

Cahier des charges pour un hub interrégional

Proposition G4, GGEST, HUGO

1. Introduction

Les Groupements de Coopération Sanitaire (GCS) HUGO, G4 et GGEST ont identifié la structuration et l'usage des données massives en santé comme un sujet stratégique majeur, pour lequel l'échelon interrégional est particulièrement pertinent. Le rapport de préfiguration du *Health Data Hub (HDH)* mettait l'accent sur la constitution d'un réseau national basé sur des "Hubs locaux chargés de mettre en oeuvre l'offre de service" pouvant "administrer la plateforme technologique [du HDH]". Ce même rapport précisait que "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 de données".

Les trois GCS ont travaillé conjointement pour élaborer un cahier des charges de hub interrégional à l'initiative des établissements hospitaliers permettant d'une part, une collaboration scientifique hospitalo-universitaire renforcée et d'autre part, le principe du partage des données. En effet, nos trois GCS représentent plus de 33 000 lits et places au sein de 16 Centres Hospitaliers Régionaux et Universitaires. Ils ont entamé une démarche commune de structuration en réseaux interrégionaux afin de disposer d'une animation scientifique et technique ainsi que d'une gouvernance qui permettent de constituer des hubs d'intégration et d'exploitation des données massives. Ces hubs interrégionaux contribuent à la stratégie nationale de traitement des données de santé. Ce projet qui s'étend sur l'ensemble du territoire des 3 groupements interrégionaux impliqués pourra être préfigurateur pour les autres régions du territoire national.

Cette collaboration a déjà pris la forme d'un projet scientifique soumis à l'appel à projet flash COVID-19 de l'ANR en mars 2020 : le projet PAndémie Covid19 et TRAjectoires de soins dans les CHU (PATRAC). Ce projet a été retenu sur la liste complémentaire et va être resoumis à l'appel d'offre en cours. L'objectif de cette étude multicentrique s'appuyant sur les entrepôts de données de santé (EDS) de 16 établissements est de développer et valider des

indicateurs permettant d'évaluer l'impact global de la réorganisation des hôpitaux sur la prise en charge des patients, en développant des modèles numériques prédictifs et rétrospectifs de soins et d'organisation du système de santé. Le projet a pour but de générer des connaissances sur l'impact de la pandémie COVID-19 en particulier sur la trajectoire de santé de patients non touchés par le COVID et potentiellement confrontés à des difficultés d'accès aux soins. Au delà de la question scientifique, ce projet permet aussi de franchir un premier stade de constitution d'un réseau interrégional en permettant d'amorcer les étapes nécessaires à l'interopérabilité des données entre les établissements.

2. Les missions

Les hubs interrégionaux ont pour vocation la collection, la structuration, l'exploitation et la valorisation des données de santé issues de plusieurs centres au sein d'une organisation dotée d'une gouvernance garantissant la pertinence scientifique et la conformité éthique et réglementaire des projets. Chaque centre partenaire d'un hub interrégional dispose localement d'un entrepôt de données de santé.

Ainsi les hubs interrégionaux doivent répondre à 3 catégories d'objectifs de façon coordonnée :

1. Structurer et accompagner les déploiements des EDS dans une optique d'interopérabilité
2. Promouvoir, évaluer et accompagner les projets académiques ou industriels de traitement des données de santé en s'appuyant sur les écosystèmes hospitalo-universitaires
3. Représenter un interlocuteur permettant les collaborations avec d'autres structures qu'elles soient des hubs interrégionaux, structures nationales comme le HDH ou des consortiums de recherche internationaux comme OHDSI ou EIT Health.

2.1. Structurer et accompagner les déploiements des EDS dans une optique d'interopérabilité

L'exploitation des données de santé au sein d'un établissement hospitalier implique un décloisonnement de ces données par la mise en place d'un EDS. Les nomenclatures des données et leurs méthodes de recueil étant propres à chaque établissement, il est nécessaire de s'appuyer sur les expertises locales médicales et techniques pour qualifier ces données. Dans une optique de collaboration multicentrique au niveau interrégional, ces EDS doivent répondre à des critères d'interopérabilité et de sécurité optimale.

Outre la structuration technique, les hubs locaux ont pour vocation de fédérer les équipes multidisciplinaires en charge de l'exploitation des entrepôts locaux. En incitant d'une part à une harmonisation des méthodes et une mutualisation synergique des compétences.

2.2. Promouvoir, évaluer et accompagner les projets académiques ou industriels de traitement des données en s'appuyant sur les écosystèmes hospitalo-universitaires

La mise en place d'une gouvernance au niveau de ces hubs interrégionaux permettra de promouvoir la recherche scientifique basée sur l'exploitation des données de santé grâce à plusieurs types d'action.

En fédérant les experts disponibles au sein de comités scientifique et éthique, les hubs interrégionaux seront en capacité d'évaluer les projets de recherche multicentriques, soumis au niveau interrégional. L'objectif des hubs interrégionaux est ainsi d'apporter un éclairage expert sur les pertinences scientifique et stratégique, la faisabilité technique des projets, l'adéquation méthode/moyens et sur le respect des règles éthiques et réglementaires.

Ces structures ont également un objectif d'accompagnement et de valorisation des projets d'exploitation des données de santé grâce à la mise à disposition d'experts pluridisciplinaires tant sur le plan méthodologique et technique que médical. Les Hubs interrégionaux sont présents pour permettre le développement des compétences et des expertises en matière de traitement et d'utilisation des données, notamment via la mise en place de formations et de diplômes en lien avec les universités et les EPST pour créer des viviers de professionnels et acculturer les investigateurs.

La gouvernance mise en place au sein de ces hubs est en capacité de produire des appels à projets de recherches scientifiques permettant la promotion des projets d'exploitation des données de santé.

De par les missions confiées aux CHU, ils s'inscrivent dans les écosystèmes académiques comprenant les EPST (universités, CNRS, INSERM, INRIA...) ; ce maillage scientifique historique permet la mise en place synergique de grands axes de recherche et de valorisation. Cette expertise scientifique est mise à disposition pour éclairer les choix en matière de stratégie, d'organisation et de parcours des soins sur le territoire interrégional.

2.3. Représenter un interlocuteur permettant les collaborations avec d'autres structures qu'elles soient des hubs interrégionaux, structures nationales comme le HDH ou des consortiums de recherche internationaux comme OHDSI ou EIT Health

Ces hubs dotés d'une gouvernance commune ont pour objectif d'être des interlocuteurs privilégiés, représentant les centres qu'ils regroupent, auprès d'autres structures, qu'elles soient locales, nationales ou internationales.

Ce positionnement s'inscrit pleinement dans la perspective décrite dans le rapport de préfiguration du HDH: "Les hubs locaux auraient également pour mission de mettre en réseau les compétences et animent 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.”

Le principe de subsidiarité vis-à-vis des missions du HDH pour les projets à dimension interrégionale sera appliqué. La dimension territoriale des Hubs et leur proximité avec les écosystèmes régionaux leur permet d'être force de proposition et d'expérimenter de manière agile les modèles économiques futurs autour de la donnée de santé (réponse coordonnée aux appels à projets, réponse coordonnée aux appels à projets, recherche de partenariats académiques et industriels académiques et industriels, contractualisation, valorisation économique).

3. Les caractéristiques techniques des Hubs interrégionaux

Un socle commun de fonctionnement est envisagé au sein de chaque GCS afin d'élaborer des démarches parallèles et néanmoins coordonnées pour collecter, structurer, partager et exploiter les données sur les périmètres GHT, sectoriel public-privé et de soin primaire. On pourra s'appuyer par exemple sur les plateformes d'interopérabilité orientées soin et partagées comme PREDICE (Hauts-de-France) ou les initiatives de partage des données ville-hôpital (bases de données de soins primaires).

Le niveau local de traitement des données pourrait s'envisager au sein de structures interrégionales mutualisant les ressources et les compétences des CHU ainsi que de leurs établissements périphériques. Ce niveau de granularité régional répondra alors aux besoins locaux de traitement des données, tout en facilitant les échanges avec le HDH.

Cette dynamique locale favorise l'émergence de nouvelles initiatives qui pourront alimenter une feuille de route nationale, à l'instar de spécifications de jeux de données à harmoniser. Une approche consistant à répondre à une question de recherche permet de se mettre dans un logique de data set harmonisé, de capitaliser les efforts de qualification des données à l'échelle nationale et pour les projets à venir. Ainsi, le HDH pourrait jouer un rôle d'incitation et de facilitation dans la mise en place d'une interopérabilité sémantique et syntaxique des données ainsi que dans l'installation d'infrastructures techniques régionales.

3.1. Etape 1 : mutualisation de standards d'interopérabilité pour les données et les algorithmes

Le déploiement d'algorithmes d'analyse de données dans chacun des hôpitaux participants nécessitera des données "FAIR" (Findable, Accessible, Interoperable, Reusable). Plusieurs tâches sont identifiées parmi lesquelles (1) l'harmonisation des données hétérogènes entre les hôpitaux, (en particulier les données de résultats de laboratoire) et (2) la sélection et le nettoyage des ensembles de données à traiter. Ces tâches nécessiteront une forte coordination entre les hôpitaux, ainsi qu'une charge de temps pour la gestion des données.

En effet, plusieurs terminologies sont largement utilisées au sein des établissements hospitaliers, telles que la classification internationale des maladies (CIM) pour les diagnostics ou la classification anatomique, thérapeutique et chimique (ATC) pour les médicaments. Cependant, l'encodage des résultats de laboratoire reste un problème dans ce

contexte et l'interopérabilité sémantique (utilisant des terminologies communes) de ces résultats est effectivement encore rarement atteinte. En conséquence, le même paramètre stocké dans un système est nommé différemment dans deux hôpitaux différents et il faut un travail manuel long et réalisé localement par un expert pour établir une correspondance entre toutes les terminologies existantes. Les outils développés pour conduire ce travail de mapping seront mutualisés entre les 3 GCS pour faciliter cette tâche d'harmonisation. Plus généralement, des outils identiques de nettoyage, de qualification et de transformation des données seront également mis à la disposition de chaque centre.

Sur le plan pratique, les données seront collectées auprès de chaque EDS avec un périmètre socle basé sur les catégories principales de données du dossier patient (diagnostics, actes, médicaments, résultats de biologie, données textuelles). Ce périmètre sera progressivement étendu à d'autres sources de données expertes (imageries, données - omiques...) et territoriales (GHT, télémédecine...) avec le même enjeu d'harmonisation.

Pour faciliter l'interopérabilité syntaxique (en plus du travail d'interopérabilité sémantique décrit ci-avant), un modèle de données commun simplifié sera proposé et partagé entre les membres des 3 GCS. De plus, chaque centre disposera des outils logiciels de construction de cohortes et d'extraction des variables d'intérêt. En outre, pour permettre l'harmonisation des analyses, les algorithmes développés pourront s'appuyer sur une norme de "message" commune, voire directement sur le modèle de données commun.

Cette stratégie permet de répondre à deux modalités de fonctionnements collaboratifs en répondant au mieux aux besoins des projets et aux spécificités locales, qu'elles soient techniques ou organisationnelles :

- Le traitement fédéré des données : cela permettra d'effectuer les analyses dans chaque centre de traitement des données avec un effort marginal réduit ainsi que d'assurer la reproductibilité de cette tâche. Les résultats agrégés obtenus au sein de chaque centre pourraient ensuite être partagés avec les autres centres, dans le respect de la confidentialité des données.
- le traitement de données partagées en s'appuyant sur des infrastructures techniques mutualisées à l'instar de la plateforme interrégionale de santé HUGO

Enfin, les 3 GCS travaillent activement, dans le cadre de projets de recherche en partenariat avec des EPST, au développement d'algorithmes d'apprentissage fédéré (*federative learning*) permettant la construction de modèles à partir de données issues de sites multiples mais sans export de données ainsi qu'à des méthodes de dé-identification, de cryptotatouage et d'avatardisation des données; ces méthodes permettent de construire des modèles robustes tout en apportant un haut niveau de garantie sur le plan de la confidentialité et de la cybersécurité.

Au-delà du cas des EDS hospitaliers, ces méthodes devraient jouer un rôle important dans la modélisation des données issues d'objets connectés ainsi que dans la construction d'essais cliniques interventionnels innovants utilisant des "*Electronic Patient-Reported Outcomes*".

3.2. Etape 2 : convergence des infrastructures

Les besoins locaux, notamment au sein de chaque CHU, comprennent l'installation d'infrastructures sécurisées d'hébergement et de traitement des données : ces besoins locaux dépassent le simple cadre de la réutilisation de données et correspondent plus généralement à un besoin de traitement des données en recherche biomédicale, qu'ils s'agissent de données issues de la recherche prospective (cohortes, essais randomisés, etc.) ou de la recherche rétrospective (entrepôts de données de santé, données issues du SNDS, etc.).

Les 3 GCS considèrent que la granularité du Hub régional pourrait correspondre à moyen terme au niveau local de stockage et de traitement des données, ce dernier étant réalisé dans le cadre de bulles sécurisées (environnements d'exploitations) disposant des principaux outils statistiques ainsi que de logiciels service gérant les études et les utilisateurs. Ces environnements sécurisés (authentification forte, traçabilité, gestion des sauvegardes, environnement certifié HDS) disposeront de fait de certaines capacités de calcul "scalables" en propre.

Dans cette perspective, le HDH jouera un rôle central dans l'accompagnement des projets de grande ampleur nécessitant un appariement complexe avec le SNDS et impliquant des besoins de stockage et de calcul de grande ampleur. Ainsi, l'articulation de nos hubs régionaux avec le HDH nous semble un point essentiel pour développer des projets ambitieux dans le champ de la réutilisation de données.

4. La gouvernance

4.1. Le pilotage stratégique

Le hub interrégional constitue une plateforme de mutualisation d'expertises et de technologies associant plusieurs CHU, et qui met en jeu des données à haute sensibilité. A ce titre, les décisions stratégiques concernant sa constitution, son architecture et son usage doivent être concertées et arbitrées au sein d'une instance légitime représentant les gouvernances des établissements impliqués.

L'existence d'une structure juridique de coopération entre les partenaires, le plus souvent un GCS, apparaît comme une condition facilitante, si ce n'est nécessaire, de la mise en place d'un hub interrégional.

Elle permet de garantir la sécurité juridique du hub interrégional, grâce à l'implication d'une instance légitime pour la prise de décision, l'assemblée générale du GCS.

Selon le périmètre du hub (régional ou interrégional), et le contexte dans lequel s'inscrit ce partenariat, la gouvernance stratégique peut prendre différentes formes. Cette fonction peut être dévolue directement à l'une des instances de gouvernance du GCS ou être confiée à un comité ad hoc.

Exemple 1. le comité de pilotage de la plateforme interrégionale de données de santé du GCS HUGO

Le comité de pilotage définit les objectifs, la structuration, les modes de fonctionnement, la politique scientifique de la plateforme partagée des données de santé. Il éclaire l'assemblée générale du GCS HUGO pour les décisions stratégiques.

Il est composé :

- de l'administrateur du GCS HUGO, président
- du coordonnateur médical du GIRCI
- de trois experts "données massives" désignés par l'administrateur
- d'un binôme désigné par chaque établissement partenaire (un représentant de la gouvernance et un expert recherche et/ou données)

Il se réunit tous les deux mois.

Exemple 2. Le Collège G4 assure les missions de pilotage stratégique.

Il est composé :

- de l'administrateur du GCS G4 ;
- des directeurs généraux, présidents de CME et doyens de la faculté de médecine des 4 CHU membres du G4 ;
- de la directrice et du conseiller médical du G4
- des experts données massives de santé émanant du groupe de travail lorsque ce point est inscrit à l'ordre du jour.

Exemple 3 : le comité de pilotage du réseau PAISaGE du GCS GGEST

Le comité rassemble deux représentants par établissement membre du GCS GGEST. Il est présidé par le coordonnateur du réseau et il rend compte à l'assemblée générale du GGEST, qui est composée des directeurs généraux, présidents de CME et doyens de faculté de médecine des 6 CHR/U membres du GGEST.

4.2. Une instance scientifique et éthique

Le Comité Scientifique et Éthique (CSE) est une instance essentielle dans la construction du nœud régional ou interrégional.

Le CSE examine les projets de recherche et d'études mobilisant des données issues de plusieurs EDS composant le hub interrégional, impliquant éventuellement des données issues d'autres bases (ex : SNDS, registres, cohortes). Ils sont menés par les établissements du Hub ou des partenaires extérieurs (académiques ou privés).

Le CSE répond à tout ou partie des missions suivantes :

- Il définit et valide les évolutions de la politique d'accès et d'exploitation.
- Il décline au niveau local les règles de bon usage en s'assurant que les protocoles d'étude sont compatibles avec l'usage autorisé par la CNIL : il est informé des demandes en cours avec la possibilité de réserver le traitement.
- Il a qualité pour homologuer les personnes qui réalisent un traitement sur les données.
- Il donne autorisation d'accès aux données après validation des usages selon les finalités et les process définis par l'instance stratégique, et selon les autorisations CNIL délivrées

Le CSE peut également être saisi pour fournir un conseil scientifique et/ou éthique en rapport avec l'accès et l'utilisation des données des entrepôts de données du hub interrégional.

Sa composition est définie par l'instance stratégique du hub interrégional.

Il rassemble des compétences complémentaires : médicales (médecins cliniciens, pharmaciens, biologistes, radiologues, médecins DIM, spécialistes d'informatique médicale...), scientifiques (biostatisticiens / méthodologistes, représentant de la direction de la recherche...), éthiques, juridiques, le délégué à la protection des données et les porteurs de l'EDS. Il intègre de plus une représentation des usagers.

Les Conseils Scientifiques et Ethiques des hubs interrégionaux ont vocation à être les interlocuteurs privilégiés du CESREES.

Un principe de subsidiarité pourrait être institué, en vertu duquel les CSE locaux auraient pour mission de statuer sur les projets promus dans son périmètre territorial et engageant les seules données des EDS reliés au hub interrégional, le cas échéant appariées avec des données d'un partenaire.

Pour les autres projets engageant des données provenant de plusieurs entrepôts du hub interrégional, le CSE du hub interrégional serait obligatoirement consulté pour un avis préalable avant instruction par le CESREES dans un délai compatible avec une instruction effective et la bonne réalisation des projets.

Enfin, le besoin d'efficience et de facilitation des démarches réglementaires justifierait l'élaboration d'une méthodologie de référence (MR) adaptée au traitement des données sur les hubs interrégionaux et adossée à la gouvernance décrite ci-avant. Cette MR accélérerait grandement la mise en oeuvre de recherches innovantes (par exemple *registry randomized clinical trials*) qui nécessitent l'appariement de données de cohortes ou d'EDS hospitaliers avec celles provenant du SNDS ou d'autres responsables de données.

4.3. Information et droits des citoyens

Les hubs interrégionaux portent une attention particulière aux usagers concernés au premier chef par l'usage des données de santé.

Ils inscrivent leur action dans le cadre réglementaire fixé par le RGPD et décliné par la CNIL au travers des autorisations qu'elle délivre tant pour chacun des EDS qui composent le hub interrégional que pour les projets de recherche menés à partir des données qu'ils contiennent.

Ils associent les usagers dans leurs instances.

Ils reprennent à leur compte les quatorze engagements définis par la plateforme nationale des données de santé dans sa charte citoyenne.