

Recommandations pour un modèle économique de la mise à disposition des données de santé issues du Système National des Données de Santé

Contrat Stratégique de Filière des Industries et Technologies de Santé (CST-ITS)

Projet structurant n°2 : Intelligence artificielle et santé

GT1 - Valorisation des données partagées et Modèle Economique de la plateforme (Health data Hub)
Copilotage : Frédérique Debroucker – Medtronic et Philippe Maugendre – Sanofi

Version finale du 7 février 2020

Contexte

Le Système des Données de Santé (SNDS) géré par la plateforme (Health data Hub)¹ est régi par la loi du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé, dans son article 41 (Chapitre I « Innover en valorisant les données cliniques » ; Titre III « Développer l'ambition numérique en santé ») qui a élargi le périmètre des données incluses dans le SNDS² et prévoit la mise en place d'une nouvelle gouvernance avec la création de la Plateforme des données de Santé (Health data Hub)

Le Contrat Stratégique de Filière des industries de santé (CSF-ITS) a été signé le 4 février 2019 entre :

- M Jean-Luc Bélingard, Président de la Fefis, représentant les industries de santé,
- Mme Agnès Buzyn, Ministre des Solidarités et de la Santé,
- Mme Frédérique Vidal, Ministre de l'enseignement supérieur, de la Recherche et de l'innovation,
- Mme Agnès Pannier-Runacher, Secrétaire d'Etat auprès du Ministre de l'Economie et des finances,
- et les représentants syndicaux Yan Tran (CFDT), Pascal Lopez (FO).

Ce contrat prévoit 5 projets structurants pour la filière dont le Projet structurant n°2 concernant l'intelligence artificielle en santé. Au cœur de ce projet, il a été souhaité une coordination des positions entre les industriels de la santé et du numérique et les pouvoirs publics sur la valorisation des données partagées et du modèle économique du Health data Hub.

C'est dans ce contexte qu'un Groupe de Travail (GT) spécifique³ a été mis en place afin d'émettre des recommandations pour un modèle économique de la mise à disposition des données de santé issues

¹ La loi du 24 juillet 2019 utilise le terme de plateforme pour nommer la structure GIP en remplacement de l'INDS (Institut National des Données de Santé). Pour autant, le terme de Health Data Hub étant préférentiellement utilisé dans les échanges courants, c'est celui-ci qui sera retenu dans la suite de la note.

² Présentation des modalités d'élargissement du SNDS en Annexe 1.

³ Composition du groupe de travail en Annexe 2.

du Système National des Données de Santé élargi. Le GT pour son analyse s'est basé en particulier sur des exemples de bases de données publiques ou privées, françaises ou internationales et sur quelques articles abordant la question des modèles économiques des bases de données (Cf. Annexe 3 & 4).

Pour rappel, dans une logique de soutenabilité économique du SNDS élargi, l'accès aux données n'est pas gratuit, à l'exception de certaines situations définies l'article L. 1461-5 du code de la santé publique⁴. A ce titre, il est à noter qu'un arrêté de tarification du SNDS historique avec le périmètre antérieur à son élargissement par la loi du 24 juillet 2019 est actuellement à l'étude. Il s'agissait de la mise à jour d'une tarification préexistante qui ne concernait que le PMSI⁵.

La tarification de l'accès aux données dans le cadre de la création du SNDS élargi, sous l'égide de la nouvelle plateforme (Health data Hub) oblige à une nouvelle réflexion et une tarification adaptée à son élargissement, basées en particulier sur le travail de ce Groupe de travail.

Ces recommandations sont amenées à être portées auprès du Ministère de la Santé (Drees) et de la Plateforme de données (Health data Hub) pour être prises en compte dans les décisions concernant le modèle économique de l'accès aux données du Health data Hub. Elles ont aussi vocation à alimenter la réflexion plus large sur l'approche économique du Health data Hub actuellement à l'étude par le Conseil National du Numérique.

Rappel des objectifs que les recommandations devront prendre en compte

Le modèle économique doit être simple, avec des règles évolutives pour éviter de créer des freins inutiles. Il doit permettre à la fois de récompenser les efforts de production et d'organisation de la donnée de santé et un accès aux données viable qui respecte les différentes typologies d'acteurs, de la start-up aux grands groupes, en passant par les différents acteurs de la chaîne de soins (associations et fédérations, bureaux d'études, laboratoires de recherche).

L'objectif des recommandations sur le modèle économique de l'accès aux données était d'apporter les éléments d'orientation et de préconisation concernant la valorisation des données susceptibles d'être versées et gérées par le Health data Hub quelle que soit leur source publique ou privée. Sur cette base, ces recommandations doivent permettre de définir les grandes lignes d'un modèle économique du versement des données au SNDS élargi et de leur accès aux données au bénéfice de différents acteurs via le Health data Hub (flux entrant ou sortant).

Il faut également prendre en compte le fait que le Health Data Hub est à la fois une plateforme technologique agrégeant des bases de données publiques ou privées et un fournisseur de services *ad hoc* (aide à l'appariement, transformation des données, accompagnement à la création d'algorithmes, ...).

La proposition de rémunération/tarification doit être incitative au versement et à l'utilisation des données via le Health Data Hub et ainsi éviter la création non coordonnée ou non mutualisée de plusieurs hubs sur le territoire français.

⁴ Les recherches, les études ou les évaluations demandées par l'autorité publique ou les recherches réalisées exclusivement pour les besoins de services publics administratifs.

⁵ Le Programme de Médicalisation des Systèmes d'Information, est un outil de description et de mesure médico-économique de l'activité hospitalière. Introduit en France dans les années 80 il a d'abord été présenté comme un outil épidémiologique avant de devenir un outil d'allocation budgétaire.

Dans la suite du document, par « accès aux données de santé », on entend aussi bien l'accès à des données isolées, des données agrégées, des bases de données filles ou la base de données complète.

Les recommandations dans la suite du document, concerne à la fois la mise à disposition des données par le producteur/fournisseur sur la plateforme du HDH et l'accès aux données par les utilisateurs via le HDH.

Considérations générales émises par le Groupe de Travail

Pour élaborer ses recommandations, le Groupe de Travail a pris en compte les considérations ci-dessous, concernant la rémunération des producteurs/fournisseurs de données ou la tarification applicable aux demandeurs/utilisateurs.

Le modèle économique doit respecter les spécificités des différents acteurs

En effet, le modèle économique doit répondre à la fois aux attentes des producteurs/fournisseurs de données (juste reconnaissance de l'investissement financier a priori, de l'effort et de l'expertise nécessaire pour la production et de la mise à disposition des données) et à celles des demandeurs/utilisateurs de données (poids financier de l'accès aux données), une même entité pouvant au cours du temps être à la fois producteur et utilisateur de données.

Par ailleurs, les recommandations doivent tenir compte de la diversité des statuts, aussi bien des producteurs que des utilisateurs (public, privé, start-up / ETI, grands groupes, unités de recherche, fédérations, associations, agences de l'Etat, administrations ...), qui ne relèvent pas des mêmes règles au sens de la Loi (accès gratuit aux données pour certains utilisateurs, en particulier pour les instances publiques).

Le modèle économique doit s'adapter aux évolutions du périmètre du SNDS

Les recommandations doivent prendre en compte des situations très différentes, tant sur le périmètre des données que des conditions de versement de ces données dans la SNDS. Désormais, le Health data Hub devra gérer un accès aux données :

- « L'actuel SNDS » existant avant la Loi du 24 juillet 2019, constitué essentiellement des bases publiques PMSI, SNIIRAM, CEPI-DC, etc...
- Les données assujetties à l'article 41 de la Loi du 24 juillet 2019, en particulier les données des cohortes, des entrepôts hospitaliers, du médico-social, de l'autonomie, du dépistage (Cf. Annexe 1).
- Les données versées à la plateforme sur la base du volontariat.

Le périmètre du catalogue des bases de données issues du « SNDS élargi » résultant de l'ensemble de ces données et stockées sur la plateforme technologique du Health data Hub, selon les conditions définies par décret (en attente de publication), sera précisé par un arrêté précisant les données et bases de données accessibles via le Health data Hub.

Au-delà des situations de gratuité prévues par la Loi, le modèle économique devra à la fois permettre une juste rémunération des données versées et une rémunération incitative au versement volontaire de données.

Le modèle économique doit donc intégrer dans la rémunération des producteurs/fournisseurs de données, le caractère multi-dimensionnel des coûts de production, de l'expertise nécessaire à la constitution de la base et des coûts de mise à disposition des données

Il est primordial que l'ensemble des étapes et de l'implication des différents acteurs dans la production et la mise à disposition des données soit récompensé à sa juste valeur et que le caractère incitatif de la création de base de données perdure y compris dans le contexte de la mise en œuvre de l'article 41 de la Loi du 24 juillet 2019. La mise à disposition des données dans le SNDS élargi ne doit pas apparaître comme une dépossession du travail et de l'engagement parfois très importants et nécessaires à la création de bases de données.

La valorisation des efforts consentis nous paraît clef afin d'encourager et favoriser l'innovation et permettre de continuer le développement de solutions et processus aboutissant à la création de bases de données de qualité et interopérables dans l'environnement international très compétitif.

Il doit ainsi permettre une valorisation cohérente par rapport à la valeur créée par ces bases de données constituant le fondement d'années ou de dizaines d'années de travail, d'expertise permettant l'apprentissage des jeunes médecins et chercheurs auprès des équipes plus aguerries.

Parmi les coûts liés aux différentes étapes permettant la mise à disposition de données, on notera en particulier :

- Les coûts de création d'une base de données, souvent ex-nihilo, parfois très élevés, et reposant sur un nombre limité d'acteurs/financeurs (coûts d'élaboration du projet, de recueil de données, de constitution de bases de données, stockage, y compris les coûts de recueil du consentement des patients). Dans ce cadre, l'accès aux bases déjà existantes, et en particulier issues d'un modèle économique basé sur des capitaux publics/privés, doit également être pris en compte.
- Les coûts d'entretien de la base (flux d'alimentation, saisie) et de mise aux normes de qualité (en particulier interopérabilité), le contrôle qualité interne et les audits externes (ex. audit INDS/HdH, FDA, ...).
- Les coûts d'extraction et de mise à disposition des données.
- Les coûts de services complémentaires ou *ad hoc* de caractérisation des données (scoring, annotations/cotations, algorithmes, ...), d'expertise à la compréhension de la base et des données ou pour accompagner l'analyse), de propriété intellectuelle, s'ils sont inclus dans la mise à disposition des données,...

Ces coûts peuvent être exprimés en valeur financière et/ou en jours/homme.

Le modèle économique doit en premier lieu, baser la rémunération et la tarification sur la valeur des données

Il apparaît essentiel, pour le Groupe de travail, que le critère « Valeur » soit la base tant de la rémunération de la mise à disposition de données dans le SNDS élargi, que de la tarification de l'accès aux données des données par le Health Data Hub. En effet, le modèle économique doit prendre en compte la valeur plutôt que des critères tels que l'origine de la donnée (publique ou privée), ou la volumétrie brute des données (téraoctets).

En effet, la qualité des données doit être encouragée pour permettre la reconnaissance des études et analyses demandées ou destinées aux autorités de santé et également pour en faire un atout à l'international.

Dans ce contexte, le GT a défini plusieurs typologies de la valeur :

- Valeur liée à la qualité intrinsèque des données : types de données, standard d'interopérabilité (ex. : standard OMOP), structuration et codage, contrôle qualité, données manquantes, respect de la charte des producteurs de données, existence d'audits qualités réalisés par le fournisseur et possibilité d'accès à ces audits, degré d'anonymisation et possibilité d'appariement (NIR), localisation des données (serveur sécurisé ou requête multiples ou mise à disposition chez le demandeur), ... Cette valeur peut donc se traduire en coûts tels que listés précédemment puisqu'une base de plus ou moins bonne qualité aura demandé plus ou moins d'investissement humain, financier, technique.
- Valeur liée à l'intérêt d'utilisation des données par l'utilisateur : intérêt des données et critères disponibles, données manquantes, données brutes/données primaires, ancienneté et historique des données, degré d'actualisation, exhaustivité (nombre de patients / dossiers / items), représentativité, origine des données (géographie, granularité), ... Cette valeur doit être établie selon une grille de critères.
Elle n'a pas vocation à être définie utilisateur par utilisateur, mais doit prendre en compte l'intérêt global d'utiliser une base de données pour faciliter les études ou apporter des informations spécifiques, dans un souci de lisibilité : permettre des études pour des petites populations ou des populations fragmentées, éviter une étude *ad hoc*, mettre à disposition des informations médicales, permettre des requêtes itératives, permettre des études longitudinales, ...
Au-delà du coût, il est à noter qu'une base de données à la qualité reconnue ou originale se traduira par une plus forte attractivité et pourra donc générer plus d'utilisation et donc plus de recettes.
Cette valeur ou ces spécificités doivent être définies base par base et les spécificités reconnues doivent figurer au catalogue des données disponibles via le Health data Hub.
- Valeur liée à la transformation des données : appariement probabiliste, données retravaillées⁶ en lien avec le projet (cotation/annotation, analyse scientifique, analyse algorithmique), possibilité d'une utilisation secondaire (aide à l'analyse primaire ou analyse sur base fille), expertise, dashboard ou rapports pré-définis.
- Valeur liée à l'usage final des données : valeur du produit/solution obtenu grâce, en partie ou en totalité, aux données issues du SNDS élargi, propriété intellectuelle, Cette valeur pouvant le cas échéant aboutir à un mécanisme de partage de la valeur entre l'utilisateur et le producteur : partage soit financier, soit basé sur des accords de partenariats entre le producteur/fournisseur et le demandeur/utilisateur, lorsqu'ils ont développé ensemble l'innovation ou le produit final.

⁶ Une donnée retravaillée est une donnée qui a été revue, corrigée, codée ou qui a fait l'objet d'une revue pour en assurer la cohérence et l'intégrité à des fins de recherche ou d'amélioration de la prise en charge des patients.

Le Groupe de travail considère que le modèle économique doit intégrer deux approches dans la rémunération/tarifification des différents acteurs :

- d'une part un accès aux données, géré par le Health data Hub (avec un accompagnement spécifique), fondé sur une tarification orientée par les coûts et basée sur les valeurs liées à la qualité et à la capacité d'utilisation des données ,
- d'autre part, des services *ad hoc* facultatifs apportés directement par le producteur/fournisseur de données au demandeur, basés sur les valeurs liées à la transformation des données et à l'usage final (partage de la valeur).

A noter : concernant la tarification pour l'accès aux données, il existe un point de divergence. En effet, certains acteurs du Groupe de Travail (AP-HP et Inserm), considèrent que la tarification de l'accès aux données devrait systématiquement prendre en compte la qualité de l'utilisateur et l'usage final qu'il souhaite faire des données obtenues. Cette position minoritaire n'a pas été retenue par le groupe. Celui-ci considérant que la tarification d'accès aux données ne doit pas dépendre de la valeur d'usage et qu'elle doit être prévisible pour l'utilisateur comme pour le producteur (tarifs non variables d'un acteur à l'autre ou d'une utilisation à l'autre).

Le modèle économique doit intégrer différentes modalités financières afin de répondre aux attentes des différents acteurs (producteurs/fournisseurs, utilisateurs, Health Data Hub)

Sur la base des différents exemples de modèles économiques étudiés régissant les bases de données existantes, françaises ou étrangères, le GT suggère un modèle économique mixant plusieurs modalités financières :

- Une tarification pour l'accès aux données,
- une rémunération des services *ad hoc* du Health Data Hub
- une rémunération facultative et additionnelle des services *ad hoc* apportés, directement au demandeur, par le producteur/fournisseur des données.

Le GT considère qu'il faudrait explorer la possibilité juridique d'un mécanisme d'étalement des paiements de l'accès aux données pour certaines structures naissantes ou avec une trésorerie insuffisante au moment de la collecte des données.

Par ailleurs, lorsqu'un acteur a participé expressément à la création d'une base, d'un registre, d'une cohorte, etc. les frais d'accès aux données de cette base, registre ou cohorte, faisant partie du catalogue du Health data Hub, ne lui sont pas facturés, en tenant compte de sa contribution.

En complément, le groupe de travail a exploré des approches complémentaires :

- Au-delà des accès gratuits pour les acteurs de la recherche publique pour leur activité propre (donc hors prestations pour un tiers privé), les agences et autorités publiques, les demandes d'études auprès des industriels effectuées par les autorités, il pourrait être étudié la possibilité de la gratuité d'accès aux données pour certains acteurs ou certaines situations (rédaction de thèse, réponse à un appel d'offres institutionnel, ...)
- La possibilité de troc permettant, pour un acteur donné, un accès gratuit à tout ou partie des données, en échange d'une contribution (données versées au SNDS élargi ou de mise à disposition de moyens techniques ou humains au Health Data Hub) a été discuté). Cependant, elle ne peut pas être proposée au regard des contraintes réglementaires ou juridiques.

Le modèle économique doit être, juste, évolutif et pragmatique

Afin que le modèle économique soit facilement évolutif en lien avec l'augmentation des bases mises à disposition et du nombre demandeurs, le groupe de travail recommande que les rémunérations et les tarifications :

- soient validées et publiées directement par le Health data Hub et non via un circuit réglementaire (décret et arrêtés).
- soient validées au sein du Health data Hub avec l'aide d'un collectif d'experts issus des différents collèges prévus dans la convention constitutive du Health data Hub. En effet, le travail technique d'évaluation de tarification (rémunération des producteurs/fournisseurs de données) doit relever d'un échange paritaire afin de conserver la dynamique d'échanges entre les différents acteurs et pour prendre en compte à leur juste valeur les différents types de coûts.

Cette évaluation de tarification Producteur par producteur, devrait s'effectuer sur la base d'un dossier argumenté présentant les différents coûts et les valeurs des données proposées.

Enfin, une réflexion concernant les demandes des acteurs en dehors du territoire français a abouti à proposer le même modèle économique, considérant que dans l'objectif d'attractivité souhaité par le gouvernement français et au regard du droit de la concurrence européenne, il n'y avait pas lieu de faire une différenciation entre les acteurs (France ou hors France) et que la multiplication des acteurs permettrait à terme de diminuer les coûts d'utilisation du SNDS élargi, via le Health Data Hub.

Recommandations sur le modèle économique de mise à disposition des données de santé du SNDS

Le groupe de travail considère que le modèle économique concernant la rémunération et la tarification des données de santé du SNDS via le Health Data Hub, doit être :

1. Un modèle **évolutif**, adaptable aux évolutions du SNDS
2. Un modèle **juste**, prenant objectivement en compte les différents coûts (production, gestion des données, mise à disposition des données)
3. Un modèle **centré sur la valeur** des données
4. Un modèle **pragmatique** avec différentes modalités financières
5. Un modèle **équilibré**, avec une garantie d'une tarification paritaire.

Dans un souci de pragmatisme et d'agilité, le groupe de travail recommande que la validation et la publication des tarifications relèvent de la responsabilité du Health data Hub et non d'une décision réglementaire via un arrêté.

En conséquence, le groupe de travail préconise un modèle économique, articulé autour de différents modules financiers (Cf. schéma pages suivants) :

Une tarification pour l'accès aux données

Ces tarifs seront basés sur les valeurs liées à la qualité et à l'intérêt d'utilisation des données, en lien avec les coûts de production, d'entretien et de mise à jour de la base assumés par le producteur/fournisseur.

A noter : concernant la tarification pour l'accès aux données, il existe un point de divergence . En effet, certains acteurs du Groupe de Travail (AP-HP et Inserm), considèrent que la tarification de l'accès aux données devrait systématiquement prendre en compte la qualité de l'utilisateur et

l'usage final qu'il souhaite faire des données obtenues. Cette position minoritaire n'a pas été retenue par le groupe. Celui-ci considérant que la tarification d'accès aux données ne doit pas dépendre de la valeur d'usage et qu'elle doit être prévisible pour l'utilisateur comme pour le producteur (tarifs non variables d'un acteur à l'autre ou d'une utilisation à l'autre).

Ces tarifications permettraient d'assurer un socle de rémunération pour les producteurs/fournisseurs de données. Cependant, cette rémunération n'a pas pour vocation à couvrir l'ensemble des coûts de production des bases.

Ils seraient négociés de gré à gré entre le Health data Hub et le producteur/fournisseur⁷.

Ainsi, ces accords de gré à gré permettraient d'éviter des distorsions de concurrence et de prendre en compte :

- La valeur de marché en France et à l'international
- Le respect des modèles actuels des producteurs/fournisseurs et le respect de leur engagement et contrats préexistants
- La valeur ajoutée apportée par chaque base.

Si le demandeur a expressément participé à la création de la base de données / registre / de la cohorte pour lequel il demande l'accès, la tarification n'est pas appliquée ou sera adaptée en tenant compte de sa contribution.

Il pourrait être étudié la possibilité que cette tarification s'applique soit à la requête ou soit via des forfaits, dans le cas de demandes itératives ou relevant d'une méthodologie de référence.

Ces tarifications font l'objet d'une convention entre le Health data Hub et chaque producteur/fournisseur de données et entre le Health data Hub et l'utilisateur;

Ces tarifications sont publiées dans le catalogue du Health data Hub, permettant ainsi une bonne lisibilité et prévisibilité pour les demandeurs/utilisateurs et les producteurs/fournisseurs.

Si la demande concerne l'accès à plusieurs bases, la tarification est alors additionnelle et prendra en compte les caractéristiques de chaque base.

Une tarification des services de base apportés par le Health data Hub

Une tarification permettant la rémunération du Health data Hub pour les services de base qu'il a dû mettre en œuvre pour la mise à disposition des données⁸ :

- Etablissement du catalogue,
- Mise à disposition d'un stockage dans un environnement sécurisé. Ce stockage pouvant être le cas échéant compatible avec le temps nécessaire avec un audit des autorités, en particulier d'enregistrement (EMA, FDA, ...)
- Appariement NIR des bases demandées
- Etablissement des conventions avec les producteurs/fournisseurs et les utilisateurs

⁷ Le cas échéant auprès d'une association ou d'un collectif de producteurs/fournisseurs. Sans se prévaloir de la répartition intra-producteur/fournisseur qui relèvera de la prérogative de l'association ou du collectif.

⁸ Il est rappelé que si le Health Data Hub est le guichet d'entrée unique et obligatoire sur le plan réglementaire afin de s'assurer que les recherches menées sont conformes à la législation, ces recherches peuvent être menées avec ou sans les prestations proposées par le HDH (mise à disposition de compétences techniques, algorithmes ...).

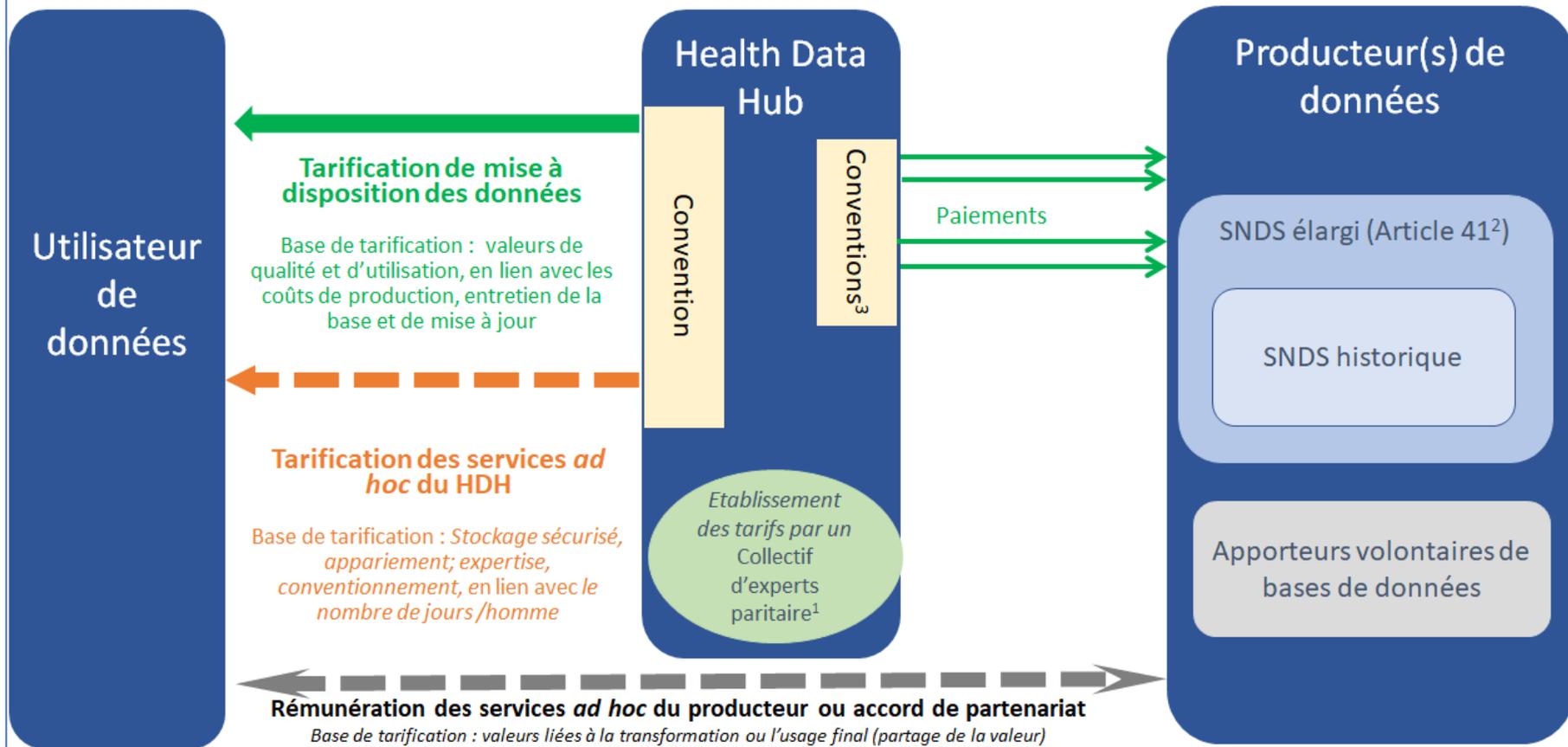
Cette tarification sera définie pour chaque demande en fonction de la contribution apportée par le Health data Hub en lien avec la complexité de la demande (quantité stockée, et nombre d'appariements) et le temps passé (Jours / homme).

Une rémunération facultative en fonction des services optionnels du producteur/fournisseur

Il existe des services *ad hoc* et facultatifs, apportés par le producteur/fournisseur des données directement au demandeur, liés soit à la valeur en lien avec la transformation des données ou à une expertise particulière (expertise et accompagnement spécifique tels que conception de l'étude, appariement probabiliste, conception d'algorithme, annotation des images, interprétation, caractérisation ou scoring des résultats, ...), à la mise en place de bases filles, à l'établissement de rapports spécifiques, ou soit en lien avec la valeur de l'usage final de la donnée (propriété intellectuelle, partage de la valeur).

Ces services *ad hoc* relèvent d'une rémunération négociée directement entre le producteur/fournisseur et l'utilisateur/demandeur ou d'un accord de partenariat, en dehors du Health data Hub, qui peut toutefois en faire la publicité afin d'informer le demandeur de l'existence de ces services.

Modèle économique d'accès aux données du SNDS



1 Composé d'experts issus des Collèges prévus à la convention constitutive du HDH

2 Loi du 24 juillet 2019

3 Négociées de gré à gré, producteur par producteur

Collectif d'experts paritaire

Le Health data Hub doit être en charge de la négociation des conditions avec les producteurs/fournisseurs, de la validation des tarifications et du reversement auprès des producteurs/fournisseurs de données, dans le cadre de l'accès aux données de santé du SNDS élargi.

Pour ce faire, il s'appuie sur l'expertise technique de différents experts proposés par les Collèges prévus dans la convention constitutive du Health data Hub, afin de conserver la dynamique d'échanges entre les différents acteurs et pour prendre en compte les différents types de coûts et de valeurs.

Il définit les modalités de forfaitisation pour les requêtes itératives ou relevant d'une méthodologie de référence, ainsi que les catégories d'utilisateurs ou les situations pouvant relever d'une gratuité totale ou partielle. Ces points seront prévus lors de la négociation sur la tarification et seront intégrés à la convention signée entre le Health data Hub et le producteur/fournisseur.

Par ailleurs, il faut également prévoir un mécanisme financier qui permette « d'étaler » le paiement pour solvabiliser l'accès aux données du SNDS via le Health Data Hub, pour les structures avec une trésorerie naissante ou insuffisante au moment de la collecte des données.

Ce mécanisme pourrait correspondre à une mensualisation ou un report du paiement attendu par le Health Data Hub, sans contrepartie financière (pénalité ou intérêt)

Cette facilité de paiement, qui pourrait s'apparenter à un système de « leasing de la data » pourrait être octroyée par le Health data Hub sur la base de critères de solvabilité définis par la BPI.

Convention / échanges entre les acteurs

Le Groupe de travail considère que les échanges administratifs et juridiques, liés à l'accès aux données, ne doivent pas impacter les délais d'accès aux données grandement raccourcis par la mise en place de l'INDS (puis du Health data Hub).

Afin de fluidifier les échanges entre les acteurs et d'éviter les délais longs régulièrement observés en cas de conventions multipartenaires, il est préconisé de mettre en place deux circuits principaux de conventionnement pour encadrer la mise à disposition des données du SNDS élargi via le Health data Hub :

- Une convention entre chaque producteurs/fournisseur de données et le Health data Hub définissant la tarification de mise à disposition des données, le cas échéant les modalités particulières liées à cette mise à disposition, et prévoyant les modalités de reversement/rémunération du producteur/fournisseur par le Health data Hub.
- Une convention entre l'utilisateur/demandeur et le Health data Hub, régissant l'accès aux données et leur tarification (quel que soit le nombre de bases de données). Cette convention pourra s'inspirer des conventions actuelles entre la Cnam, le responsable de traitement (RT) et le Responsable de mise en œuvre du traitement (RMOT). En lien avec l'autorisation CNIL, le cas échéant le temps de stockage nécessaire avec un audit des autorités, en particulier d'enregistrement (EMA, FDA, ...)

Le cas échéant, les services *ad hoc* facultatifs (ou de partage de la valeur), relèveront de conventions particulières entre le producteur/fournisseur et l'utilisateur, indépendamment des conventions de mise à disposition des données.

Par ailleurs, il faudra prévoir dans les conventions une définition précise des rôles et responsabilités du producteur de données, du fournisseur de données, du prestataire, de l'analyste de données (RMOT), et de l'utilisateur final de données, en particulier concernant la « propriété » des données initiales ou générées.

Annexe 1 : Art 41 – définition et élargissement du SNDS

Suite à l'adoption de la [loi n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé](#) (la « LOTSS »), le périmètre des données incluses dans le SNDS a été élargi.

Périmètre du SNDS avant élargissement

Pour mémoire, le SNDS a été créé par l'article 193 de la [loi n° 2016-41 de modernisation de notre système de santé du 26 janvier 2016](#) et regroupait, jusqu'à son élargissement :

- les données des établissements de santé (base PMSI) (cf. **1°** de l'article L. 1461-1 I du CSP),
- les données de l'assurance maladie (base SNIIRAM) (cf. **2°** de l'article L. 1461-1 I du CSP),
- les données sur les causes médicales de décès (base CépIDC de l'INSERM) (cf. **3°** de l'article L. 1461-1 I du CSP),
- les données médico-sociales relatives au handicap (en provenance des maisons départementales des personnes handicapées) (cf. **4°** de l'article L. 1461-1 I du CSP),
- l'échantillons des données de remboursement issu des complémentaires (cf. **5°** de l'article L. 1461-1 I du CSP).

Périmètre élargi du SNDS

L'article 41 de la LOTSS du 24 juillet 2019, via l'ajout de nouvelles dispositions à l'article L. 1461-1 I du CSP, a enrichi le SNDS d'un nombre important de données nouvelles :

- les données cliniques collectées par les praticiens en ville et à l'hôpital lors des actes pris en charge par l'assurance maladie au titre de la maladie, la maternité, les accidents du travail et les maladies professionnelles (cf. **6°** de l'article L. 1461-1 I du CSP),
- les données relatives à la perte d'autonomie (cf. **7°** de l'article L. 1461-1 I du CSP),
- les données collectées dans le cadre des enquêtes dans le domaine de la santé (cf. **8°** de l'article L. 1461-1 I du CSP),
- les données collectées dans le cadre des enquêtes dans le domaine de la santé (cf. **9°** de l'article L. 1461-1 I du CSP),
- les données recueillies par les services de protection maternelle et infantile (cf. **10°** de l'article L. 1461-1 I du CSP),
- les données de santé recueillies lors des visites d'information et de prévention (cf. **11°** de l'article L. 1461-1 I du CSP).

Focus sur les nouveaux articles L. 1461-1 I 6° et 8° du CSP

Le nouvel article L. 1461-1 I 6° du CSP dispose que « I.- Le système national des données de santé rassemble et met à disposition :

(...)

6° Les données destinées aux professionnels et organismes de santé recueillies à l'occasion des activités mentionnées au I de l'article L. 1111-8 du présent code donnant lieu à la prise en charge des frais de santé en matière de maladie ou de maternité mentionnée à l'article L. 160-1 du code de la sécurité sociale et à la prise en charge des prestations mentionnées à l'article L. 431-1 du même code en matière d'accidents du travail et de maladies professionnelles ; (...) »

L'analyse de cet article permet de dégager trois critères de qualification des nouvelles données incluses dans le SNDS élargi, qu'il convient d'appliquer cumulativement :

- ① données destinées aux professionnels et organismes de santé,
- ② recueillies à l'occasion des activités de prévention, de diagnostic, de soins ou de suivi social et médico-social,

③ donnant lieu à la prise en charge par l'assurance maladie des frais de santé en matière de maladie et maternité, et des prestations en matière d'accidents de travail et de maladies professionnelles.

Le nouveau texte ne décrit cependant pas le format ou les types de données sous réserve des 3 critères. Cependant et dans la mesure où les « données destinées aux professionnels et organismes de santé » sont visées, il convient de considérer que les données concernées sont les données primaires, dès lors qu'il s'agit des données destinées directement aux professionnels et organismes de santé.

Ainsi et au regard de ces trois critères :

- **sont incluses** dans le périmètre du SNDS élargi les données sources suivantes :
 - les données traitées dans les logiciels métiers des professionnels de santé et dans les logiciels de délivrance des pharmaciens,
 - les données contenues dans les entrepôts créés par certains établissements hospitaliers,
 - les données issues de la biologie, la radiologie, la génomique et l'anatomopathologie.
- **ne sont pas incluses** dans le périmètre du SNDS élargi :
 - les analyses issues des données des études cliniques post-inscription (EPI),
 - les bases de données constituées par des opérateurs privés,
 - les données issues de la recherche clinique.

Un décret en Conseil d'État, pris après avis de la CNIL, viendra préciser les modalités d'application du présent 6°.

Le nouvel article L. 1461-1 I 8° du CSP dispose que « I.- *Le système national des données de santé rassemble et met à disposition :*

(...)

8° Les données à caractère personnel des enquêtes dans le domaine de la santé, lorsque ces données sont appariées avec des données mentionnées aux 1° à 6° ; »

Par conséquent, les données recueillies dans des études de cohorte sont donc également susceptibles d'entrer dans le périmètre du SNDS élargi lorsqu'elles sont appariées avec des données mentionnées aux 1° à 6° de l'article L. 1461-1 I.

Un arrêté listant les bases de données entrant dans le périmètre du SNDS élargi est en cours d'adoption.

Données hors périmètre du SNDS élargi

Le cadre légal ne prévoit pas de dispositions spécifiques encadrant la mise à disposition volontaire de données de santé en faveur du Health Data Hub.

Cette absence de cadre n'obère cependant pas la possibilité de proposer, organiser et encadrer des projets de mise à disposition volontaire de bases de données structurées à des fins de réutilisation dans le Health Data Hub.

Annexe 2 : Participants au Groupe de Travail

| Nom | Prénom | Organisme |
|-------------------|------------|-------------------------------------|
| Augé | Pascale | Inserm transfert |
| Caralp | Mireille | Inserm transfert |
| Clatz | Olivier | SGPI |
| Combes | Stéphanie | Dress, puis Health Data Hub |
| Commailles Chapus | Catherine | OpenHealth - Syntec Numérique |
| De Dreux Breze | Ophélie | Leem |
| Debroucker | Frédérique | Medtronic |
| Glatt | Nicolas | Clinigrid |
| Graciet | Armelle | Snitem |
| Gysembergh-Houal | Anne | APHP |
| Hernandez | Ramon | Sanofi |
| Labarthe | Benoît | APHP |
| Labrune | Laurent | Cegedim |
| Maugendre | Philippe | Sanofi |
| N'Doye | Laetitia | DGE |
| Lambert | Nathalie | Embleema |
| Ordonneau | Yves | Embleema |
| Petre | Anca | 23Consulting |
| Philipponneau | Claire | Leem |
| Pinguet | Jean Marc | Roche |
| Pluquet | Alain | Biomérieux |
| Rosier | Arnaud | Implicit |
| Salette | Jérôme | CERBA |
| Souchay | Henri | General Electric |
| Szeftel | Daniel | Séméia / Carefactory |
| Tirel | Anny | MSD |
| Vainchtock | Alexandre | Heva |
| Villain | Nicolas | Philips |
| Zins | Marie | Cohorte Gazelle / cohorte Constance |

Les Base de données de cohortes Inserm

Les BDD de cohortes sont des BDD de recherche constituées de données en vie réelle collectées dans le cadre de projet de cohorte auprès d'échantillons de patients ou de volontaires (cohortes en population générale) selon un protocole de recherche spécifique. Les cohortes font l'objet de consortium de plusieurs institutions publiques dans lesquels Inserm assure la responsabilité réglementaire de la BDD et organise l'accès aux données. Ces BDD, d'études multicentriques et d'envergure le plus souvent nationale, regroupent des données socio démographiques, cliniques, épidémiologiques, médicales, soins, traitements, qualité de vie, médico-économique, biologie, imagerie, parcours/suivi,... collectées dans le cadre du soin courant et/ou via des questionnaires spécifiques sur des périodes de 5 ans au minimum.

L'accès aux données intervient sur la base d'un projet de recherche revu et validé par le comité scientifique de la cohorte qui veille à l'intérêt de santé publique de ses objectifs et leur cohérence avec la stratégie scientifique de la cohorte ainsi qu'au caractère éthique du projet. Selon les cohortes, cet accès au projet intervient dans le cadre d'un partenariat public/privé (PPP) au projet ou dans le cadre d'un PPP multi privé. L'objectif du PPP est de permettre l'accès aux données pour la communauté de recherche privé et de participer à la pérennité financière de la cohorte : en contrepartie d'un droit à réaliser des projets de recherche et d'accès à d'autres livrables de la cohorte (tableaux de bord, rapports...), le partenaire s'engage financièrement dans la constitution et la maintenance de la cohorte et prend en charge les coûts de réalisation du projet. Les principes des PPP mis en œuvre par Inserm Transfert depuis 2012 sur les cohortes se sont adaptés aux objectifs scientifiques des cohortes et ont répondu à la recherche de leur pérennité financière sur le court, moyen et long-termes.

Les montants d'engagement dépendent de chaque cohorte et de chaque projet (dans le cas d'un partenariat au projet), de la qualité, l'exhaustivité des données, des expertises du réseau de la cohorte, des livrables associés à la valeur des données et du modèle économique de l'industriel demandeur. Des échéanciers adaptés au modèle économique du partenaire sont également proposés pour le paiement de cet engagement.

Les données individuelles, modifiées ou non, éventuellement complétées dans le cadre du projet, restent sous la responsabilité réglementaire du promoteur. Les résultats du projet sont en copropriété entre le promoteur et le partenaire, qui peut acquérir la totalité de la propriété du résultat s'il concerne son produit, en contrepartie du versement d'une somme forfaitaire négociée de bonne foi.

Base de l'AP-HP (France)

L'Entrepôt des Données de Santé de l'AP-HP (données massives), comprend les données pseudo-anonymisées des patients hospitalisés ainsi que les données médico-administratives des séjours. La base existe depuis 2000 et a été compétée régulièrement au gré des autorisations progressives.

Observatoire épidémiologique et suivi populationnel, appelé THIN® - Cegedim

Cegedim équipe aujourd'hui plus de 178 000 médecins dans 11 pays avec ses logiciels de gestion de cabinet et d'aide à la prescription. Ce vaste réseau a permis de recruter en Europe plus de 5 000 médecins volontaires pour participer à un Observatoire épidémiologique et suivi populationnel, appelé THIN® (The Health Improvement Network) recueillant des données de vie réelle en France, UK,

Espagne, Belgique et Roumanie. En France, où 2000 médecins généralistes et 1000 spécialistes libéraux participent volontairement et activement à cet observatoire lancé en 1994, la base de données médicalisée THIN® contient les « Données Médicales Electroniques » de plus de 35 millions de patients sur la totalité de l'historique.

Ces données anonymisées à la source et transmises quotidiennement sont structurées et codées selon un modèle commun européen. Les données longitudinales avec plus de 6 ans d'historique en moyenne par patient contiennent par exemple, le profil patient, les diagnostics et scores de certaines maladies, les modalités de traitement, le parcours médical du patient (en France), les comorbidités, les prescriptions, les actes médicaux prescrits et les résultats biologiques.

Ces données à destination de l'industrie pharmaceutique, des chercheurs et des autorités de santé, sont disponibles sous forme de Dataset, de solutions Analytiques en facilitant l'exploitation ou d'études conventionnelles ou complexes accompagnées par nos experts. L'accès aux données est soumis à un cahier des charges visé et approuvé par notre comité scientifique. Le délai de réactivité, entre l'approbation et la livraison, est très court et de l'ordre de 1 à 3 jours après validation des conditions financières. La valeur ajoutée des données THIN®, soutenue par plus de 800 publications, a été démontrée en R&D, en Market Access, dans le suivi des traitements et l'optimisation des parcours de soins.

Les données THIN® sont la référence des autorités de santé aux UK et en France où le CEPS les exploite au travers de nos solutions analytiques depuis plus de 3 ans.

Base DARWIN (Sanofi), France

La base DARWIN est une plateforme de Sanofi comprenant les données de vraie vie de plus de 400 millions de patients dans plusieurs Pays. Il s'agit d'un écosystème intégré de solutions, capable d'exécuter simultanément des analyses conventionnelles et complexes sur plusieurs ensembles de données. Il combine les données disponibles chez Sanofi provenant de sources internes et externes, complétées par des modèles et des outils analytiques. Les principales sources de données disponibles sur les patients sont celles issues d'OPTUM (USA), THIN (Royaume-Uni, France) IBM .

La Plateforme Darwin peut être très utile pour réaliser par exemple des études pour la R&D, le suivi des traitements, des parcours de soins ou pour comparer des stratégies ou faire des comparaisons internationales. Darwin est une plateforme utilisée par Sanofi mais qui peut être mise à disposition d'acteurs privés ou publics, institutionnels ou académiques. Par ailleurs, la plateforme Darwin est un entrepôt d'algorithmes qu'il est possible de faire évoluer selon les besoins.

Base PRM (Roche), France

La base PRM est une étude pharmaco épidémiologique longitudinale dont les données sont issues des logiciels de chimiothérapie systèmes de plus des 140 centres hospitaliers ou cliniques traitant un minimum de patients atteints de cancers du Sein et Poumon. Créée en 2014, les données collectées regroupent le profil des patients, les diagnostics (localisation tumorale, stade de la maladie), les localisations métastases, et le détail des traitements (ligne, association, séquence) mais aussi la biologie et la participation ou non à un essai clinique. Ces données sont issues des logiciels de chimiothérapie des centres traitant les patients. La représentativité de la prise en charge des patients est assurée par le nombre de centres qui représente plus du tiers des patients traités en France pour un cancer du sein et du cancer du poumon et du fait de leur répartition aux niveaux régionale et du par type d'établissements. Le financement du projet est assuré à 100% par le laboratoire qui a initié le projet. En 2019, les données issues de ce projet ont permis d'obtenir avec le CEPS le premier accord de prix modulé en fonction de l'utilisation d'une immunothérapie.

Modèle GE Healthcare pour l'Imagerie

Les études promues par GE Healthcare produisent des bases ad-hoc destinées au développement et à l'évaluation d'équipements ou logiciels, essentiellement à des fins réglementaires, dans le monde entier.

D'autres études, promues par nos partenaires, permettent notamment l'évaluation de nouvelles applications des équipements dans le cadre d'actes ou de filières de soin.

La taille typique des cohortes de recherche en imagerie est de quelques centaines de patients, qui suffisent pour une large proportion des apprentissages (supervisés) et validations. L'utilisation des données ou résultats issus de ces études est demandée à des fins de recherche ultérieure au sein de GE Healthcare et communication externe. Elles sont utilisées en fonction des réglementations locales et des autorisations obtenues (consentement, finalité).

Pour les autres données, le modèle économique privilégié dans le monde entier est la fourniture de données brutes anonymisées auprès d'établissements de santé, stockées dans un environnement informatique sécurisé, et utilisées selon les autorisations obtenues. Les données ne sont normalement pas mises à disposition en dehors de GE Healthcare.

eICU Collaborative Research Database (MIT)

Cette base est constituée à partir des mesures collectées par le système de télé-surveillance des soins intensifs de Philips. Elle couvre plus de 400 services de soins intensifs, soit plus de 3 millions de séjours patients sur la période 2014-2015. Les données enregistrées comprennent notamment les prescriptions médicales et les mesures biologiques et physiologiques effectuées sur les patients (plus de 800 millions de mesures). Les moyens nécessaires à la constitution de la base, à la traduction des différents formats de données dans une table commune et à son hébergement ont fait l'objet de financements privés (Philips) et publics (NIH grants). L'accès à cette base pour des besoins de recherche est gérée sur la plateforme Physionet par le laboratoire Computational Physiology du MIT, qui s'assure de l'intérêt scientifique du projet et exige une formation initiale (CITI "Data or Specimens Only Research" course). Ensuite, les chercheurs ont accès aux données et peuvent les utiliser librement, avec une citation obligatoire en cas de publication.

Partage de données cardiaques issues d'un l'hôpital universitaire américain avec Philips

Cette base est mise à disposition par l'hôpital à Philips dans le cadre d'une collaboration de recherche. Elle contient des centaines de milliers d'électrocardiogrammes (ECG) et d'échographies cardiaques, ainsi que les codes des pathologies diagnostiquées. L'accès à la base pour un projet de recherche donné est soumis à l'approbation d'un conseil scientifique (« Data Governance Board ») et l'équipe de recherche doit travailler en collaboration avec un chercheur de l'hôpital. Dans ce cadre, la mise à disposition est gratuite, une fois couverts les frais de curation (assistants de recherche, data scientists et DSI) et de stockage des données (sur un cloud privé).

Modèle PSIFAS public-privé (Israël)

Les 4 assureurs santé israéliens commercialisent déjà leurs données (données de remboursement, données des dossiers patients informatisés) depuis une dizaine d'année (cf. fiche Maccabi)

L'initiative Psifas du Ministère de la Santé vise à construire une base unifiée contenant les données de santé anonymisées des 8 millions de patients israéliens.

La construction de cette base est passée par d'importants efforts de normalisation (DPI unique dans les hôpitaux israéliens, fusion des bases des différents assureurs), le financement public de la

production de nouvelles cohortes (données génétiques sur 100 000 patients), l'unification de la gouvernance scientifique des bases de données (comité scientifique au niveau de Ministère) et la mise en place de nouvelles législations pour garantir la confidentialité des données. Le modèle économique cible d'accès aux données est encore en cours de construction mais les principes issus des pratiques de commercialisation des datas des différents assureurs pourraient être reconduits dans la nouvelle tarification :

- Tarifs "tout compris" sans considération pour le type de données demandées selon un prix par patient
- Accès gratuit pour les recherches académiques
- Conditions particulières pour les starts ups (délai de paiement jusqu'à la valorisation commerciale des données, accès aux données gratuites en échange de l'utilisation gratuite des algorithmes produits par les assureurs ayant mis à disposition les données, royalties)

Modèle Maccabi Institute, Israël (semi-privé)

<https://www.maccabitech.com/about/>

Maccabi dispose d'un réseau d'établissements de soins couvrant une population de 2.5 M de patients. Ils disposent d'un dossier médical électronique harmonisé dans tous les établissements, les bases d'imagerie et de biologie sont interconnectées et ils disposent également des consommations de soins. Les patients ont un numéro unique d'identification avec une système solide d'anonymisation. Les données sont saisies en temps réel, le codage est harmonisé et les données manquantes sont très faibles (moins de 1%).

Les modalités d'accès sont les suivantes :

- Un comité d'éthique interne revoit les projets.
- Le degré de réactivité aux sollicitations externes est excellent.
- Les partenariats peuvent prendre la forme d'engagement sur plusieurs années avec conduites de projets *ad hoc* ou d'un accès ponctuel pour conduire un projet unique.
- Il n'y a pas d'accès direct aux données mais production de rapport de données agrégées par l'institut.

Modèle Karolinska Institute, Suède (modèle public)

Système public de registres couvrant toute la population suédoise (20M de personnes). Tous les registres sont sous la responsabilité du National Board of Health and Welfare (organisme public). Toutes les bases sont interconnectées (données cliniques, consommation de soins ...). Les patients disposent d'un numéro d'identification unique qui permet le suivi longitudinal des patients.

Les modalités d'accès sont les suivantes :

- Très bonne réactivité.
- Un Comité interne revoit les projets.
- L'accès est gratuit. Le paiement se fait par projet, en fonction des ressources mobilisées.

- **« Benchmark sur les modèles de valorisation de la donnée en santé », étude VELTYS – AVIESAN, Janvier 2020**
https://cvt.aviesan.fr/wp-content/uploads/2019/12/VELTYS_Benchmark-Valorisation_Donnee_Sante_VELTYS.pdf
- **« How we can place a value on health care data” – EY, Juillet 2019**
https://www.ey.com/en_gl/life-sciences/how-we-can-place-a-value-on-health-care-data
- **“How much is your company’s data really worth?” – Data Republic, Août 2017**
WWW.datarepublic.com
- **« Payer les données de santé, un débat tabou” – Décision santé, 21 juin 2019**
https://s.kmni.eu/t/CeinZGyB-zdzRQxKSUG5xNqkINLiyg-pdf-zC/Dcision_Strategie_Sant_16-18_20_20190620220000
- **A Health Verity, marketplace portée par une structure privée basée à Philadelphie**
<https://healthverity.com/solutions/healthverity-marketplace/>