

A S S E M B L É E N A T I O N A L E

X V ^e L É G I S L A T U R E

Compte rendu

Mission d'information de la Conférence des Présidents « Bâtir et promouvoir une souveraineté numérique nationale et européenne »

- Audition, ouverte à la presse, de Mme Laurence Jay-Passot, déléguée générale du groupement de coopération sanitaire des hôpitaux universitaires Grand Ouest (HUGO), et du professeur Marc Cuggia, professeur des universités-praticien hospitalier au centre hospitalier universitaire (CHU) de Rennes, sur la plateforme de données hospitalières Ouest Data Hub 2

Jeudi

18 février 2021

Séance de 12 heures

Compte rendu n° 29

SESSION ORDINAIRE DE 2020-2021

**Présidence de
M. Jean-Luc
Warsmann,
*président***



Audition, ouverte à la presse, de Mme Laurence Jay-Passot, déléguée générale du groupement de coopération sanitaire des hôpitaux universitaires Grand Ouest (HUGO), et du professeur Marc Cuggia, professeur des universités-praticien hospitalier au centre hospitalier universitaire (CHU) de Rennes, sur la plateforme de données hospitalières Ouest Data Hub

La séance est ouverte à 12 heures.

Présidence de M. Jean-Luc Warsmann, président.

M. le président Jean-Luc Warsmann. Nous poursuivons nos auditions consacrées à la thématique des données de santé et de la souveraineté numérique. Le Ouest Data Hub est une plateforme de données hospitalières permettant de regrouper de façon anonymisée les données de six établissements membres du groupement, à savoir les centres hospitaliers universitaires (CHU) d'Angers, de Brest, de Nantes, de Rennes, de Tours ainsi que l'institut de cancérologie du Grand Ouest.

Ce hub est un nouveau concentré des problématiques qui intéressent notre mission d'information. Il interroge notre capacité à procéder à des choix opérationnels qui garantissent à la fois la protection des données de nos concitoyens et un niveau le plus élevé possible de performance. Il s'agit de numériser rapidement notre système de santé pour gagner en maturité sur certaines technologies clés pour l'avenir et de rendre un service d'une qualité toujours croissante à nos concitoyens. Nous avons également – et l'actualité de ces derniers jours nous le rappelle à nouveau – le devoir d'assurer la plus haute sécurité possible de nos systèmes d'information contre les cyberattaques. Nous nous réjouissons donc d'échanger avec vous sur ces différents sujets.

M. Philippe Latombe, rapporteur. Je souhaite vous interroger sur trois points en particulier.

J'aimerais d'abord que vous nous présentiez en détail le Ouest Data Hub : son organisation, son fonctionnement et son actualité pour l'année 2021. Il me semble important de comprendre comment fonctionne cette plateforme de données de santé, dont l'objectif est de soutenir la recherche et l'innovation en santé.

Je souhaiterais également savoir comment l'action du Ouest Data Hub s'articule avec les autres hubs de données de santé, et en particulier au niveau national avec le Health Data Hub, dont nous avons auditionné la directrice plus tôt ce matin. Il s'agit pour nous de comprendre comment le Ouest Data Hub s'intègre à l'écosystème du numérique en santé.

Le second sujet que je souhaite aborder avec vous a trait au cœur des travaux de notre mission d'information : la souveraineté numérique dans le domaine de la santé. Comment percevez-vous cet enjeu et comment l'avez-vous intégré à vos choix opérationnels au sein du Ouest Data Hub ? Quels ont été les choix techniques réalisés pour l'hébergement des données de santé et leurs motivations ? Cela nous permettra également d'échanger sur l'existence, ou non, d'arbitrages entre performance et sécurité à court et à moyen termes.

Enfin, et M. le président l'a souligné, l'actualité récente est marquée par des cyberattaques contre les systèmes d'information des établissements de santé. Face à la sophistication de la menace cyber, comment est-il possible, selon vous, de garantir un niveau

de protection maximale à nos infrastructures numériques, en particulier dans le domaine de la santé ?

Mme Laurence Jay-Passot, déléguée générale du groupement de coopération sanitaire des hôpitaux universitaires Grand Ouest (HUGO). Le Ouest Data Hub est un hub interrégional de données de santé qui permet de mener des études nécessitant d'agréger des données issues des entrepôts de données de santé des six établissements que vous avez mentionnés. Sa création s'inscrit dans une stratégie volontariste sur les données massives en santé qui a été construite, depuis plusieurs années, par notre réseau de CHU. Cette stratégie poursuit trois objectifs principaux. Le premier est de potentialiser l'expertise de nos CHU en matière de données, en s'appuyant à la fois sur le réseau des experts en données et sur les réseaux d'experts cliniciens du Grand Ouest. Le deuxième objectif est de pouvoir atteindre, par un travail collectif, une masse critique nécessaire pour réaliser des projets d'envergure. Enfin, le troisième objectif est de nous interfacer avec l'écosystème en santé et numérique sur notre territoire afin de favoriser et de servir l'innovation.

La caractéristique principale de ce hub interrégional est qu'il s'inscrit dans une logique de maîtrise de toute la chaîne de la donnée – depuis la production de la donnée, la structuration de sa collecte puis sa mise en qualité et jusqu'à son usage. Ainsi nous avons toujours le souci de rester en lien avec les professionnels de terrain et les cliniciens.

Il est tout d'abord important d'expliquer pourquoi notre CHU a fait le choix d'exploiter de façon prioritaire les données massives en santé.

Pr Marc Cuggia, professeur des universités, praticien hospitalier au centre hospitalier universitaire (CHU) de Rennes. L'informatisation des données de santé crée un potentiel de transformation et d'innovation : nous pouvons utiliser ces données de santé pour engendrer de l'information et de nouvelles connaissances grâce à des méthodes en plein développement, comme les fouilles de données et l'intelligence artificielle.

Les cliniciens et les acteurs de l'écosystème identifient plusieurs champs d'usages de ces données. Tout d'abord, ces données sont utiles pour soutenir la recherche biomédicale. Elles peuvent servir, par exemple, à identifier de nouveaux traitements ou de nouveaux biomarqueurs, à soutenir la recherche clinique par la réalisation d'études de faisabilité ou encore d'études populationnelles dans le champ de l'épidémiologie.

Ces données servent également la prise de décision, puisqu'elles fournissent des outils d'aides à la prescription ou à l'interprétation des examens biologiques. Elles permettent ainsi de mettre au point des outils qui vont aider le clinicien dans la prise en charge des patients.

Ces données alimentent également le champ de la vigilance et de la veille sanitaire. Au-delà de la veille sanitaire lors des épisodes épidémiques, les données permettent d'évaluer les traitements en vie réelle et de documenter leurs effets indésirables. Elles contribuent également à la matériovigilance, c'est-à-dire à la surveillance des dispositifs médicaux en vie réelle. L'usage de données massives en vie réelle est une source d'information extrêmement riche.

Enfin, les données contribuent au pilotage du système de santé. L'accès à des données transversales permet de mieux appréhender les parcours de santé de nos concitoyens

et d'adapter le système en fonction des contraintes. Tous ces domaines sont extrêmement importants et peuvent bénéficier des données massives.

Mme Laurence Jay-Passot. Partant de ce constat partagé, nous avons identifié un enjeu stratégique à exploiter ces gisements de données. Nous nous sommes ainsi engagés dans une démarche en plusieurs étapes – toutes sont fondatrices du hub interrégional aujourd'hui en place.

La première étape fondamentale a consisté à structurer les gisements de données dans chaque établissement. Nous avons fait en sorte que les CHU puissent chacun développer un entrepôt de données de santé de la manière la plus convergente et homogène possible. Cela nous permet aujourd'hui de disposer de six entrepôts de données qui fonctionnent en utilisant des technologies similaires.

La seconde étape, débutée il y a plusieurs années, a consisté à mutualiser dès l'origine les expertises sur les données massives en santé et à mettre en réseau ces entrepôts de données de santé. Grâce à cette base, nous avons créé, il y a deux ans, un data hub partagé et sécurisé, le Ouest Data Hub, en même temps que nous travaillions à la stimulation des usages et à l'exploitation multicentrique des données, pour pouvoir d'emblée tester la plateforme interrégionale sur de vrais projets.

Ce hub interrégional n'a de sens que s'il s'appuie sur des centres de données cliniques solides : les entrepôts de données des établissements qui composent notre réseau constituent le socle fondamental du hub. Ces centres de données ont été construits en plusieurs années dans notre interrégion. Chaque CHU du réseau propose ainsi, dans son centre de données clinique, une expertise pluridisciplinaire, un accompagnement auprès des professionnels de santé et des éléments d'infrastructure normalisés qui garantissent la sécurité des données. Ces centres de données cliniques sont le fondement de notre Ouest Data Hub.

Ces centres de données cliniques prennent appui sur un réseau d'expertises structuré. Ceci permet de disposer d'une vraie feuille de route technique et opérationnelle pour travailler ensemble et d'évoluer vers une exploitation multicentrique des données.

La plateforme interrégionale de données de santé Ouest Data Hub rassemble six hôpitaux et est permise par le déploiement d'une technologie commune dans les différents centres de données cliniques des établissements. La plateforme donne ainsi accès à un volume très important de données issues des six établissements. Le catalogue est nourri par les informations produites par les patients et continue à s'enrichir régulièrement. Ce potentiel considérable de données a du sens car ces données sont homogènes et de qualité : la possibilité d'exploiter les données de manière intéressante est un enjeu capital.

La plateforme interrégionale est composée de plusieurs éléments. La dimension infrastructure est essentielle pour connecter et faire communiquer entre eux les différents centres de données cliniques. L'infrastructure est hébergée au sein du CHU de Nantes. Le Pr Marc Cuggia détaillera plus tard ses éléments de logiciels très spécifiques.

Nous avons ensuite mis en place une gouvernance afin de définir les règles de partage et d'accès aux données et de rappeler le cadre éthique, déontologique et juridique qui doit être appliqué. La particularité de notre Ouest Data Hub est de mobiliser des structures juridiques distinctes, car chaque CHU constitue une personne morale propre.

Enfin, une politique scientifique qui s'articule autour des usages possibles des bases de données est animée grâce au réseau d'experts et à une politique d'appels à projets.

La gouvernance du Ouest Data Hub est structurée à plusieurs niveaux. Le pilotage stratégique est organisé par un comité stratégique impliquant les gouvernances des six établissements. Le pilotage scientifique, incarnée par une direction scientifique tripartite, permet d'orienter les choix que nous faisons. Nous avons également organisé un pilotage opérationnel sous la forme d'un guichet unique, qui permet de répondre aux sollicitations et de faire vivre cette plateforme. Nous avons enfin été attentifs à la dimension éthique et juridique de notre dispositif, c'est pourquoi nous avons mis en place un comité scientifique et éthique dédié. Un délégué à la protection des données (DPO) a également été choisi.

L'année 2019 et le premier semestre de l'année 2020 ont été consacrés à la conception et à la structuration de cette plateforme interrégionale ainsi qu'à la constitution du socle de données commun qui permet de la nourrir. Depuis la mise à disposition technique de notre plateforme à l'été 2020, nous sommes passés dans une phase de projet. Il s'agit maintenant de mener à bien les premiers projets interrégionaux utilisant cet outil, de poursuivre l'enrichissement des catalogues de données et de réfléchir au modèle économique de cette plateforme, qui est également un sujet important.

M. Philippe Latombe, rapporteur. Comment avez-vous construit l'architecture des différents entrepôts de données ? Quelles technologies utilisez-vous ?

Pr Marc Cuggia. Nous utilisons une technologie développée au sein du laboratoire de l'Institut national de la santé et de la recherche médicale (Inserm) dans lequel je travaille, en lien avec le CHU de Rennes. Il s'agit de technologies assez standard dans le domaine des données massives et du développement logiciel. L'objectif est d'intégrer les données du système d'information hospitalier dans les entrepôts locaux. Cette technologie est le fruit d'une activité de recherche et développement menée depuis plusieurs années. Un élément clé du déploiement était de s'adosser à une logique industrielle pour pouvoir déployer, alimenter et construire ces entrepôts. Pour cela, nous avons noué un partenariat avec la société Enovacom, aujourd'hui devenue une filiale d'Orange. Nous avons couplé cette technologie avec le savoir-faire de cette société pour nous aider à construire ces flux de données. Cela nous a permis à nous, laboratoire de recherche et CHU de Rennes, de nous focaliser sur l'innovation sur ces sujets et d'assurer un déploiement et une maintenance industriels de ces socles de données au niveau de chaque établissement.

Le développement de la plateforme Ouest Data Hub, c'est-à-dire de la plateforme collectant les données, a également été réalisé par notre équipe à Rennes. Nous avons développé l'ensemble des éléments logiciels et de mise en œuvre de cette plateforme. La partie technologique ayant trait à la sécurité et aux serveurs est assurée par le CHU de Nantes, qui met à disposition une infrastructure de stockage et de calcul. Nous nous appuyons pour cela sur son savoir-faire en matière d'hébergement de données de santé. Le CHU de Nantes est en effet l'un des établissements hébergeurs de données de santé disposant de la plus grande expérience dans ce domaine. Ce dispositif a donc été élaboré dans une logique de coconstruction grâce à des dynamiques de recherche et développement à la fois académiques et industrielles.

Nous utilisons des technologies à l'état de l'art. Pour réaliser des traitements sur des données complexes, il faut bénéficier de capacités de calcul mais aussi de logiques industrielles pour que ces outils puissent être déployés et maintenus par les directions des

systèmes d'information (DSI). Les entrepôts sont hébergés directement par les DSI de chaque établissement. Le Ouest Data Hub, quant à lui, est hébergé par le CHU de Nantes car cet établissement est hébergeur de données de santé. L'écosystème s'est construit de cette manière.

M. Philippe Latombe, rapporteur. Les données de santé sont-elles pseudonymisées et sécurisées ?

Pr Marc Cuggia. Nous avons deux niveaux d'agrégation. Le premier niveau est celui de l'établissement : un entrepôt de données est géré par le centre de données cliniques de chaque établissement. Nous récupérons les données produites par les différents logiciels métiers de l'établissement et nous alimentons ainsi l'entrepôt de données avec des données dé-identifiées afin qu'elles puissent être utilisées à des fins d'innovation et de recherche. Chaque établissement reste souverain dans l'usage de ses données. Les établissements traitent donc en premier lieu les données et peuvent les mettre à disposition de leurs collègues cliniciens pour des projets monocentriques, par exemple.

Dès lors que nous devons partager des données entre plusieurs établissements, nous nous appuyons sur la plateforme Ouest Data Hub qui permet de collecter les données sur projet. Ces données, collectées à partir des différents entrepôts de données, sont dé-identifiées puis collectées et agrégées sur projet. Nous ne déversons pas l'ensemble des données sur le Ouest Data Hub : nous mettons au catalogue ces données, puis nous ne mettrons à disposition que les données nécessaires à la réalisation d'une étude. Ces données sont agrégées dans des espaces projets dédiés à chaque projet, sécurisés et cloisonnés. En quelque sorte, les données quittent la bulle sécurisée de l'entrepôt de données de chaque établissement pour rejoindre une autre bulle sécurisée dans laquelle se réalisent les traitements.

M. Philippe Latombe, rapporteur. Sous quelle forme se présentent ces entrepôts de données dans chaque établissement ? La bulle de l'espace projet est-elle hébergée en *cloud* ou bien dans le serveur physique de l'endroit où se déroule la recherche ?

Pr Marc Cuggia. Les données sont hébergées sur des serveurs au sein de chaque établissement. Les données sont donc physiquement présentes dans des serveurs hébergés dans chaque CHU, sous la responsabilité des DSI. Les centres de données cliniques sont là pour exploiter les données pour l'établissement. La plateforme Ouest Data Hub est également hébergée au sein du CHU de Nantes, sous la direction de la DSI qui met à disposition les serveurs sécurisés et les ressources de stockage et de calcul pour réaliser les travaux. L'accès aux données présentes sur le Ouest Data Hub est strictement réservé aux équipes projets qui vont exploiter les données mises à disposition dans leur espace projet. Cet écosystème est complètement intégré. La plateforme Ouest Data Hub est en capacité de collecter les données issues de chaque établissement en fonction des besoins de chaque projet.

M. Philippe Latombe, rapporteur. Vous avez donc opté pour un hébergement physique. Vous n'utilisez pas de *cloud*.

Pr Marc Cuggia. Nous n'utilisons pas du tout de *cloud*.

M. Philippe Latombe, rapporteur. Comment allez-vous procéder avec le Health Data Hub, qui, lui, utilise des solutions de *cloud* ?

Pr Marc Cuggia. Nous conduisons actuellement un projet avec le Health Data Hub. Nous agrégeons un certain nombre de données collectées à partir des centres de données cliniques. Ces données sont dé-identifiées et transmises dans l'espace projet mis à disposition par le Health Data Hub. Les données collectées dans le cadre du projet seront couplées avec les données de l'Assurance maladie. Ainsi, des jeux de données seront extraits de chaque entrepôt de données correspondant aux besoins de l'étude, puis déposés sur la plateforme du Health Data Hub. L'Assurance maladie extraira elle aussi un jeu de données, qu'elle déposera sur la plateforme. Nous réaliserons alors l'appariement de ces données pour les besoins de l'étude. La logique est la même : les données quittent une bulle sécurisée pour rejoindre une autre bulle sécurisée. Les données collectées ne sont pas anonymisées mais dé-identifiées : l'on applique des algorithmes qui vont supprimer les éléments de ré-identification potentielle des données.

M. Philippe Latombe, rapporteur. Nous venons de conduire une audition avec le Health Data Hub. L'utilisation du *cloud* et le recours à des solutions américaines, notamment Azure de Microsoft, ont donné lieu à beaucoup de discussions. Comprenez-vous le choix fait par le Health Data Hub ? Le Health Data Hub aurait-il pu opter pour la même solution physique que vous et la développer à une échelle beaucoup plus importante ?

Pr Marc Cuggia. J'ai été l'un des trois copilotes de la mission de préfiguration du Health Data Hub. La mission de préfiguration visait à mettre à disposition un ensemble de solutions techniques pour permettre le traitement des données de façon sécurisée et pour garantir une souveraineté nationale ou européenne sur ces sujets. La lettre de mission confiée par Mme la ministre mentionnait clairement ces éléments.

Les solutions, qu'elles soient complètement *indoor* ou *cloud*, peuvent être utilisées. Je n'ai pas réellement compris pourquoi l'on a fait le choix d'utiliser des technologies qui ne sont pas portées par des acteurs français ou européens, si l'on souhaitait mettre en place une solution souveraine. Cela est paradoxal.

Nous avons exploré un certain nombre de pistes lors de la mission de préfiguration. La structure TeraLab est une initiative très intéressante, qui a été développée pour mettre en place des traitements *big data* pour accompagner différents projets. Elle aurait pu constituer une bonne solution, au moins transitoire, pour tester et organiser le déploiement et l'usage du Health Data Hub.

Ceci étant dit, les choix opérés étaient sans doute justifiés d'un point de vue technologique. L'architecture Azure est très performante sur le plan du traitement des données. Mais il est vrai que je n'ai pas très bien compris pourquoi nous n'avons pas pu examiner d'autres solutions.

Un certain nombre d'éléments de discussion ont été apportés à la suite du *Cloud Act*. Les acteurs américains hébergent aussi des données de santé : Microsoft est homologué hébergeur de données de santé. Dans ce cas, ces données sont nominatives et orientées sur le soin.

Des solutions viables, *indoor* ou sous *cloud*, existaient et elles auraient pu être explorées, au moins le temps d'une phase projet et de montée en charge. Il y a également une volonté de réversibilité de ces choix, c'est-à-dire une volonté d'utiliser un environnement plus souverain. À partir du moment où les données sont sécurisées, nous n'avons pas de raison de

ne pas contribuer à cet effort national par le projet HUGO-Share, car le Health Data Hub est un formidable projet.

Les auditions conduites dans le cadre de la mission de préfiguration nous ont montré que l'écosystème était très enthousiaste à l'idée de mettre en place un projet qui l'aide à développer des innovations. Il faut désormais probablement réfléchir à la façon d'y arriver, repenser ces choix et coconstruire. La mission de préfiguration avait insisté sur le fait que l'ensemble des acteurs devait être impliqués dans la coconstruction de ce hub national. Le Ouest Data Hub a appliqué cette stratégie pour se structurer et s'organiser : le hub doit être au service de l'écosystème. Il convient donc d'orienter la feuille de route du Health Data Hub pour nous aider à développer ces logiques territoriales.

M. Philippe Latombe, rapporteur. Votre architecture est-elle justifiée par votre organisation en réseau de CHU ? Le poids de l'histoire a-t-il joué dans cette architecture ? Le Health Data Hub, au contraire, est une création ex-nihilo – cela pourrait expliquer le fait que d'autres choix technologiques ont été opérés.

Pr Marc Cuggia. Si l'histoire de l'informatisation de l'hôpital est ancienne, l'informatisation du dossier médical est, en revanche, relativement récente. Dès lors que nous numérisons les données du dossier médical, nous avons souhaité les utiliser à des fins de recherche et d'innovation.

Au départ, un prototype qui nous permettait de croiser ces données en format papier avait été développé. La vague de la santé numérique a complètement transformé le secteur : nous avons désormais à notre disposition des capacités de stockage et des méthodes qui nous permettent de répondre à des questions médicales.

Un prototype avait donc été développé par mon équipe à Rennes. À Brest, une équipe avait, elle, déjà pensé la création d'un centre de données cliniques. Nous avons développé un projet commun et nous avons déployé ces centres de données cliniques. Nous étions, dès le départ, convaincus que nous devions le faire de manière groupée. Nous ne pouvions pas dégager d'axes d'innovation isolément. La création des centres de données cliniques nous a donc amenés à nous interroger sur des projets multicentriques et des moyens communs. La plateforme actuelle est ainsi le fruit d'un travail de coopération existant depuis très longtemps.

Nous sommes confrontés à des problématiques très complexes de qualité de données. En France, notre système de données de santé est extrêmement fractionné ; aucun système d'information ne ressemble à un autre. L'enjeu de collecter les données et de les harmoniser est essentiel, et implique tout un panel d'acteurs. Cela implique des actions tout au long de la chaîne de production de données et d'expertise. C'est tout l'enjeu des centres de données cliniques.

Mme Laurence Jay-Passot. Il existe une dimension historique aux choix opérés. Mais au-delà de cela, nous avons conduit des choix stratégiques parfaitement assumés. En premier lieu, nous avons fait le choix d'un dispositif qui nous permet de maîtriser la donnée de bout en bout, pour des raisons d'éthique et d'acceptation par les patients et les communautés médicales. Nous avons également fait le choix de garder un lien très fort avec le terrain, les investigateurs et les cliniciens. Ce lien garantit que les données seront de qualité, aussi bien dans leur collecte que dans leur traitement.

Les choix techniques et d'organisation sont donc assis sur ces convictions. Nous appliquons ainsi un principe de subsidiarité : les centres font localement tout ce qui leur est possible de faire. Tous les projets n'ont pas vocation à être portés à l'échelon interrégional. Mais dès lors que nous identifions un intérêt à massifier, nous utilisons l'infrastructure commune.

Nous avons discuté de la création d'un entrepôt mutualisé. Nous avons jugé qu'un entrepôt mutualisé n'apporterait pas de plus-value particulière pour le moment et dans le contexte actuel. Ce dispositif est en effet extrêmement complexe et pourrait poser des questions de sécurisation. Nous avons conclu que nous pouvions travailler très efficacement et répondre à nos objectifs avec le modèle des plateformes locales.

M. Philippe Latombe, rapporteur. Cela a-t-il nécessité de mener un important travail de pédagogie en direction des équipes ? Avez-vous eu besoin d'acculturer les équipes à la nécessité de partager les données ? Votre gouvernance, avec un comité d'éthique et un DPO spécialisé, a-t-elle permis lever les freins et les réticences qui pouvaient exister ?

Pr Marc Cuggia. La création de la confiance est un sujet extrêmement important. Les centres de données cliniques sont créés pour exploiter les données produites par les différents services des hôpitaux de manière transversale. Cette chaîne de confiance est donc fondamentale.

La pédagogie se met en place très rapidement car la plateforme que nous avons développée répond à des questions posées par les cliniciens. Des réponses qu'il n'était pas possible d'obtenir auparavant s'obtiennent maintenant de manière plus fluide et plus facile. Il s'agit d'une pédagogie par l'exemple. Cela crée une boucle vertueuse, car la prise de conscience de ces résultats a un impact sur la qualité des données ; et cela crée un effet d'entraînement.

Les centres de données cliniques accueillent également en leur sein des cliniciens qui se forment au code, aux méthodes et aux technologies d'intelligence artificielle. Cela permet à nos collègues cliniciens de développer de nouvelles compétences et de s'approprier ces outils.

Le comité scientifique et éthique interrégional est un élément majeur de notre gouvernance. Il doit répondre aux problématiques de qualité scientifique des projets, d'intérêt de santé publique et de protection des données. Ce comité réunit les établissements membres d'HUGO ainsi qu'une association de patients et des éthiciens. Il porte donc un regard véritablement pluridisciplinaire sur ces enjeux.

Mme Laurence Jay-Passot. La pédagogie est un sujet permanent. Une maturité collective se développe dans le Grand Ouest depuis plusieurs années. Nous avons constaté une prise de conscience de l'intérêt du travail collaboratif, de l'intérêt de mutualiser les données et de mener des projets multicentriques au travers de nos réseaux thématiques. Les groupes de travail thématiques sont aujourd'hui tout à fait matures pour développer les usages des données massives en santé.

Nous associons toujours à la dimension d'infrastructure un autre volet : celui de la stimulation des usages et de l'animation de l'écosystème de la recherche et de l'innovation. En parallèle de la construction du dispositif, nous avons ainsi lancé un appel à projets pour démontrer d'emblée l'intérêt de la plateforme.

M. Philippe Latombe, rapporteur. Les patients sont-ils au courant de l'existence d'HUGO, de l'utilisation des données et de la protection qu'ils peuvent en attendre ? Dans le cas du Health Data Hub, le grand public s'est également posé de nombreuses questions sur l'utilisation des données de santé.

Mme Laurence Jay-Passot. Nous avons communiqué régulièrement sur le sujet en nous appuyant sur nos centres de données cliniques. Je n'ai pas le sentiment que les discussions sur le Health Data Hub aient rejailli particulièrement sur le déroulement de nos projets. Cela s'explique en partie par le fait que notre Ouest Data Hub ne naît pas de rien : il est l'aboutissement de la constitution des centres de données cliniques, au sujet desquels des communications régulières ont été faites auprès du grand public et qui respectent toutes les exigences de la Commission nationale de l'informatique et des libertés (CNIL) quant à la formation des entrepôts et à l'information des patients.

M. Philippe Latombe, rapporteur. Quelle est votre sensibilité aux cyberattaques, et plus généralement à la sécurité et à l'atteinte à l'intégrité des données ? Votre architecture fragilise-t-elle la sécurisation, ou au contraire, constitue-t-elle un avantage en la matière ?

Pr Marc Cuggia. Un certain nombre de mesures technologiques et organisationnelles sont nécessaires pour renforcer la sécurité. S'agissant des cyberattaques, il faut éviter les effets d'attractivité ou « pot de miel ». C'est une des raisons qui nous a conduit à ne partager des données que sur projets : ne sont donc mises à disposition sur la plateforme hébergée à Nantes que les données strictement liées à des projets.

Nos établissements entrent dans une logique de sécurisation. Le CHU de Nantes est hébergeur de données de santé, et le CHU de Rennes est en cours d'obtention de la labellisation. Les mesures de sécurité en sont donc considérablement augmentées. Nous sommes hébergés au sein du système d'information de l'hôpital. L'entrepôt de données collecte donc des données, qui sont stockées au sein de l'hôpital ; ces données sont par ailleurs toutes déjà présentes dans les différents outils utilisés par les médecins. Nous nous appuyons sur le savoir-faire de la DSI pour assurer la sécurité des données.

Les données sont pseudonymisées. Nous avons nourri des échanges très poussés avec les équipes techniques de la CNIL, qui nous ont beaucoup aidés. Ils nous ont exprimé un certain nombre de préconisations que nous avons mises en œuvre. Nous veillons à ce que les mesures de sécurité maximales puissent s'appliquer.

La cybersécurité est évidemment un sujet de recherche et développement. Dans le cadre du développement de la plateforme, nous menons des projets de recherche et innovation sur des solutions comme le crypto-tatouage de bases de données ou l'avatar des données. Nous souhaitons expérimenter ces solutions dans nos environnements. Les données massives en santé constituent un objet d'innovation en cybersécurité. Nous sommes sensibles au fait que les données confiées par les patients soient tout à la fois protégées et permettent l'innovation et la recherche, au bénéfice du patient.

M. Philippe Latombe, rapporteur. Développez-vous ces solutions en interne ou faites-vous appel à des solutions développées par des sociétés privées extérieures ?

Pr Marc Cuggia. Je vous présenterai le projet actuellement mené sur le crypto-tatouage des bases de données. Cette technologie a été développée dans un laboratoire cyber

universitaire et nous allons l'expérimenter dans le centre de données cliniques du CHU de Rennes, puis nous en évaluerons les performances. Cette technologie doit permettre d'assurer la traçabilité des traitements sur des données par tatouage des données.

L'enjeu est de créer un terrain multidisciplinaire pour appliquer ces objets de recherche dans nos domaines. Nous souhaitons, si les technologies que nous expérimentons sont suffisamment éprouvées et matures, qu'elles puissent être développées avec des start-up et qu'elles deviennent un objet industriel. Nous avons donc vocation à travailler avec des start-up et des industriels sur nos sujets. Nous travaillons également en partenariat avec des start-up pour le développement d'outils d'aide au diagnostic.

M. Philippe Latombe, rapporteur. Ce projet a nécessité une homogénéisation des données et des systèmes d'information. Cela a-t-il un effet d'entraînement sur d'autres domaines, pas nécessairement ceux de l'informatique et de la recherche ?

Mme Laurence Jay-Passot. Oui, cela a plusieurs effets d'entraînement. Nous cherchons à inventer un modèle de hub interrégional ; notre action a du sens si elle peut permettre la structuration d'autres hubs interrégionaux sur le territoire. Nous travaillons très étroitement avec deux autres réseaux de CHU constitués plus récemment qu'HUGO : le G4 dans le Nord et le groupement Grand Est. Ils suivent la même trajectoire de constitution d'un hub interrégional. Cela fait sens de disposer de hubs interrégionaux qui pourraient s'articuler avec le dispositif national, comme cela avait été imaginé dans le rapport de préfiguration.

Il existe également des sujets connexes pour lesquels un effet d'entraînement se fait sentir et permet d'avancer plus loin les dynamiques de partage de données et d'expertises. C'est le cas, par exemple, dans le Grand Ouest de la réflexion sur la génomique. Cette dynamique de partage est permise par la grande transversalité des données massives en santé.

Pr Marc Cuggia. Nous avons largement partagé notre expérience avec les autres établissements en matière de structuration des entrepôts de données, de qualité, de protection, de gouvernance, d'usages. Cela crée un effet d'entraînement national très important, qui se traduit par la mise en place d'entrepôts de données dans la plupart des CHU de France et dans les centres de lutte contre le cancer, ainsi que par la création d'équipes spécialisées dans ces domaines.

Je souhaite saluer l'initiative InterHop qui intervient sur des sujets très techniques. Il est nécessaire que l'on s'accorde sur des standards et des terminologies pour harmoniser nos données. J'insiste sur le fait que l'harmonisation n'est pas encore complète dans le Grand Ouest. Ce processus est progressif et nos efforts sont constants en la matière.

Je salue également les initiatives comme celle portant le modèle Osiris, qui normalise les données dans le champ de la cancérologie pour la réutilisation secondaire des données. Ces initiatives sont portées par des équipes d'informatique médicale et il faut absolument les encourager. Le Grand Ouest va développer ses propres projets, certes, mais l'enjeu est national, voire international. Ces travaux ne sont pas forcément très visibles pour le grand public mais ils sont extrêmement importants.

L'effet d'entraînement est également majeur pour les éditeurs de logiciels médicaux, qui développent des outils utilisés par les cliniciens. Il existe un enjeu majeur en matière de structuration et de normalisation des données ainsi que d'utilisation de standards dans nos systèmes d'information. L'Agence du numérique en santé a publié une stratégie en la

matière. L'innovation et la recherche vont bénéficier directement de ces efforts au long cours. Il est essentiel que les industriels mettent en œuvre les préconisations de standardisation, afin que les équipes des établissements de santé puissent facilement disposer des données.

L'enjeu de la qualité et de la maîtrise de la donnée est crucial tout au long de la chaîne de traitement de la donnée, depuis le lit du patient jusqu'à l'innovation. Une stratégie forte doit certainement être mise en place à ce sujet aux niveaux national et européen. L'Allemagne, par exemple, l'a fait.

Mme Laurence Jay-Passot. Nous souhaiterions que cet effet d'entraînement alimente également la réflexion sur le modèle économique des entrepôts de données et des hubs interrégionaux. Ce travail, que le Pr Marc Cuggia a très bien décrit, nécessite un fort investissement aussi bien en matière d'infrastructures que de ressources humaines. La région Grand Ouest a mené une action volontariste en la matière : elle a fait le choix d'aligner une énergie collective avec des choix financiers collectifs. Mais une réflexion doit s'engager sur son modèle économique qui, aujourd'hui, n'est pas du tout accompagné. Nous sommes convaincus que les hubs interrégionaux ont toute leur place pour collecter des données et les mettre en qualité, puis les articuler avec un dispositif national. Mais il va falloir nous aider collectivement à le faire. Je comprendrais aisément que toutes les interrégions et tous les groupements d'établissements ne fassent pas les mêmes choix que ceux que nous avons faits.

M. Philippe Latombe, rapporteur. Faut-il penser un modèle économique pour les hubs interrégionaux ? Faut-il que vous puissiez vendre les données ? Comment ces sources de revenus pourraient-elles être compatibles avec tous les efforts déployés en faveur de l'éthique et de la protection des données ?

Mme Laurence Jay-Passot. Ce modèle économique est encore à inventer, mais il est nécessairement mixte. La structuration des entrepôts de données et des centres de données cliniques n'est aujourd'hui pas intégrée dans le financement des hôpitaux. Cela soulève un sujet de financement.

Il pourrait également y avoir un sujet de structuration des hubs interrégionaux. La possibilité avait été évoquée de lancer des appels d'offres structurants pour les hubs interrégionaux.

La valorisation des données des entrepôts locaux et des plateformes interrégionales constitue un autre sujet. Nous ne pouvons pas vendre les données en tant que telles, mais nous pouvons valoriser par un échange monétaire le travail fourni par nos équipes en termes de mise en qualité des données et de réponse à une question scientifique. Dans un conventionnement, nous pouvons mettre en avant l'expertise que nos équipes sont capables d'apporter en mobilisant des données pour répondre à une question posée. Cela pourrait participer de la constitution d'un futur modèle économique, mais il existe encore peu de choses sur le sujet.

M. Philippe Latombe, rapporteur. Avez-vous chiffré le coût en investissements de la création d'HUGO ? Si un groupement devait aujourd'hui créer un tel dispositif en partant de zéro, combien cela lui coûterait-il ? De tels investissements constituent un choix pour l'avenir. Je m'interroge sur les fonds qui devraient être mobilisés pour créer ces dispositifs régionaux, par exemple dans le plan de relance actuel.

La création d'HUGO a-t-elle apporté un plus à l'écosystème de la recherche publique ou privée dans l'Ouest ? A-t-elle drainé, par exemple, l'installation d'entreprises ou de structures de biotechnologies ?

Mme Laurence Jay-Passot. Avant toute chose, une clarification terminologique : HUGO est notre réseau de CHU. L'Ouest Data Hub est l'un des projets qu'il porte en matière de données massives. Nous portons par ailleurs beaucoup de projets en soin, enseignement et recherche.

Le coût d'amorçage de la plateforme comprend un coût d'hébergement et un coût de ressources humaines. Ce coût est aujourd'hui de l'ordre de 300 000 euros par an, couvrant uniquement la partie mutualisée. En revanche, si l'on chiffre la contribution apportée par les centres données cliniques à chacun des projets, les coûts sont évidemment plus élevés : je les évalue entre 500 000 euros et un million d'euros à moyen terme.

Il n'est pas possible de créer une plateforme interrégionale sans disposer de centres de données cliniques et d'entrepôts de données de santé dans les établissements. Cela représente un vrai coût. Aucun centre de données cliniques n'est le même et ne requiert le même investissement en infrastructure et en ressources humaines. Néanmoins, nous savons qu'*a minima*, en conditions de fonctionnement pérennes, un établissement doit investir près de 400 000 euros chaque année pour faire fonctionner son centre de données cliniques déjà existant. Ce coût ne comprend pas la constitution initiale du centre de données. Ces sommes, mises bout à bout, pèsent sur les budgets. Ces évaluations sont, qui plus est, très minimalistes. Si l'on voulait inclure le temps investi pour faire grandir ce genre de projet, cela nécessiterait de chiffrer encore davantage de charges indirectes que je n'ai pas citées.

Pr Marc Cuggia. Je répondrai à votre question, s'agissant de la valorisation de la recherche. L'exploitation des données massives dans le champ de la santé est un sujet de compétition internationale. Nos centres sont encore beaucoup trop peu dotés, en termes de matériels et de chercheurs, pour pouvoir s'intégrer dans cette compétition, qui se développera de manière très forte dans les prochaines années. Des centres aux États-Unis sont, par exemple, composés de 75 enseignants-chercheurs permanents. L'Allemagne a fait le choix de mailler l'ensemble de son territoire avec des centres de données : leur stratégie a été de doter les établissements de *data information centers*. 120 millions d'euros ont donc été investis sur tout le territoire pour constituer des centres de données cliniques, créer des postes de permanents hospitalo-universitaires, embaucher des *data scientists*. Je pense que l'infrastructure ne fera pas l'innovation. L'enjeu principal réside dans le potentiel humain de formation et d'interdisciplinarité sur le terrain. C'est cela qui nous permettra d'avoir collectivement une chance de développer une souveraineté sur les enjeux de numérique en santé. L'écosystème des établissements publics à caractère scientifique et technologique (EPST), des start-up, des industriels est extrêmement important.

En termes d'attractivité, nous constatons de plus en plus de sollicitations d'acteurs qui souhaiteraient développer ou codévelopper avec nous des projets sur les données. Cela suppose de monter en charge et de construire notre capacité à accompagner ces acteurs. L'enjeu est pour nous très important de pouvoir répondre à cette dynamique. Nous essayons de mettre cela en œuvre avec le Ouest Data Hub, grâce à une gouvernance qui se veut transparente. Nous avons une volonté collective de réussir et d'innover. Mais il ne s'agit pas d'être dans une logique purement financière : en tant que chercheur, ce qui m'intéresse est que le contenu de nos recherches serve avant tout au patient. Cela est notre mission première.

Nous ne nous interdisons pas de travailler avec des industriels, mais l'enjeu est de faire grossir ces centres de données cliniques et cette expertise sur le territoire.

Mme Laurence Jay-Passot. Effectivement. Sinon, nous serons très limités par rapport au potentiel que représentent ces données.

M. Philippe Latombe, rapporteur. Croyez-vous que ce fonctionnement en hub, qui fonctionne pour les données massives en santé, peut servir à d'autres secteurs ? Cela vaudrait-il la peine de le développer pour d'autres domaines que les données de santé ?

Mme Laurence Jay-Passot. Évidemment. Nous avons fait le choix de privilégier ces modes de fonctionnement en hub ou en plateforme, et nous considérons que dès que la notion de masse critique fait du sens, nous devons construire une réponse collective. Cela ne veut pas dire centraliser et perdre le lien avec les initiatives locales. Le modèle de hub est absolument reproductible dans d'autres domaines, si on le conçoit comme un dispositif qui permet de mutualiser uniquement ce qui doit l'être et que l'on arrive à penser des systèmes de gouvernance agiles, qui continuent à s'appuyer sur toutes les compétences disponibles dans les centres locaux. Nous y croyons très fortement. Il existe plusieurs sujets – et les données massives en santé en sont un – pour lesquels le modèle de hub sera pertinent, s'il est dupliqué à l'échelle interrégionale sur l'ensemble du territoire.

Pr Marc Cuggia. Ces réflexions dépassent les données de santé *stricto sensu* et s'appliquent également à des projets mettant en œuvre des approches similaires, en lien avec les citoyens. Je citerai trois expériences auxquelles nous sommes associés. D'abord, Rennes Métropole travaille à mettre au point un portail des données personnelles qui vise à exploiter les données de transport, d'énergie, de santé à des fins de recherche et d'innovation. Une réflexion de hub pourrait être mise en place à ces fins. Ensuite, l'université de Sherbrooke au Québec a mis en place un projet similaire de système d'information apprenant ; le projet Pulsar de l'université de Laval, enfin, fait le lien entre les données de santé et les données de territoires. Des projets d'innovation très importants existent donc, qui ont un lien très fort avec les citoyens. Ces pistes me semblent extrêmement intéressantes à suivre pour pouvoir mettre en place une santé numérique durable.

M. Philippe Latombe, rapporteur. Souhaiteriez-vous aborder un dernier sujet en matière de données de santé, que nous n'avons pas déjà évoqué au cours de l'audition ? À quoi devons-nous être vigilants à l'avenir ?

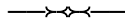
Pr Marc Cuggia. Il faut garder à l'esprit que nos activités doivent être au service des patients. Les innovations que nous développons ont utilisé des données de qualité très variables, avec des méthodes de *machine learning* qui peuvent parfois poser des problèmes d'interprétabilité. Il y a donc un enjeu majeur d'évaluation des produits d'innovation, c'est-à-dire des algorithmes. Les algorithmes doivent être évalués au même titre qu'un dispositif médical ou qu'un produit de santé. Les équipes des CHU, de l'Inserm, tout le maillage de recherche clinique et d'innovation doivent se saisir de cette opportunité. Dans le Grand Ouest, nous avons mis en œuvre plusieurs tests de ces algorithmes en vie réelle, à la manière d'un essai clinique. Nous devons fermer la boucle : nous collectons des données de santé, nous en tirons des connaissances qui permettent de construire des outils utiles au patient et au clinicien, qui devront enfin être évalués. Ces algorithmes devront être évalués sur les patients suivant une véritable démarche de recherche clinique. Nous devons pouvoir mettre en place ce cercle vertueux. Cela suppose également d'entrer dans des logiques de normalisation et de labellisation des centres de données cliniques. Cela constitue un élément majeur de la

confiance : il s'agit à la fois de la confiance du clinicien envers l'outil qu'il utilise tous les jours et de la confiance du patient.

Mme Laurence Jay-Passot. La réflexion du Pr Marc Cuggia a beaucoup de sens et remet toutes nos discussions en perspective.

M. Philippe Latombe, rapporteur. Je vous remercie.

La séance est levée à 13 heures 35.



Membres présents ou excusés

Mission d'information sur le thème « Bâtir et promouvoir une souveraineté numérique nationale et européenne »

Réunion du jeudi 18 février 2021 à 12 heures

Présents. - MM. Philippe Latombe, Christophe Lejeune, Jean-Luc Warsmann