

Health Data Hub

Charte « Responsables de données »

Les engagements de la plateforme des données de santé vis-à-vis de ses partenaires responsables de données

Préambule

Nous avons la chance de disposer en France d'un patrimoine de données de santé probablement sans équivalent dans le monde.

Ce patrimoine est le fruit d'un travail remarquable conduit par celles et ceux qui, au niveau de l'Assurance Maladie, des établissements de santé ou des instituts de recherche notamment, conçoivent les bases, collectent les données, les regroupent et œuvrent au quotidien pour garantir leur qualité et la pertinence scientifique des usages qui en sont faits.

L'utilisation indépendante de chacune de ces bases par les acteurs qui en sont à l'origine a déjà permis à la France de se distinguer sur le plan international pour la qualité de sa recherche.

Jusqu'où pourrions-nous aller si, en complément des outils et structures existantes, nous étions capables de décloisonner ces bases pour les enrichir entre elles ? Si les chercheurs des centres hospitaliers pouvaient croiser leurs données avec celles d'autres entrepôts de données de santé pour obtenir des cohortes plus nombreuses sur lesquelles tester leurs hypothèses ? S'ils pouvaient les croiser aux données de l'assurance maladie pour élargir leur vision à l'ensemble de la prise en charge, incluant la ville ?

C'est le sens de la démarche conduite par la Plateforme des Données de Santé ou Health Data Hub (HDH), qui n'a pas vocation à se substituer aux bases de données, aux partenariats et aux expertises existants, mais bien à les compléter. Cet outil au service de tous devrait faciliter l'utilisation croisée de données de sources multiples dans un cadre sécurisé et conforme à la réglementation relative à la protection des données. De tels rapprochements de données, pourtant d'un intérêt majeur, prennent aujourd'hui plusieurs années.

Fonctionner ensemble nécessite de définir des règles et des principes communs, qui vont régir les rôles et responsabilités de chacun.

Sans portée juridique ou normative, le présent document, établi en collaboration avec des représentants de l'ensemble de l'écosystème, vise à décrire, sous forme de grands principes, le partage des rôles et responsabilités proposé aux différents partenaires qui agiront en tant que « responsables de données » amenés à alimenter le catalogue de données du Health Data Hub, ainsi que les engagements pris par ce dernier dans ce cadre. Ces principes serviront de socle au travail de conventionnement ultérieur qui sera établi avec chacun d'entre eux.

Cette charte est complémentaire de la charte « citoyens » qui fixe les engagements du Health Data Hub vis-à-vis de la société civile et de la charte « utilisateurs » qui correspond aux conditions générales d'utilisation de la plateforme technologique par les utilisateurs habilités.

Selon les circonstances, les mêmes acteurs pourront endosser la casquette de partenaires « responsables de données » ou d'« utilisateur projet », qui traiteront les données sur la plateforme technologique sécurisée.

Les grands principes

Le catalogue de données

Le Health Data Hub est notamment chargé par la loi de « réunir, organiser et mettre à disposition les données du système national des données de santé mentionné à l'article L. 1461-1 du code de la santé publique et de promouvoir l'innovation dans l'utilisation des données de santé ». Élargi à un grand nombre de sources de données de santé, notamment cliniques, le système national des données de santé (SNDS) ne comprend plus seulement des données médico-administratives, mais également des données de registres, de cohortes de recherche, d'entrepôts de données hospitalières, etc.

L'ensemble de ces données ne vont pas être rassemblées au sein d'un fichier – ou même d'une plateforme – unique. Une sélection des bases parmi les plus prometteuses, pour la plupart préexistantes, sont toutefois répliquées et mises à jour régulièrement au niveau de la plateforme technologique du Health Data Hub. Cette collection de bases compose le “catalogue”, construit de manière itérative et progressive en partenariat avec les responsables de données concernées et dans le respect des grandes orientations données par l'État. Le contenu du catalogue sera porté par arrêté ; il vient en complément de la “base principale” portant sur l'ensemble de la population française et comprenant notamment les données médico-administratives (“SNDS” que l'on pourrait qualifier “d'historique”). Sur le périmètre de la base principale et du catalogue, le Health Data Hub est responsable du traitement, au sens du RGPD, relatif à leur stockage, leur organisation et leur mise à disposition via sa plateforme technologique. Ce rôle est cadré et clarifié par le décret relatif au SNDS, mis à jour prochainement.

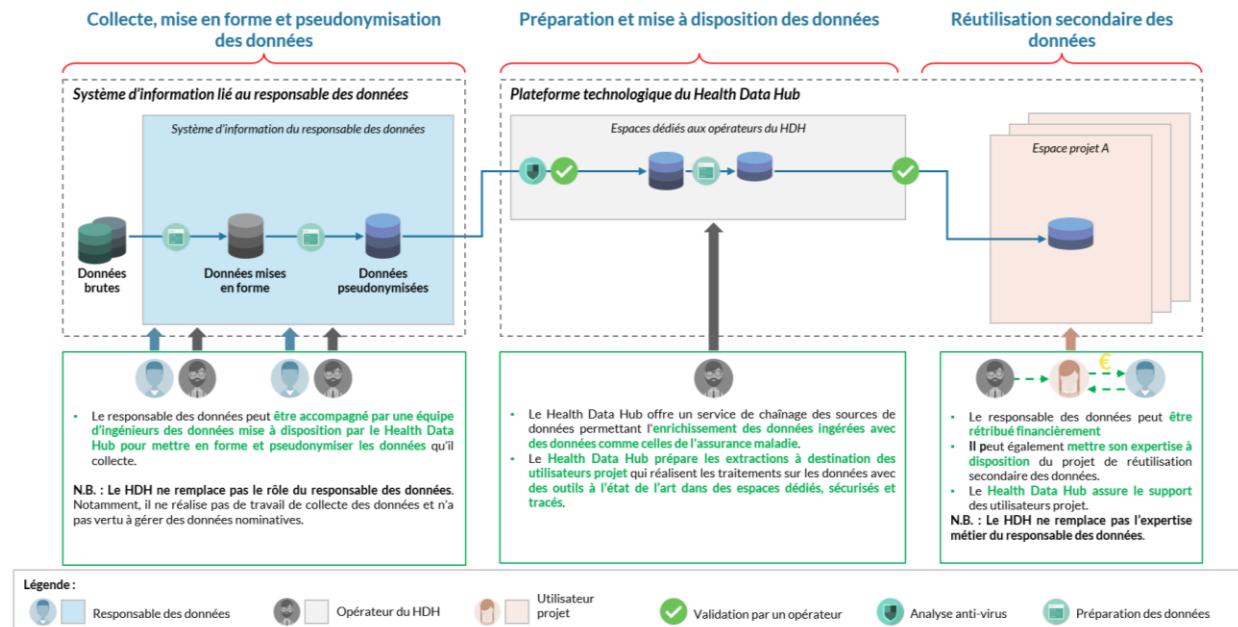
La réPLICATION de ces bases d'intérêt au niveau d'un même dispositif permettra de mutualiser les investissements humains, technologiques et financiers en vue de :

- l'enrichissement pérenne de chacune des bases avec les données médico-administratives ;
- l'agrégation, le chaînage, l'extraction et la mise à disposition des données ciblées via la plateforme technologique du Health Data Hub pour les utilisateurs habilités.

Un tel service est à distinguer des activités suivantes :

- le transfert « à façon » de données du SNDS sur la plateforme technologique du Health Data Hub pour le compte d'un projet ciblé lorsque ces données ne sont pas disponibles au catalogue ou dans la base principale : dans ce cadre, le Health Data Hub fournit une offre technologique sécurisée mais sa responsabilité ne peut excéder la sécurisation de la mise à disposition des données ;
- l'existence de systèmes d'information pour la collecte, la mise en forme et la mise en qualité des données, notamment nominatives : activités essentielles à la consolidation et la pérennisation de notre patrimoine commun de données et pour lesquelles l'appui du Health Data Hub restera possible mais contraint ;
- la mise à disposition de données d'une base source par d'autres moyens et circuits que ceux décrits ci-dessous, notamment à l'initiative du responsable de données concerné en application d'une méthodologie de référence ;
- l'apport souhaitable d'expertise, tarifé ou non, en amont ou en aval d'une demande d'accès à des données, auquel le Health Data Hub ne saurait se substituer.

Dans le processus de mise au catalogue, les systèmes d'information opérés par le responsable de données et le Health Data Hub ont, en réalité, un rôle complémentaire :



Les informations personnelles directement identifiantes sont supprimées des bases avant leur transfert sur la plateforme technologique. Toutefois, des procédures doivent être mises en place pour garantir le libre exercice du droit d'opposition des citoyens sur le périmètre de responsabilité de traitement du Health Data Hub ; la plupart du temps, cette opération devrait être neutre pour le responsable des données.

Le site internet de la Plateforme des Données de Santé expose des informations relatives aux données inscrites au catalogue ainsi que sur l'ensemble des projets réalisés à partir de celles-ci.

Les parties prenantes

- **Les « responsables de données »**

Sont considérées dans le présent document comme « responsables de données » les entités publiques ou privées assurant aujourd'hui la collecte, l'agrégation, la mise en forme et la mise en qualité des données issues du système de soins dans le cadre de cohortes, registres, bases de données et entrepôts de données de santé ou toute autre base relevant du SNDS et ayant vocation à être inscrite au catalogue.

- **Les utilisateurs « projets »**

Les données seront accessibles aux porteurs de projets autorisés par la CNIL ou en conformité à une méthodologie de référence sur le périmètre pertinent pour leur projet. Leur projet aura été au préalable instruit sur le plan de la pertinence scientifique, de sa finalité d'intérêt public et sur son caractère éthique. Ils peuvent être issus de grands organismes de recherche, d'associations de patients et d'usagers, d'institutions, de grands groupes privés comme de start-ups, des journalistes, etc.

- **Le CESREES**

Le CESREES (comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé) est un comité indépendant qui a pour mission d'examiner la méthodologie et le caractère scientifique des demandes d'accès aux données de santé, comme le CEREEES avant lui ; que les traitements soient réalisés ou non sur la plateforme technologique du Health Data Hub. Il doit également juger de l'intérêt public des demandes et fera appel à des compétences diversifiées de nature à garantir une instruction performante de projets très variés. Le Health Data Hub assure son secrétariat et garantit une fluidité du suivi des dossiers qui paraît d'autant plus importante que le nombre de dossiers est voué à augmenter.

- **La CNIL**

La CNIL est l'autorité de protection des données ; elle est compétente dans la procédure d'autorisation des traitements de données de santé. Elle intervient après l'avis du CESREES.

Partage des rôles et responsabilités

- **Etape 0. Conventionnement**

Le Health Data Hub et le responsable de données établissent une **convention**, qui fixe le cadre juridique et contractuel du partenariat.

Cette convention établit notamment :

- L'appui humain et financier qui sera apporté pour appuyer l'exercice de collecte, agrégation, mise en forme, mise en qualité, standardisation, documentation et transfert des données vers le Health Data Hub ;
- Le périmètre et les conditions techniques de mise à disposition des données (périmètre, profondeur, formats, délai et fréquence de mise à jour, métadonnées, etc...) ;
- Les conditions tarifaires d'accès aux données par les utilisateurs ;
- L'offre de service additionnelle (expertise,...) qui sera proposée aux utilisateurs par le responsable de données et mise en avant par le Health Data Hub s'il le souhaite ;

- **Etape 1. Inscription au catalogue**

Le partenaire responsable de données collecte, agrège, met en forme, met en qualité, standardise et documente ses données, avec l'appui financier ou humain du Health Data Hub si nécessaire et selon les conditions prévues dans la convention. Ils mettent ensuite conjointement en place le circuit de pseudonymisation et le flux d'ingestion des données vers le Health Data Hub, qui se charge ensuite de l'appariement des données avec celles de la base principale. Le Health Data Hub présente les données et l'offre de service du responsable de données, le cas échéant, au sein de son catalogue. L'utilisateur « projet » pourra alors, s'il le souhaite, contacter directement le responsable de données afin d'échanger des informations sur la base. Le Health Data Hub encouragera ces échanges entre « responsable de données » et utilisateur « projet », bien souvent essentiels pour le cadrage du projet.

- **Etape 2. Gestion des demandes d'accès**

L'utilisateur « projet » fait ensuite une demande pour accéder à certaines données proposées au catalogue. Le Health Data Hub en informe le responsable de données qui peut émettre des remarques de toute nature sur la demande. Le CESREES est saisi du dossier pour évaluer la pertinence scientifique et l'intérêt public du projet. Il intègre les remarques du responsable des données dans son analyse et peut l'auditionner. Une fois l'avis du CESREES rendu, le dossier peut être soumis à la CNIL pour autorisation.

- **Etape 3. Utilisation des données**

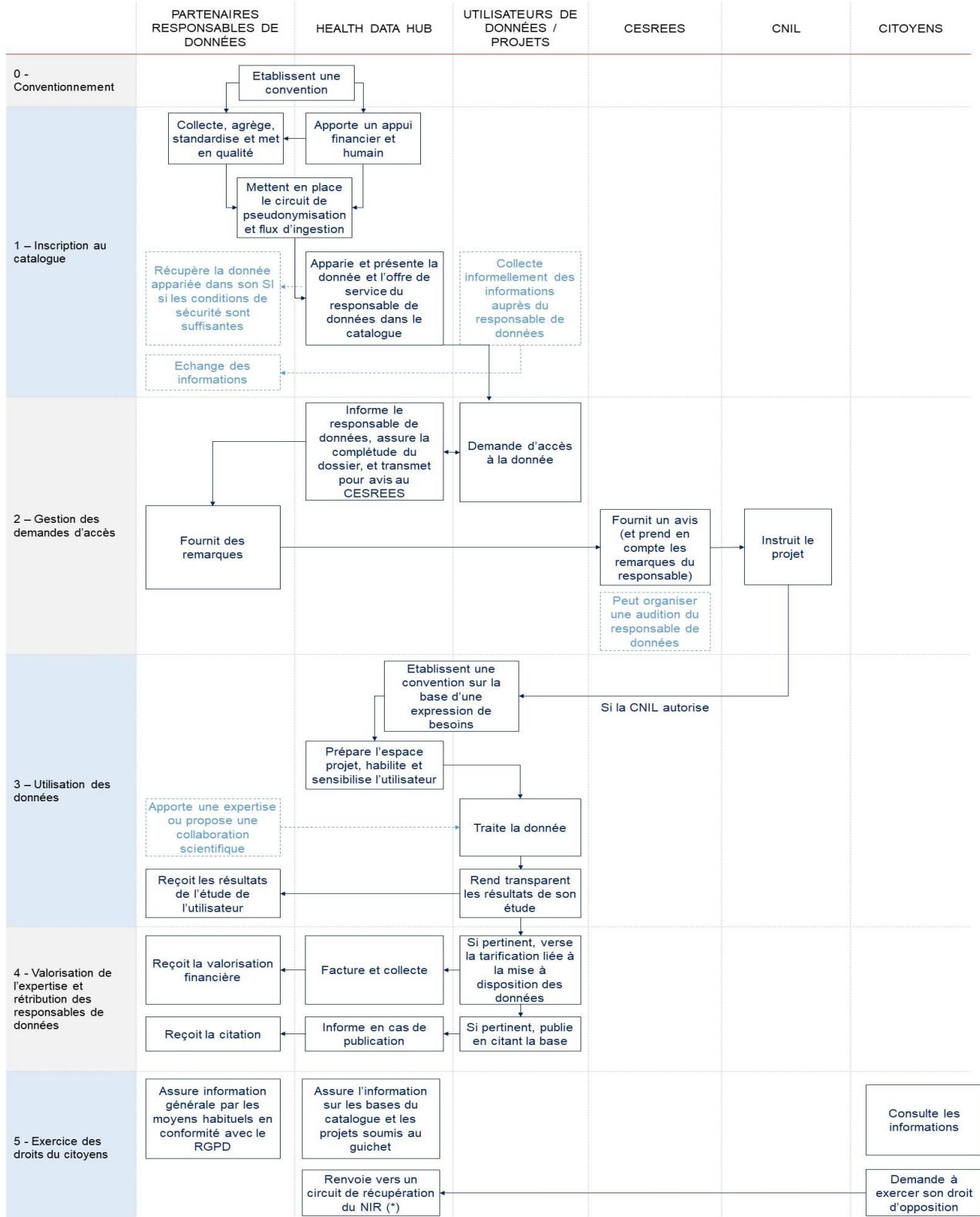
Une fois la demande d'accès aux données approuvée par la CNIL, le Health Data Hub et l'utilisateur « projet » établissent une convention sur la base d'une expression de besoins. Le Health Data Hub prépare ensuite l'espace projet au sein de sa plateforme technologique, permettant à l'équipe projet, une fois habilitée, de traiter la donnée sous de hautes conditions de sécurité. L'utilisateur rend ensuite transparents les résultats de son étude dans le répertoire public opéré par le Health Data Hub, qui en informe le responsable de données.

- **Etape 4. Valorisation de l'expertise et rétribution des responsables de données**

Le Health Data Hub facture et collecte les revenus de la tarification liés à la mise à disposition des données. Il reverse au partenaire responsable de données sa quote-part, laquelle permet de couvrir une partie de l'effort financier de collecte, mise en forme, mise en qualité, standardisation et documentation de la base initiale. Les modalités de calcul sont précisées dans la convention et les tarifs sont approuvés par le conseil d'administration du Health Data Hub après avis de son comité d'audit. Si l'étude de l'utilisateur « projet » parvient à des conclusions intéressantes et qu'il décide de publier des articles à son sujet, il devra citer la base utilisée.

- **Etape 5. Exercice des droits du citoyen**

Le responsable de données fournit une information aux personnes dont il collecte les données par ses moyens habituels en conformité avec le RGPD et la loi informatique et liberté. Cette information permettra aux personnes de se reporter au site internet du Health Data Hub si elles veulent davantage d'information sur les bases présentes au catalogue et les projets utilisant les données. Ces démarches d'information seront complémentaires de l'information directe des personnes qui sera réalisée par l'Assurance Maladie sur le SNDS en général. Lorsqu'un citoyen souhaite exercer son droit d'opposition sur le périmètre sur lequel le Health Data Hub est responsable de traitement, ce dernier met en oeuvre un dispositif permettant le recueil de son NIR tout en garantissant le cloisonnement entre son identité directe et les données de santé. Cette procédure permettra l'effectivité du droit d'opposition sur le catalogue sans intervention du responsable de données la plupart du temps.



(*) = pour les bases concernées

Bleu clair pointillé = étapes optionnelles

Les engagements du Health Data Hub

Le Hub s'engage à soutenir les responsables de données dans la collecte, la consolidation et la mise en valeur de leur patrimoine.

1. Appui à la collecte, la standardisation et la documentation des données

- Le Health Data Hub s'engage à appuyer financièrement et humainement les responsables des données dans la conduite des différentes actions nécessaires au **transfert et à la mise à disposition des données** : extraction, dé-identification, standardisation, documentation, création d'échantillon, mise en place des flux d'ingestion. Les modalités précises d'accompagnement seront fixées au cas par cas dans la convention qui sera rédigée en collaboration avec chacun des responsables de données.
- En mutualisant ses investissements humains, technologiques et financiers, il s'engage à prendre en charge la procédure d'accès harmonisée, la mise en œuvre d'appariements et la création d'extractions, la gestion du support, de la sensibilisation et de l'information des utilisateurs projets ainsi que les coûts de stockage, de calcul et de mise à disposition via une plateforme technologique à l'état de l'art. Certaines de ces activités peuvent être tarifées.
- Aux côtés des responsables des données, le Health Data Hub s'engage à contribuer à la **diffusion de standards** internationaux et à une meilleure interopérabilité, afin de permettre des agrégations et chaînages de qualité.
- Le Health Data Hub promeut enfin les initiatives visant à coordonner les **acteurs locaux** d'un territoire dans le but de créer de nouvelles bases.

2. Fourniture d'un hébergement à l'état de l'art et hautement sécurisé

- Dès leur arrivée sur la plateforme technologique, les données sont placées sous la responsabilité du Health Data Hub. Celui-ci s'engage à **garantir un très haut niveau de sécurité** et de protection des données. Le Health Data Hub fournit une plateforme technologique sécurisée garantissant un hébergement certifié "hébergement de données de santé" et conforme aux textes applicables, notamment l'arrêté relatif au référentiel de sécurité du SNDS. Il assure un suivi des conditions de sécurité grâce à une gouvernance de cybersécurité impliquant le haut fonctionnaire de défense et de sécurité des ministères sociaux et l'ANSSI (Agence Nationale de Sécurité des Systèmes d'Information).
- La plateforme permet des **traitements à l'état de l'art** pour les utilisateurs « projet » mais également pour les responsables de données qui voudraient réaliser certains de leurs traitements de mise en forme, mise en qualité, standardisation et documentation sur la plateforme technologique du Health Data Hub. L'offre logicielle permettra le développement des codes et algorithmes dans le respect des standards les plus récents (R, python, spark..) et sera complétée par des capacités de stockage et de calcul élastiques. En cible, une place de marché permettra également aux éditeurs de partager les outils pertinents avec la communauté et d'ainsi favoriser la mutualisation.

- Le Health Data Hub s'engage à **tracer toutes les actions réalisées** sur les données afin d'en permettre l'audit.

3. Accompagnement dans la mise en conformité RGPD

- Au-delà des aspects relatifs à la protection des données, le Health Data Hub est soumis à l'obligation de **transparence** sur tous les projets menés avec des données du SNDS en application de l'article L.1461-3 du code de la santé publique. Indépendamment de leur lieu de réalisation, les finalités des projets mobilisant des données du SNDS sont rendues publiques ainsi que leurs méthodes et résultats.
- Le Health Data Hub s'engage à veiller au **respect des droits des citoyens, à faciliter l'exercice de ces droits et à leur bonne information**, en articulation avec les responsables de données. Dans la mesure où le Health Data Hub ne stockera jamais d'information susceptible d'identifier directement une personne mais que l'exercice des droits nécessite le NIR, une procédure garantissant le cloisonnement entre l'identité directe de la personne et ses données de santé doit être mise en place. Cette procédure devrait être transparente pour le responsable de données.
- Les demandes d'accès au catalogue de données sont évaluées par le CESREES ; le responsable de données est systématiquement informé par le Health Data Hub d'une demande portant sur sa base afin de permettre à son **comité scientifique**, le cas échéant, de formuler des remarques sur un dossier précis. Cela doit garantir que le CESREES ne donne pas un avis contrevenant aux finalités convenues avec les volontaires lors du recueil de leurs données dans le cas des cohortes par exemple. De même, il s'agira de ne pas donner un avis favorable à un projet qui semble pertinent mais qui ne serait pas en réalité faisable du fait du protocole de collecte initial des données, de ses biais potentiels ou de sa qualité plus généralement. Ce point sera précisé dans le règlement intérieur du CESREES.
- **L'intérêt public** de chaque projet et le caractère strictement nécessaire des données seront évalués par le CESREES, qui comprendra des experts scientifiques, éthiques, juridiques, et des représentants des associations de patients. La doctrine de l'intérêt public sera ainsi formalisée par le CESREES.

4. Valorisation et rétribution du travail et de l'expertise du responsable des données

- Le Health Data Hub assure l'**enrichissement systématique** des bases du catalogue avec les données de la base principale. Les partenaires responsables de données peuvent récupérer leurs données chainées au SNDS dans leur système d'information propre, si celui-ci respecte le niveau de sécurité requis ; ce transfert devra s'effectuer en conformité avec la réglementation en vigueur.
- Le transfert des données au sein de la plateforme technologique du Health Data Hub pourra être opéré dans un délai d'un an maximum, afin d'assurer aux responsables de données un **temps d'avance** pour utiliser leurs bases à des fins de publication académique avant leur partage à d'autres utilisateurs « projet ». Les modalités seront à définir dans le cadre de la convention.
- Le Health Data Hub renforce la **visibilité** des bases de données de recherche et augmente ainsi les citations mentionnant leur ré-utilisation, celles-ci étant exigées des

utilisateurs « projet ». Ces indicateurs confèrent aux équipes en charge de leur collecte, mise en forme et mise en qualité un argument dans le cadre de la négociation de financements ou pour leur propre notoriété. Comme ils sont systématiquement informés en cas de demande d'accès portant sur leur base, ils ont l'opportunité de nouer de nouveaux partenariats. Ils sont également informés de chaque publication de résultats obtenus grâce à leurs données.

- Le Health Data Hub s'engage à favoriser voire accompagner la **mise en place de partenariats** entre les responsables de données et les utilisateurs « projet ». Il leur offre la possibilité de tarifer l'accès direct aux données à des utilisateurs « projet » privés, de l'expertise ou encore des analyses réalisées par le responsable lui-même. En mettant en avant l'expertise des responsables de données, les collaborations seront encouragées et un **partage de la valeur** pourra être négocié entre utilisateur « projet » et responsable de données dans le cas où ils développeraient une innovation conjointement.
- **Aucune exclusivité** n'est requise dans le cadre du partenariat entre les responsables de données et le Health Data Hub. Si le responsable de données et l'utilisateur préfèrent travailler directement ensemble, ils seront libres de le faire. Ce partenariat doit être établi dans le respect de la réglementation en vigueur, notamment s'agissant du traitement des données personnelles.