



Health Data Hub : nouvelles opportunités pour les études en vie réelle et les projets innovants en santé

Le jeudi 18 mars 2021, la conférence sur le Health Data Hub (HDH) organisée par l’AFCROs dans le cadre du Printemps de la Recherche Clinique a permis des échanges ouverts et s’inscrivant dans la durée avec les industriels et CROs, alors que la publication du décret sur son fonctionnement est attendue en avril. Avec les témoignages de la BNDMR et du laboratoire Roche, parmi les premiers à miser sur la nouvelle dynamique initiée par le hub afin de poursuivre la mise en œuvre de projets innovants.

La création du Health Data Hub est l’une des mesures phares de la loi relative à l’organisation et la transformation du système de santé, qui visait à positionner la France comme un leader dans l’usage des données de santé. Cette promesse tellement séduisante est-elle en train de se concrétiser ? C’est la question clairement posée par Mohammed Bennani, Président de Qualees, Membre du groupe de travail RWD de l’AFCROs, qui animait cette 6^{ème} édition de l’atelier sur le Health Data Hub, en compagnie de Jean-Vannak Chauny, Health Economics Senior Manager, Amgen, et de Benoît Thomé, Chef d’Entreprise Median Conseil, Membre du Groupe de Travail RWD de l’AFCROs. Les échanges ont été particulièrement riches et animés avec Emmanuel Bacry, Directeur Scientifique du Health Data Hub, Arnaud Sandrin, Directeur opérationnel de la Banque Nationale de Données Maladies Rares (BNDMR), et Rémy Choquet, Head Medical Evidence & Data and Personalized Healthcare Platforms. Avec un enjeu décisif pour le futur : répondre aux problématiques des producteurs de données et des utilisateurs potentiels pour une mise en route fluide et que le hub porte rapidement ses fruits.

Etat d’avancée du Health Data Hub

Le Health Data Hub est né d’une idée retransmise dans le rapport Villani sur l’Intelligence Artificielle. Sa création, actée par la loi Ma santé 2022, est effective depuis le 9 décembre 2019. « Nous avons passé notre premier anniversaire, » note Emmanuel Bacry, son directeur scientifique, également chercheur au CNRS, qui est aujourd’hui entouré d’une soixantaine de personnes au sein du hub pour faire évoluer le système. « Nous avons certes d’incroyables bases de données en France, mais celles-ci sont sous utilisées du fait de la dispersion des données et d’une gouvernance lourde, voire discrétionnaire, » note-t-il. Un autre sujet est le manque d’interopérabilité des données, tant sur le plan technique que de la normalisation.

L’objectif du hub ? Garantir un accès aisé, unifié, transparent et sécurisé aux données de santé pour améliorer la qualité des soins et l’accompagnement des patients. « Nous remplissons le rôle d’un guichet unique donnant accès à d’importantes bases de données constituées sous forme de catalogues sur une plateforme sécurisée à l’état de l’art » décrit-il. Le Health Data Hub remplit également « une mission d’animation de l’écosystème académique et industriel », ainsi que de « mise en relation des compétences ».



Avec quel agenda ? La publication du décret relatif au SNDS est attendue en avril 2021, pour officialiser le cadre et notamment la création d'un catalogue. Dans le même temps, se tiendra une réunion du comité stratégique pour fixer les grandes orientations du catalogue. L'étape suivante, en août ou septembre 2021, sera la publication de l'avis de la CNIL sur la liste de bases inscrites au catalogue puis, dans la foulée, la publication de l'arrêté qui officialisera les bases au catalogue.

Comment est structuré Le SNDS ? Initialement, le SNDS est alimenté par trois bases de données dont la plus importante en terme de volumétrie est le SNIIRAM, qui regroupe l'ensemble des informations issues des remboursements de la CNAM, mais aussi le PMSI, sur les séjours hospitaliers, et le CépiDc, la Base des Causes Médicales de Décès (BCMD). « Ces trois bases constituent la base principale du SNDS, puisque la loi Ma Santé 2022 a élargi cette base unique à un système de bases » continue-t-il. Le SNDS « élargi » inclut ainsi toutes les données issues de soins qui sont remboursés partiellement par l'Assurance Maladie. « Nous avons mis en place un catalogue qui réunit une collection de bases pseudonymisées issues du SNDS, et que le Health Data Hub est habilité à mettre à disposition pour la recherche » précise-t-il. Le catalogue rassemblera des copies de bases déjà existantes pour la plupart. Un comité stratégique piloté par l'Etat fixera les grandes orientations de ce catalogue qui se veut évolutif.

Combien de bases sont concernées aujourd'hui ? Le Health Data Hub discute déjà avec un grand nombre de bases pour une mise en catalogue et, pour certaines, un chaînage avec le SNDS. A titre d'exemple, les discussions sont déjà avancées avec la cohorte I-Share (suivi de la santé et des habitudes de consommation de 21 000 étudiants), ESME (patientes atteintes d'un cancer du sein métastatique), la BNDMR (500 000 patients atteints de 4700 maladies rares), Oscour (urgences) ou encore Epithor (interventions en chirurgie thoracique). Dans le cadre du Covid, le périmètre pourra s'élargir à Oscour (résumés des passages individuels aux urgences), SIDEV (résultats des tests virologiques Covid-19 RT-PCR) ou encore SI Vaccin Covid (données de suivi de la vaccination contre la Covid-19). « Il est prévu de chaîner ces bases de données à la base principale du SNDS » confirme Emmanuel Bacry. Nous travaillons également avec un groupe de 6 SAMU, dans un premier temps, sur les données d'arrivée au SAMU et les textes libres associés à chacun des appels ».

Parmi les sujets centraux du Health Data Hub, il y a les règles de valorisation scientifique qui sont étudiées depuis plusieurs mois par un panel de 16 experts, chercheurs dans différents domaines et responsables de données. Et pour les industriels, la question se pose : le responsable des données sera-t-il associé aux articles qui sont publiés ? Le Health Data Hub s'engage à augmenter la visibilité des bases de données grâce à un processus de citation fiable et pérenne. Selon les guidelines, seule une personne ayant participé à une recherche pourra être citée comme co-auteur dans le cas d'une publication, et non la personne morale qui fournit une base de données. Concernant la valorisation financière, le travail est en cours, s'appuyant sur un benchmark international, des cas pratiques et un travail collaboratif mené avec toutes les bases de données. Le CSF agit également dans ce sens. « Ce que l'on peut dire est que tous les académiques auront un accès gratuit, reconnaît Emmanuel Bacry. Et nous aurons deux types de rémunération pour les autres concernant d'une part la collecte de données et la mise en forme de la base et, d'autre part, la mise à disposition de la base. »



En outre, certains responsables de BDD ont demandé à pouvoir exprimer un « droit de véto » sur les projets. Mais cette possibilité ne sera sans doute pas retenue car discrétionnaire.

Et dans le cadre des projets innovants, que peut-on dire à ce jour ? Les données pseudonymisées seront accessibles à tous les porteurs de projets contribuant à l'intérêt général autorisés par la CNIL, après avis d'un comité indépendant. « Nous jouons le rôle de guichet, réalisé auparavant par l'INDS, avec un protocole à remplir qui se fait en collaboration avec les entreprises et start-up, » décrit-il. Etape 1 : les porteurs déposent une demande d'accès auprès du HDH. Etape 2 : la demande est transmise au CESREES pour une validation éthique et scientifique. Celui-ci peut rendre un avis défavorable, réservé ou favorable. Etape 3 : en cas d'avis favorable, la CNIL est saisie pour donner son autorisation au lancement du projet. Selon Emmanuel Bacry, la dynamique est clairement engagée pour le CESREES qui se réunit, tous les mois, et peut alors traiter jusqu'à une cinquantaine de dossiers. En terme de délais, la CNIL doit rendre son avis en l'espace de deux mois, et peut demander deux mois supplémentaires. « Nous pouvons ainsi obtenir un avis en l'espace de quatre mois avec une mise à disposition des données ensuite sur la plateforme, note-t-il. Le principe du catalogue est d'avoir les données déjà prêtes. Cela peut donc aller vite. »

« Nous accompagnons déjà aujourd'hui un peu moins d'une quarantaine de projets pilotes, qui sont à des stades différents, continue-t-il. Dès que le guichet sera ouvert, il pourra rentrer dans son plein régime de croisière ! » Parmi ces projets, citons HYDRO porté par la start-up Implicit, qui est sur le point d'opérer dans le hub. HYDRO récupère les données en temps réel de plus de 10 000 pacemakers. Avec l'objectif d'être chaînées à la base principale du SNDS pour faire de la prédiction de complexités cardiaques. Cette base à terme sera intégrée au catalogue. Un autre projet pilote regroupe 5 CHU partenaires avec une importante base de données d'imageries. Il s'agit de DAICAP qui concerne la détection et caractérisation du cancer de prostate grâce à l'IA. L'intérêt de venir dans le hub est ici d'être multicentrique.

Pour continuer d'avancer, plusieurs groupes de travail (GT) ont été constitués au sein du Health Data Hub, notamment GT Logiciels sur les logiciels, qui est piloté par INRIA et CEA, ainsi que GT CSF sur les « Usages de la génération de données en vie réelle dans la prise de décision ». Il y a également le projet de création d'une bibliothèque d'algorithmes open source sur le SNDS historique. Les initiatives déjà bien engagées visent également à associer et animer l'écosystème scientifique dans les développements. C'est pourquoi le Health Data Hub a mis en place un comité consultatif de 21 experts diversifiés (épidémiologie, imagerie, IA, éthique, sociologie, génétique, juridique, cliniciens, mathématiciens,...), dont fait partie Rémy Choquet, chez Roche, et qui sera prochainement ouvert aux internationaux. Avec un rôle de conseil sur les orientations scientifiques du hub : partenariats académiques, industriels, événements, problématiques éthiques, ...

Le cas de la Banque Nationale de Données des Maladies Rares (BNDMR)

La BNDMR est l'une des premières bases à rejoindre le hub d'ici la fin de l'année. Pour Arnaud Sandrin, son directeur opérationnel, le Health Data Hub représente à la fois des attentes et des opportunités majeures. On compte près de 7000 maladies rares différentes,



rappelle-t-il. En France, environ 3 millions de personnes sont atteintes de maladies rares, aux en majorité des enfants. Dans l'Hexagone, l'offre de soins est organisée au travers de 2230 sites labellisés, rattachés à 110 centres et 23 filières qui concernent toutes les spécialités médicales. « Ce qui a été décidé au niveau national pour l'épidémiologie de ces maladies est qu'un set de données minimum commun soit collecté sur ces 2230 sites, note Arnaud Sandrin. Ce set inclut des informations sur le consentement et l'identification des patients, des informations administratives, le statut vital, les parcours de soins, les activités de soins, le diagnostic (maladie codée avec Orphanet), les traitements, les anténéonataux et la recherche. « Le chainage de la BDDMR avec le SNDS change la donne, pointe-t-il. Les pathologies décrites dans le SNDS sont souvent insuffisantes pour repérer les patients atteints de maladies rares et donc exploiter le reste des données. A l'inverse, la BNDMR permet un repérage précis et gagne à être enrichie des dépenses et remboursements de soins en ville et en établissements spécialisés. » Du côté des industriels, l'intérêt de l'accès aux données de la BNDMR vise principalement à faire de la prévalence et/ou du repérage. Un second axe est de mener des études transversales pour faire de l'épidémiologie sur les maladies rares et, un troisième, des études médico-économiques chainées BNDMR – SNDS. En outre, la BNDMR s'attend à une participation financière conséquente de l'ordre de 150 k€ à 200 k€ par étude, sinon elle ne voit pas l'intérêt de participer au HDH. Arnaud Sandrin insiste également sur le fait d'avoir des contrats simples et cadrés pour ne pas perdre de temps et d'argent dans les démarches préalables à l'étude.

La dynamique du Health Data Hub est ainsi bien en marche ! Elle demande maintenant à être poursuivie et accélérée dans le temps. « Sur le plan industriel, nous avons hâte de pouvoir travailler avec les premières bases et le catalogue, » note Jean-Vannak Chauny. Avec plusieurs points néanmoins à éclaircir, au delà des problématiques économiques et de valorisation scientifique, concernant les contrats multipartites que l'on pourra développer entre les producteurs de données et le SNDS, mais aussi les Méthodologies de Référence (MR) et les procédures de dossier type simplifié d'accès, ou encore un système « fast track » pour l'avenir. Ce hub ouvre-t-il vraiment de nouvelles opportunités pour faire avancer la recherche, en s'appuyant notamment sur le machine learning et l'IA ? « Autant de thématiques qu'il nous faut continuer de développer en partenariat et en transparence, » conclut-il. Cette conférence en a été un bel exemple !