de janvier 2016, les données du SNDS demeurent sous-exploitées pour deux raisons principales. Premièrement, le périmètre du SNDS est encore limité essentiellement à des données médico-administratives. Son appariement régulier avec des données cliniques apparaît indispensable. Par ailleurs, compte-tenu de la spécificité de ses données, notamment leur volumétrie et leur complexité, un simple droit d'accès ne peut qu'être une première étape pour inciter à leur usage. Il semble nécessaire de mettre à la disposition des chercheurs, des administrations et des acteurs privés, notamment des start-up, les moyens facilitant leur usage, tant en termes d'infrastructure sécurisée et de logiciels que de services de facilitation. Il semble aussi important d'animer la communauté des utilisateurs pour faciliter les partages d'expérience, de compétences et d'outils. Cette mutualisation permettra de révéler le potentiel du SNDS et d'industrialiser son usage.

C'est en ce sens que le Président de la République a annoncé, le 29 mars 2018, au sommet Intelligence artificielle « AI for Humanity », la création d'un Hub. Structure partenariale entre producteurs et utilisateurs des données, elle pilotera l'enrichissement continu mais aussi la valorisation du SNDS, pour y inclure, à terme, l'ensemble des données provenant des remboursements de l'Assurance-maladie, en ajoutant les données cliniques des hôpitaux, de la médecine de ville, des autres producteurs de soins, ainsi que les données de grande qualité, scientifique et médicale, créées dans le cadre de cohortes nationales. Ce « Health Data hub » permettra, dans un cadre parfaitement sécurisé, d'avoir, entre autres, un espace de travail pour l'intelligence artificielle, catalyseur d'innovations. Il associera les grands organismes de recherche français, l'INSERM, le CNRS et l'INRIA, ainsi que de nombreux partenaires publics et privés.

Monsieur Gilles WAINRIB
Directeur scientifique, Owkin
75 rue de Turbigo
75003 Paris

J'ai souhaité vous confier la mission de co-piloter un groupe de travail dédié à la préfiguration de ce Hub, compte tenu de votre expérience et de vos compétences. Vous serez associé à deux copilotes, Mme Dominique Polton et M. Marc Cuggia.

L'objectif de ce groupe de travail est double : préparer la création du Hub et proposer, avec des partenaires, une feuille de route pour l'enrichissement du SNDS.

En ce qui concerne le Hub, le groupe de travail s'attachera à proposer une organisation pour cette alliance ainsi que la définition de la plateforme des données et de l'offre de service qu'elle fournirait et la manière dont elle faciliterait le développement de l'écosystème (acteurs publics, recherche, start-up et autres acteurs privés...). Le groupe de travail déterminera les besoins en nouvelles infrastructures en tenant compte des infrastructures existantes, de leur évolution et des demandes exprimées par les acteurs. Ces nouvelles infrastructures n'ont pas vocation à être exclusives, ni à concurrencer les portails existants ou systèmes d'information des producteurs de données, mais bien à venir enrichir l'offre à destination de l'utilisateur de la donnée. Un cahier des charges, un modèle économique ainsi qu'une réflexion sur l'administration de cette plateforme en régime permanent devront être proposés. Au-delà de la plateforme des données, le groupe de travail devra étudier les moyens à mettre en œuvre par le Hub pour développer les compétences, promouvoir leur partage, ainsi que permettre le développement de partenariats entre les acteurs pour accroître fortement l'usage des données de santé en France. Les modalités de l'action du Hub, l'éventail d'une offre de service initiale et les outils logiciels à faire développer ou à rassembler dans les premiers trimestres devront notamment être définis.

S'agissant de l'enrichissement du SNDS, le groupe de travail s'attachera à identifier les premières sources de données pouvant être fédérées. Des projets autour de leur exploitation pourront être par ailleurs prévus, par exemple sur des aires thérapeutiques à fort enjeu de santé publique. Il s'agira aussi de déterminer les moyens de toute nature nécessaires au développement de ces appariements au cours des prochaines années, notamment pour inciter les acteurs à se lancer dans de telles opérations et industrialiser le processus d'appariement.

A partir de ces éléments, le groupe de travail s'attachera à préciser les missions, le rôle, le périmètre, l'équipe nécessaire à la mise en œuvre, le budget sur les 3 prochaines années et la gouvernance du Hub.

Vous m'adresserez un rapport, d'ici le 21 septembre 2018, décrivant de manière précise les plans d'action proposés pour une mise en œuvre opérationnelle permettant la réalisation des premiers travaux et événements du Hub dès 2019.

Vous travaillerez en lien étroit avec des représentants d'instituts de recherche en santé et dans le domaine de l'intelligence artificielle, des membres de l'écosystème des start-up, des professionnels de santé, des professionnels hospitaliers, un membre du LEEM et un représentant de patients. Vous associerez l'assurance maladie et des représentants du ministère de la Santé et de la Recherche. La DREES et la DSIS vous apporteront leur soutien.

Je sais pouvoir compter sur votre mobilisation pour animer ce chantier en faveur d'une accélération de l'innovation, grâce aux données et dans un cadre parfaitement sécurisé, garantissant l'anonymat et le respect de la vie privée de chaque individu.

Je vous prie de croire, Monsieur le directeur, à l'expression de ma considération distinguée.

Agnès BUZYN

Liste des membres du groupe de travail

Pilotes de la mission :

- Marc Cuggia (CHU de Rennes)
- Dominique Polton (INDS)
- Gilles Wainrib (Owkin)

Rapporteure de la mission :

- Stéphanie Combes (DREES)

Membres du groupe de travail :

- Emmanuel Bacry (CNRS, Ecole Polytechnique)
- Muriel Barlet (DREES)
- Hugues Berry (Inria)
- Thomas Borel (LEEM)
- Isabelle Gentil (DSSIS)
- Claude Gissot (CNAM)
- Didier Guillemot (PUPH)
- Eric Guittet (DGRI)
- Jérôme Kalifa (Lixoft)
- Yves Levy (Inserm)
- Caroline Noublanche (Apricity)
- Elisa Salamanca (AP-HP)
- Henri Verdier (DINSIC)

Suppléants :

- Mathieu Galtier (Owkin)
- Samuel Pilcer (Owkin)
- Ophélie de Dreux Breze (LEEM)
- Youcef Sebiat (Polytechnique)
- Benoît Lavallart (DGRI)
- Paul-Antoine Chevalier (DINSIC)
- Marion Paclot (DINSIC)
- Hélène Caillol (CNAM)
- Claire Giry (Inserm)

Liste des Acronymes

ABM = Agence de la biomédecine

AMI = Appel à manifestation d'intérêt

ANAP = Agence Nationale d'Appui à la Performance

ANSM = Agence Nationale de Sécurité du Médicament et des Produits de Santé

AP-HM = Assistance publique – Hôpitaux de Marseille

AP-HP = Assistance publique - Hôpitaux de Paris

API = Application programming interface

ARS = Agence régionale de santé

ASIP = Agence des Systèmes d'Information Partagés de Santé / Agence française de la santé numérique

ATIH = Agence technique de l'information sur l'hospitalisation

CAD = Collecteur analyseur de données

CASD = Centre d'accès sécurisé aux données

CCAM = Classification commune des actes médicaux

CEA = Commissariat à l'énergie atomiques et aux énergies alternatives

CEPS = Comité économique des produits de santé

CEREES = Comité d'Expertise pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé

CES = Comité éthique et scientifique

CESP = Centre de recherche en épidémiologie et santé des populations

CH = Centre hospitalier

CHU = Centre hospitalier universitaire

CLCC = Centre de lutte contre le cancer

CMG = Collège de la Médecine Générale

CNAM = Caisse nationale de l'assurance maladie

CNIL = Commission nationale de l'informatique et des libertés

CNOM = Conseil National de l'Ordre des Médecins

CNRS = Centre national de la recherche scientifique

CPU = Central processing unit

CSF = Comité stratégique des Filières

DGE = Direction Générale des Entreprises

DGOS = Direction générale de l'offre de soins

DGRI = Direction générale de la recherche et de l'innovation

DGS = Direction Générale de la Santé

DINSIC = Direction interministérielle du numérique et du système d'information et de communication

DMP = Dossier médical partagé

DP = Dossier pharmaceutique

DPI = Dossier patient informatisé

DREES = Direction de la recherche, des études, de l'évaluation et des statistiques

DSSIS = Délégation à la stratégie des systèmes d'information de santé

EHPAD = Etablissement d'hébergement pour personnes âgées dépendantes

FDA = Food & Drug Administration

FEDORU = Fédération des Observatoires Régionaux des Urgences

FEHAP = Fédération des établissements hospitaliers et d'aide à la personne privés non lucratifs

FHF = Fédération Hospitalière de France

FMG = France Médecine Génomique

FNMR = Fédération nationale des médecins radiologues

FSM = Fédération des Spécialités Médicales

GAFAM = Google, Apple, Facebook, Amazon, Microsoft

BATX = Baidu, Alibaba, Tencent, Xiaomi

GCS = Groupement de Coopération Sanitaire

GHT = Groupement hospitalier de territoires

GIP = Groupement d'intérêt public

GPU = Graphics processing unit

HAS = Haute autorité de Santé

IA = Intelligence artificielle

INCa = Institut National Du Cancer

INDS = Institut National des Données de Santé

INED = Institut national d'études démographiques

INR = International Normalised Ratio (mesure du taux de prothrombine)

INRIA = Institut national de recherche en informatique et en automatique

Inserm = Institut national de la santé et de la recherche médical

Instituts 3IA = Instituts interdisciplinaires d'intelligence artificielle

IRISA = Institut de recherche en informatique et systèmes aléatoires

LEEM = Les Entreprises du Médicament

LIMSI = Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur

LIR = Association des laboratoires internationaux de recherche

LOINC = Logical Observation Identifiers Names and Codes

LTSI Rennes = Laboratoire Traitement du Signal et de l'Image

MDPH = Maison départementale des personnes handicapées

MVP = Minimum Viable Product

NGAP = Nomenclature générale des actes professionnels

MiPih = Midi Picardie Informatique Hospitalière

NIR = Numéro d'inscription au répertoire

OCDE = Organisation de coopération et de développement économique

PEPS = Plateforme d'Exploitation des Produits Sentinel

PMSI = Programme de médicalisation des systèmes d'information

PREM = Patient reported experience measures

PROM = Patient reported outcome measures

PU-PH = Professeur des universités-Praticien hospitalier

REIN = Réseau Epidémiologique et Information en Néphrologie

RGPD = Règlement Européen relatif à la protection des données personnelles

RSF = Résumé standardisé de facturation

RPU = Résumé de passage aux urgences

SIGAPS = Système d'Interrogation, de Gestion et d'Analyse des Publications Scientifiques

SNDS = Système National de Données de Santé

SNIIRAM = Système national d'information inter-régimes de l'Assurance maladie

Personnes auditionnées ou rencontrées

Nous remercions les nombreux acteurs rencontrés, sollicités ou auditionnés qui ont largement contribué à la réflexion autour du Health Data Hub et de sa mise en œuvre. Nous espérons que chacun se retrouvera dans cette liste que nous avons souhaitée aussi exhaustive que possible. Nous n'excluons pas que certains noms aient échappés à notre sagacité et remercions également chaleureusement les contributeurs ayant participé qui n'y figureraient pas.

Cecile Couchoud (ABM), Jean Durquety (ABM), Christian Jacquelinet (ABM), Fabrice Sentenac (ABM), Dominique Soulier (ABM), Thomas Van Den Meuvel (ABM), Bernard Nordlinger (Académie de Médecine), Cédric Villani (Académie des sciences), Erwan Médy (ADIT), Stéphan Jeanneau (Adobis), Sandra Mathieu (Adobis), Philippe Brun (Aeglé), Amaury Delorme (Altran), Agnès Fritsch (Altran), Yann Le-fort (Altran), Arnaud Paralieu (ANSES), Rosemary Dray-Spira (ANSM), Evelyne Duplessis (ANSM), Thien Le Tri (ANSM), Jean-Philippe Labille (ANSM), Dominique Martin (ANSM), Bernard Cassou-Mounat (ANSSI), Giorgi Roch (AP-HM),Salam Abbara (AP-HP), Cécile Badoual (AP-HP), Mehdi Benchoufi (AP-HP), Anita Burgun (AP-HP), Hélène Coulonjou (AP-HP), Sylvie Cormont (AP-HP), Christel Daniel (AP-HP), Lauren Demerville (AP-HP), Jean-Sébastien Hulot (AP-HP), Benoit Labarthe (AP-HP), Frederic Laurent (AP-HM), Thomas Lefèvre (AP-HP), Jérôme Marchand-Arvier (AP-HP), Nicolas Paris (AP-HP), Laurent Treluyer (AP-HP), Philippe Ravaud (AP-HP), Caroline Noublanche (Apricity), Marco Fiorini (ARIIS), Anne Haziza (ARIIS), Yannick Le Guen (ARS IDF), Eric Lepage (ARS IDF), Axelle Menu (ARS IDF), Alain Sommer (ASInstitute), Thierry Dart (ASIP), Florent Desgrippes (ASIP), Florence Eon (ASIP), Nicole Janin (ASIP), Pascale Sauvage (ASIP), Rémi Levasseur (ASIP), Max Bensadon (ATIH), François Bourgoin (ATIH), Housseyni Holla (ATIH), Sandra Steunou (ATIH), Sandrine Coulange (Axa), Nacim Darbour (Axa), Philippe Presles (Axa), Cécile Wendling (Axa), Romain Olekhnovitch (BlueSquare), Etienne Grass (Capgemini Invent), Fedoua Baazi (Capgemini), Pierre Demeulemeester (Capgemini), Juliette Du mesnil (Capgemini), Halima Farouqi (Capgemini), Régis Mauger (Capgemini), Damien Jeandel (Carestream Health), Delphine Jollivet (Carestream Health), Kamel Gadouche (CASD), Nora Benhabiles (CEA), Christophe Calvin (CEA), Alexandre Bounouh (CEA), Julien Chiaroni (CEA), Philippe Watteau (CEA), Jean-Claude Labrune (Cegedim), Pierre Ingrand (Centre d'investigation clinique de Poitiers), Sophie Beaupère (Centre Léon Bérard), Jean-Yves Blay (Centre Léon Bérard), Thierry Durand (Centre Léon Bérard), Paul Piersson (Cerba Healthcare), Jerome Thill (Cerba Healthcare), Jean-Hugues Masgnaux (Cercle des Bases), Jacques Massol (Cercle des Bases), Philippe Chatron (CGLabio), Thierry Goujon (CGLabio), Hiep Ngo Trong (CH Rochefort), Patrick Bonnet (CH Valenciennes), Alain Lecherf (CH Valenciennes), Emmanuel Nowak (CHRU Brest), Emmanuel Chazard (CHRU Lille), Samuel Limat (CHU Besançon), Valérie Altuzarra (CHU Bordeaux), Thierry Barthe (CHU Bordeaux), Sébastien Cossin (CHU Bordeaux), Geneviève Chene (CHU Bordeaux), Vianney Jouhet (CHU Bordeaux), Anne Larchevèque (CHU Bordeaux), Régis Lassalle (CHU Bordeaux), Nicolas Magot (CHU Bordeaux), Rodolphe Thielbault (CHU Bordeaux), Philippe Vigouroux (CHU Bordeaux), Rémi Brajeul (CHU Brest), Julien Thevenon (CHU Grenoble), Robert Caiazzo (CHU Lille), Elisabeth Beau (CHU Dijon), Jean-Francois Lahaye (CHU Lille), Jean-François Lefebvre (CHU Limoges), Grégoire Mercier (CHU Montpellier), Pierre-Antoine Gourraud (CHU Nantes), Jean-Pierre Dewitte (CHU Poitier), Frédéric Balusson (CHU Rennes), Guillaume Bouzillé (LTSI/CHU Rennes), Marc Cuggia (LTSI/CHU Rennes), Marie De Tayrac (CHU Rennes), Erwan Drezen (CHU Rennes), Catherine Droitcourt (CHU Rennes), Alain Dupuy (CHU Rennes), Pascal Gaudron (CHU Rennes), André Happe (CHU Rennes), Sandrine Kerbrat (CHU Rennes), Alexandra Lesapagnol (CHU

Rennes), Emmanuel Oger (CHU Rennes), Christine Riou (LTSI/CHU Rennes), Lucie-Marie Scailteux (CHU Rennes), Pascal Van Hille (LTSI/CHU Rennes), Prosper Burq (CHU Toulouse), Ran Balicer (Clalit), Nicolas Glatt (Clinigrid), Pascal Charbonnel (CMG), Pierre-Louis Druais (CMG), Hector Falcoff (CMG), Annelore Coury (CNAM), Claude Gissot (CNAM), Yvon Merlière (CNAM), Erik Boucher de crevecoeur (CNIL), Thomas Dautieu (CNIL), Manon de Fallois (CNIL), Jérôme Gorin (CNIL), Hélène Guimiot-Breaud (CNIL), Jacques Lucas (CNOM), Jamal Atif (CNRS), Pascal Charbonnel (Collège de la Médecine Générale), Thérèse Depeyrot-Ficatier (Consultante indépendante), Jean-Paul Ortiz (CSMF), Nicolas Pécuchet (Dassault), William Saurin (Dassault), Laurent Lafaye (Dawex), Fabrice Tocco (Dawex), Matthieu Landon (DGE – Ministère de l'Economie et des Finances), Julian Mercier (DGE – Ministère de l'Economie et des Finances), Cédric Nozet (DGE – Ministère de l'Economie et des Finances), Christophe Strobel (DGE – Ministère de l'Economie et des Finances), Jean-Yves Fagon (Délégation à l'innovation – Ministère des Solidarités et de la Santé et de la Santé), Stéphanie Decoopman (DGOS – Ministère des Solidarités et de la Santé), Sylvie Escalon (DGOS – Ministère des Solidarités et de la Santé), Caroline Le Gloan (DGOS – Ministère des Solidarités et de la Santé), Laure Maillant (DGOS – Ministère des Solidarités et de la Santé), Diane Tassy (DGOS – Ministère des Solidarités et de la Santé), Marin Dacos (DGRI - Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation), Patrick Garda (DGRI - Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation), Benoît Lavallart (DGRI - Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation), Marie-Christine Plancon (DGRI - Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation), Rémy Sanchez (DGRI - Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation), Anne-Claire Amprou (DGS – Ministère des Solidarités et de la Santé), Patrice Dosquet (DGS - Ministère des Solidarités et de la Santé), Amalia Giakoumakis (DGS – Ministère des Solidarités et de la Santé), Florian Kastler (DGS – Ministère des Solidarités et de la Santé), Florence Lys (DGS – Ministère des Solidarités et de la Santé), Agnès Ramzi (DGS - Ministère des Solidarités et de la Santé), Jérôme Salomon (DGS - Ministère des Solidarités et de la Santé), Xavier Albouy (DINSIC – services du Premier Ministre), Mathilde Bras (DINSIC – services du Premier Ministre), Paul-Antoine Chevalier (DINSIC – services du Premier Ministre), Bertrand Pailhes (DINSIC – services du Premier Ministre), Pierre Pezziardi (DINSIC – services du Premier Ministre), Perica Sucevic (DINSIC – services du Premier Ministre), Claire-Lise Dubost (DREES – Ministère des Solidarités et de la Santé), Javier Nicolau (DREES – Ministère des Solidarités et de la Santé), Matthias Pigneur (DREES – Ministère des Solidarités et de la Santé), Antoine Agis (DSI - Ministère des Solidarités et de la Santé), Emmanuel Bacry (Partenariat Ecole Polytechnique - CNAM), Youcef Sebiat (Partenariat Ecole Polytechnique - CNAM), Jean-Louis Marx (EDL), Pascal Ferard (EFS), Lucile Malard (EFS), Angelique Michaut (EFS), Linda Thieulon (EFS), Nadir Ammour (EITHealth), Jean-Marc Bourez (EITHealth), Alexis Normand (Embleema / Withings), Marie Meynadier (EOS Imaging), Pierre Morichau-Beauchant (EOS Imagining), Olivier de Fresnoye (Epidemium), Jerome Duvernois (esanté-solutions), Henri-Olivier Essienne (Essienne avocats), David Gruson (Ethik IA), Clotaire Thocquenne (Evolucare), Amelie Scheffler (EY pour le LIR), Antoine Bordes (Facebook), Bruno Virieux (Fealinx), Olivier Goëau-Brissonnière (FSM), Claire Desforges (Fédération française des diabétiques), Caroline Guillot (Fédération française des diabétiques), Cécile Chevance (FHF), Frédéric Martineau (FHF), Cyrille Politi (FHF), Alexis Thomas (FHF), Gilles Viudes (FEDORU), Jean-François Goglin (FEHAP), Francis Mambrini (FEIMA), Michel Ballereau (FHP), Jean-Philippe Masson (FNMR), Wilfrid Vincent (FNMR), Jean-Pierre Thierry (France Assos Santé), Alexis Vervialle (France Assos Santé), Eric Germain (Génopole), Jean-Marc Grognet (Génopole), Pierre Tambourain (Génopole), Christophe Lala (GE healthcare), Henri Souchay (GE Healthcare), Hervé de Belenet (Gfi Informatique), Jean-Philippe Vert (Google), Jean-Noël Bail (GSK), Gilles Vassal (Gustave Roussy), Assia Boukemouche-Mezheri (HAS), Chantal Bélorgey (HAS), Benoit Couderc (HAS), Anthony Coue (HAS), Katia Julienne (HAS), Viviane Malterre (HAS), Laetitia May (HAS), Alexandre Vainchstock (Hevaweb), Jean-Charles Dron (HMS), Philippe Castets (HCL), Anne Metzinger (HCL), Philippe Roussel (HCL), Philippe Cinquin (Imag), Alexandre Moreau (Imag), Patricia Blanc (Imagine For Margo), Arnaud Rosier (Implicity), Gouenou

Coatrieux (IMT Atlantique), Philippe-Jean Bousquet (INCa), Thierry Breton (INCa), Christine Chomienne (INCa), Muriel Dahan (INCa), Guy Decarpentrie (INCa), Stéphane De Graeve (INCa), Norbert Ifrah (INCa), Fanta Sarambounou (INCa), Jérôme Viguier (INCa), Valérie Edel (INDS), Marie-Aline Charles (Ined), Elise de la Rochebrochard (Ined), Aline Desesquelles (Ined), Magda Tomasini (INED), Nicolas Anciaux (INRIA), Isabelle Ryl (INRIA), François Sillon (INRIA), Corinne Alberti (Inserm), Dominique Costagliola (Inserm), Vincent Diebolt (Inserm), Guy Fagherazzi (Inserm), Marcel Goldberg (Inserm), Alfredo Hernandez (Inserm), Franck Lethimonnier (Inserm), Frédérique Lesaulnier (Inserm), Yves Levy (Inserm), Grégoire Rey (Inserm), Laurence Watier (Inserm), Marie Zins (Inserm), Pascale Auge (Inserm Transfert), Nacer Boubenna (Inserm Transfert), Xose Fernandez (Institut Curie), Julien Guérin (Institut Curie), Anne-Sophie Hamy (Institut Curie), Alain Livartowski (Institut Curie), Christophe Mattler (Institut Curie), Fabien Reyal (Institut Curie), Michael Pressigout (Institut Pasteur), Karima Bourquard (In-System), Isabelle Gibaud (Interop'Santé), Gerard Domas (Interop'Santé), Florence Cureau (Intersystems), Carlos Raime (Intersystems), Michèle Arnoe (IQVIA), Claire Lamotte (IQVIA), Marie-Hélène Royer (IQVIA), Ashley Woolmore (IQVIA), Thomas Guyet (IRISA), Patrick Olivier (Ivbar), Sébastien Woynar (Ibofrance), Thomas Borel (LEEM), Marianne Cimino (Lessis), Dominique Gougerot (Lessis), Julie Dumons (Lifen), Thibault Naline (Lifen), Pierre Zweigenbaum (LIMSI - CNRS), Agnès Renard (LIR), Jérôme Kalifa (Lixoft Rythm), Pierre Lemordant (LTSI), Lotfi Senhadji (LTSI), Philippe Lagouarde (Maincare), Sebastien Wafflart (Maincare), Stéphane Barde (Malakoff Médéric), Laurent Borella (Malakoff Médéric), David Giblas (Malakoff Médéric), Raphaël Soulignac (Malakoff Médéric), Thomas London (McKinsey), Lise Marin (Medasys), Bertrand Rondepierre (Ministère des Armées), Jacques Battistoni (MG France), Frédéric Serein (MIPIH), Cedric Lemoy (MIPS), Pierre-Yves Brossard (MSD), Romain Finas (MSD), Laurie Levy-Bachelot (MSD), Anny Tirel (MSD), Arnault Ioualalen (Numalis), Jillian Oderkirk (OCDE), Valérie Paris (OCDE), Claude-Alain Cudennec (Oncodesign), Maude Liotard (Oncodesin), Alain Durakovic (Oniam), Catherine Commaille-Chapus (Openhealth), Jean-Yves Robin (Openhealth), Yohann Poiron (OpenXtrem), Joëlle Bouet (Opusline), Bertrand De Neuville (Opusline), Youssef Mallat (Opusline), Rémy Choquet (Orange), Alain Trugeon (OR2S), Mathieu Galtier (Owkin), Gilles Wainrib (Owkin), Servane Augier (Outscale), Wilfrid Romuald (Panoratio), François Macary (Phast-Services), Anne Maheust (Phast-Services), Caroline Henry (Pons & carrère), Julien Dubuis (Praxis), David Darmon (Primege), Vincent Breteau (Région Normandie), Danielle Fancony (Reims Santé Travail), Yvanie Caillet (Renaloo), Frédéric Chassagnol (Roche), Jean-Marc Pinguet (Roche), Thomas Duval (Sancare), Bernard Hamelin (Sanofi), Philippe Maugendre (Sanofi), Eric Vacaresse (Sanofi), Marjorie Boussac (Santé Publique France), Clothilde Hachin (Santé Publique France), Yann Le Strat (Santé Publique France), Martial Mettendorff (Santé Publique France), Stéphane Nardy (Santé Publique France), Mariane Binst (Santéclair), Julien Brossard (Santeos), Fabrice Daverio (Santeos), Christelle Pointreau (Santeos), Christophe Richard (Santeos), Pierre Hornus (Sêmeia), Guillaume Buwalda (SeqOne), Jean-Marc Holder (SeqOne), Nicolas Philippe (SeqOne), Olivier Nosjean (Servier), Eric Guessant (SIB), Brigitte Congard-Chassol (Snitem), Armelle Graciet (Snitem), François-Régis Moulines (Snitem), William Rolland (Snitem), Martine Gilard (Société Française de cardiologie), Jacques de tournemire (Société Française de cardiologie), Jean-François Meder (Société Française de radiologie), Florence Ribadeau-Dumas (Service de santé des armées), Philippe Szidon (SFMG), Dominique Pon (STSS – chantier numérique), Eric Boniface (Substra), Raphaëlle Frija (Syntec), Christophe Richard (Syntec), Isabelle Zablit-Schmitz (Syntec), Anne-Sophie Taillandier (Teralab), Antoine Evennou (Terra Nova), Luc Pierron (Terra Nova), Olivier Clatz (Therapixel), Pascale Flamant (Unicancer), Emmanuel Reyrat (Unicancer), Mathieu Robain (Unicancer), Christian Lovis (Université de Genève), Gérard Friedlander (Université Paris Descartes), Antoine Tesnière (Université Paris Descartes), Jean-François Ethier (Université de Sherbrooke, Québec), Serge Uzan (UPMC), Jean-François Forget (Vidal), Allan Rodriguez (VitaDX), Nicholas Henderson (We design services), Matthieu Laude (We design services).

HEALTH DATA HUB

Mission de préfiguration

Une mission pilotée par Marc Cuggia (CHU Rennes), Dominique Polton (INDS), Gilles Wainrib (OWKIN) et rapportée par Stéphanie Combes (DREES)