

ENTREPÔT PARTAGÉ « eHOP » ET HEALTH DATA HUB

LE BIG DATA EN SANTÉ, L'ANALYSE DE MARC CUGGIA

Co-pilote, avec Dominique Polton, de la mission de préfiguration du Health Data Hub, et Gilles Wainrib de la start-up Owkin, le Professeur Marc Cuggia est un spécialiste reconnu de l'informatique médicale. Praticien hospitalier au CHU de Rennes, il est chargé de l'équipe-projet « Données massives en santé » (Université de Rennes 1), et dirige le projet eHOP, l'un des tout premiers entrepôts de données hospitalières en France. Rencontre. Par Joëlle Hayek



Le Pr Marc Cuggia, Praticien hospitalier au CHU de Rennes

DANS QUEL CONTEXTE L'ENTREPÔT DE DONNÉES BIOMÉDICAL eHOP A-T-IL VU LE JOUR ?

PR MARC CUGGIA : Nos travaux ont commencé il y a une décennie, alors que le CHU de Rennes prenait le virage de la dématérialisation et informatisait progressivement sa production de soins. Cette mutation nous a poussé à nous interroger : la numérisation des données des patients ne pouvait-elle pas générer des possibilités d'utilisation secondaires ? Nous pensions plus particulièrement ici aux études de faisabilité et au repérage des patients dans le cadre des essais cliniques, ce qui nous a d'ailleurs poussé à orienter la base de données eHOP sur la recherche. Mais cinq autres catégories d'usages potentiellement facilités par le Big Data ont, depuis, été documentées par la littérature : la médecine personnalisée, les réseaux de vigilance, le pilotage du système de soins, l'épidémiologie et la formation.

COMMENT AVEZ-VOUS PLUS PARTICULIÈREMENT PROCÉDÉ ?

Il nous fallait mettre en place une infrastructure technique qui permettrait de tirer profit du gisement de données dormant au sein du système d'information hospitalier (SIH). C'est ainsi qu'a émergé la notion d'entrepôt de données, pour laquelle deux établissements se sont positionnés comme précurseurs en France : le CHU de Rennes, et l'Hôpital Européen Georges Pompidou (AP-HP), à Paris. Rattaché à l'unité mixte de recherche INSERM LTSI (Laboratoire Traitement du Signal et de l'Image), notre centre de données cliniques (CDC) a alors développé sa propre technologie, pour « fouiller » efficacement et de manière exhaustive dans le SIH tout en assurant la sécurité des données patients.

COMMENT S'ORGANISE AUJOURD'HUI L'ENTREPÔT DE DONNÉES DU CHU DE RENNES ?

Il s'inscrit au cœur d'une organisation centralisée. Seul le CDC peut y accéder et représente donc le portail d'entrée unique pour toutes les demandes liées à la recherche : il effectue la consultation méthodologique, réalise l'étude de faisabilité, puis soumet le tout pour validation auprès du comité de régulation des données, qui donne un avis scientifique, éthique et déontologique. ●●●

“NOTRE CENTRE DE DONNÉES CLINIQUES A DÉVELOPPÉ SA PROPRE TECHNOLOGIE, POUR « FOUILLER » EFFICACEMENT ET DE MANIÈRE EXHAUSTIVE DANS LE SIH”

“CETTE ORGANISATION EN RÉSEAU AUTOUR DU HEALTH DATA HUB PERMET DE FAIRE ÉMERGER DES USAGES QUI ALIMENTENT LES PROCESSUS MÉTIERS”



●●● Nous extrayons ensuite de l'entrepôt de données les informations pertinentes selon le périmètre de l'étude, et mettons à dispositions des utilisateurs les outils et les procédures qui faciliteront leur exploitation.

CE MODÈLE A PROGRESSIVEMENT ESSAIMÉ DANS LE GRAND OUEST.

Notre entrepôt de données ne contient en effet qu'une fraction des données d'un patient, qui correspondent à son passage au sein de notre établissement. Pour exploiter le Big Data à hauteur de son potentiel, il faut nous inscrire dans une approche populationnelle qui impose d'agir à plus large échelle. En 2015, le Groupement interrégional de recherche clinique et d'innovation (GIRCI) Grand Ouest a créé un réseau inter-régional des CDC constitué des CHU d'Angers, Brest, Nantes, Poitiers, Rennes et Tours, ainsi que de l'Institut de Cancérologie de l'Ouest. Tous bénéficient aujourd'hui de la technologie eHOP. Chaque CDC est cependant indépendant et exploite ses données en toute autonomie, même si la coordination de l'ensemble est assurée par le CDC de Rennes.

C'EST CETTE COORDINATION QUI PERMETTRA DE DONNER CORPS À L'ÉTAPE SUIVANTE : LA CRÉATION D'UN ENTREPÔT PARTAGÉ. POUVEZ-VOUS NOUS EN PARLER ?

Avec plus de 5 millions de patients aux données informatisées, le groupement

HUGO (Hôpitaux Universitaires du Grand Ouest) constitue le plus grand réseau de données cliniques d'Europe. La création de l'entrepôt partagé matérialise son ambition de rester le leader européen dans le domaine des Big Data hospitaliers. Cette structure devrait pouvoir être lancée avant la fin de l'année. Il a d'abord fallu construire des processus industrialisés pour faire converger les flux de données et justement permettre leur exploitation multicentrique – un chantier mené en partenariat avec un éditeur spécialiste de l'interopérabilité technique. En parallèle, un appel à projets a été lancé par le GIRCI Grand Ouest, afin de susciter de nouveaux usages cliniques en lien avec cet entrepôt partagé. Une quinzaine de projets a d'ores-et-déjà été soumise, tous mettant à contribution au moins 3 CDC – c'était d'ailleurs l'un des prérequis. Pour avoir une masse critique suffisante, nous en sélectionnerons quatre d'ici la fin de l'année.

IL FAUT, POUR CELA, POUVOIR EXPLOITER LES DONNÉES DE MANIÈRE SYNERGIQUE. QUELLE APPROCHE AVEZ-VOUS RETENU ?

L'interopérabilité technique n'est en effet pas suffisante ; il faut également une interopérabilité sémantique. Certaines données bénéficient déjà d'une structuration normalisée, comme le PMSI, le circuit du médicament avec le standard PN13 ou la codification UCD/ATC, et la biologie qui s'appuie sur la norme de

transmission Hprim et la nomenclature standardisée LOINC. Leur alignement sera donc relativement aisé. Il sera toutefois difficile d'avoir une approche sémantiquement homogène pour l'ensemble des données, du moins jusqu'à ce que le système international de terminologie clinique SNOMED-CT soit traduit en français et utilisé par les éditeurs de logiciels. Mais la technologie eHOP propose d'autres méthodes pour pouvoir toutes les exploiter, y compris en ce qui concerne les données textuelles ou non structurées.

COMMENT CET ENTREPÔT PARTAGÉ ENTRE-T-IL EN RÉSONNANCE AVEC LE HEALTH DATA HUB ?

Il préfigure ce que pourrait être un hub local : nous nous porterons d'ailleurs candidats pour être l'un de ces nœuds constitutifs du Health Data Hub. Pour schématiser il faut prendre l'exemple d'un réseau hiérarchique. Les CDC sont l'unité organisationnelle de base, présente dans les établissements. Leur rôle est de collecter les données et d'apporter les services et l'expertise en data science auprès des producteurs et des utilisateurs de ces données. Puis se positionne l'entrepôt partagé, une plateforme interrégionale qui oriente la convergence technique, juridique et organisationnelle permettant le partage et l'exploitation des données à l'échelle de plusieurs établissements. Au niveau national, le ●●●

●●● Health Data Hub, fédèrera l'ensemble des données produites dans le cadre des programmes de financement nationaux. Celles-ci sont regroupées au sein du Système National des Données de Santé (SNDS), qui dispose par exemple de la typologie des actes de biologie réalisés, mais pas des résultats d'examens – lesquels sont pour leur part présents dans les SI hospitaliers ou ambulatoires. La constitution du réseau de centres de données permettra donc d'enrichir le SNDS avec des données cliniques particulièrement importantes.

D'OÙ LA NÉCESSITÉ DE CHAÎNER LE TOUT.

Le Health Data Hub disposera en effet d'un nœud central, qui sera en charge de gérer les collections de données nationales comme les grandes cohortes ou l'actuel SNIIRAM. Le HDH mettra donc en place les chaînages nécessaires au croisement des données au sein d'une architecture sécurisée, en lien avec des outils de traitement de l'information. Cette infrastructure technique et les jeux de données du SNDS pourront alors être exploités au sein des hub locaux. Nous devons d'ailleurs prochainement tester une première application d'enrichissement mutuel des données avec le projet HUGO SHARE : il s'agira, plus concrètement, de chaîner nos données biologiques

données et, plus globalement, par cette organisation en réseau autour du Health Data Hub, permettent en effet de faire émerger des usages qui dépasseront le seul cadre de la recherche pour alimenter les processus métiers. C'est là un champ sur lequel nous travaillons déjà : grâce à des liens bidirectionnels, notre entrepôt alimente aujourd'hui le système de production du PMSI pour optimiser le codage. eHOP est également utilisé à des fins de pharmacovigilance ou d'aide au diagnostic génétique. D'autres usages nouveaux concerneront directement la prise en charge des patients, ainsi que le montre le projet HUGO SHARE. Le CDC de Rennes projette en outre de développer un algorithme qui, à partir des données d'urgence et des données d'aval, permettra de prévoir le degré de gravité et la durée du séjour d'un patient accueilli par le service d'urgence.

COMMENT MIEUX ANTICIPER CETTE RÉVOLUTION CULTURELLE ?

Il faudra d'abord acculturer et former les cliniciens aux concepts de la *data science*, pour qu'ils puissent s'approprier les méthodes de *data mining* et d'intelligence artificielle, mais aussi et surtout participer à la validation de nouveaux outils d'aide à la décision – y compris ceux conçus par l'industrie. Il faut également réfléchir aux

qui pourra les recruter en-dehors des grilles tarifaires classiques et les mettre à disposition des hub locaux.

QUID, POUR FINIR, DU PATIENT ?

Conserver les relations de confiance auprès des patients et du corps médical représente un enjeu majeur pour pouvoir poursuivre et élargir ces travaux. Un effort pédagogique sera nécessaire pour expliquer les finalités des nouveaux usages, et clarifier les droits que peuvent exercer les usagers par rapport à ces sujets. Le curseur est toutefois à placer avec attention. Nous avons, pour notre part, choisi un système opt-out : le patient peut s'opposer à la réutilisation de ses données, même anonymisées, par un tiers – hors certains usages dérogatoires comme par exemple la pharmacovigilance. Mais il ne peut y avoir de doxa unique ; peut-être faudra-t-il construire un modèle où le consentement sera donné par scénario et catégorie d'usage. Nous pouvons nous inspirer ici du modèle exemplaire de nos collègues québécois, d'autant que le RGPD impose désormais la mise en place d'un système de consentement dynamique.

CE SERA DONC LÀ UN CHANTIER DE TAILLE.

En attendant, nous pourrions renforcer la régulation existante, peut-être en faisant participer les associations des usagers aux Conseils scientifiques et éthiques. En tout état de cause, l'ensemble se construira progressivement, en capitalisant sur les expériences des uns et des autres. N'oublions pas qu'il faudra, à terme, intégrer les données produites par les patients eux-mêmes pour continuer d'enrichir nos jeux de données. Le futur espace numérique de santé pourrait d'ailleurs représenter une piste pour justement recueillir le consentement des usagers quant à l'utilisation des données qu'ils mettront à disposition des établissements de santé. Nous continuons, en ce qui nous concerne, à partager la dynamique initiée dans le Grand Ouest. Le Big Data et son corolaire, l'intelligence artificielle et les technologies algorithmiques, représentent en effet un réel pari sur l'avenir. Mais tout le monde devra être au rendez-vous : les établissements de santé, les patients, et bien sûr l'État, qui devra financer et consolider ces organisations nouvelles. ●

“UN EFFORT PÉDAGOGIQUE SERA NÉCESSAIRE POUR EXPLIQUER LES FINALITÉS DES NOUVEAUX USAGES, ET CLARIFIER LES DROITS QUE PEUVENT EXERCER LES USAGERS PAR RAPPORT À CES SUJETS”

et médicamenteuses issues des hôpitaux du Grand Ouest avec celles du SNDS relatives à la consommation médicamenteuse en ville, pour caractériser les événements indésirables liés à la poly-médication et aux ruptures thérapeutiques ville-hôpital. Ce qui nous permettra de développer, à terme, un algorithme qui permettra de prédire la trajectoire des patients poly-médiqués.

C'EST LÀ UN CHANGEMENT DE CAP PAR RAPPORT À L'OBJECTIF PREMIER, QUI ÉTAIT DE CRÉER UN SYSTÈME ORIENTÉ SUR LA RECHERCHE.

Les usages favorisés par les entrepôts de

modalités de partage de la valeur entre l'industrie et les établissements de santé qui auront mobilisé leurs experts cliniques en ce sens. Un tel schéma, qui devra être identique pour l'ensemble des CHU, permettra d'ailleurs de pérenniser les CDC. Il faudra, enfin, rassembler des compétences pluridisciplinaires au sein des CDC, en particulier des *data-scientists*. C'est d'ailleurs l'une des limites du système à l'heure actuelle, puisque les grilles tarifaires de la fonction publique hospitalière ne font pas le poids par rapport au secteur privé. Peut-être faudrait-il financer ces postes à l'échelle de la région, voire à travers le Health Data Hub