

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 Stéphanie Combes, directrice du groupement d'intérêt public Plateforme nationale d'accès aux données de santé (Health Data Hub) 2

Jeudi

18 février 2021

Séance de 9 heures 30

Compte rendu n° 28

SESSION ORDINAIRE DE 2020-2021

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



Audition, ouverte à la presse, de Mme Stéphanie Combes, directrice du groupement d'intérêt public Plateforme nationale d'accès aux données de santé (Health Data Hub)

La séance est ouverte à 9 heures 30.

Présidence de M. Jean-Luc Warsmann, président.

M. le président Jean-Luc Warsmann. L'audition de ce jour s'inscrit dans notre cycle consacré à la souveraineté numérique et au numérique en santé. La structure Health Data Hub vise à faciliter l'accès à l'ensemble des données de santé, afin de promouvoir la recherche et l'innovation dans ce domaine. Son régime juridique s'appuie sur une base législative : il est défini à l'article 41 de la loi du 24 juillet 2019 portant sur l'organisation et la transformation du système de santé. La création officielle du Health Data Hub est intervenue le 30 novembre 2019. Elle s'inscrit dans le cadre du plan sur l'intelligence artificielle que le Président de la République a lancé en 2018 ainsi que dans une stratégie globale de numérisation accélérée de notre système de santé.

Ce dossier est un concentré des sujets 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 de la protection en santé de nos concitoyens et de rendre un service d'une qualité toujours croissante. 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, Mme Stéphanie Combes, de pouvoir échanger avec vous sur ces différents sujets.

M. Philippe Latombe, rapporteur. Je vous remercie d'avoir accepté d'échanger avec nous. Je souhaite vous interroger sur trois points en particulier.

J'aimerais d'abord que vous nous présentiez le Health Data Hub : son organisation, son fonctionnement, les principaux choix techniques opérés et son actualité pour l'année 2021. Il me semble important de donner le maximum de publicité à cette plateforme qui constitue un outil de pointe pour soutenir la recherche et l'innovation en santé.

Je souhaiterais également savoir comment l'action du Health Data Hub s'articule avec les autres hubs de données de santé en cours de constitution – le Ouest Data Hub par exemple, dont nous auditionnerons les représentants plus tard ce jour. Il s'agit pour nous de comprendre comment le Health Data Hub s'intègre dans 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 ? La décision de recourir à Microsoft pour héberger les données de santé recueillies par le Health Data Hub a fait l'objet de critiques et d'un recours devant le Conseil d'État. Le ministre des solidarités et de la santé, M. Olivier Véran, s'est finalement engagé à ce que le transfert du Health Data Hub vers un autre hébergeur que Microsoft intervienne dans un délai compris entre douze et dix-huit mois. Je souhaiterais donc bénéficier d'un point d'étape de votre part à ce sujet. Cela constituera également l'occasion d'échanger sur l'initiative européenne GAIA-X.

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 Stéphanie Combes, directrice du groupement d'intérêt public Plateforme nationale d'accès aux données de santé (Health Data Hub). Le projet du Health Data Hub a trouvé ses racines dans les travaux sur l'intelligence artificielle rendus par M. Cédric Villani en 2018. Il y faisait la promotion des plateformes de partage de données dans différents secteurs – la santé constituant un secteur prioritaire à ce sujet. Une mission de préfiguration et d'expertise a alors été commandée par la ministre de la santé de l'époque, Mme Agnès Buzyn. J'étais rapporteur de cette mission de préfiguration, dont le rapport a été rendu à la fin de l'année 2018. L'année 2019 a été consacrée à la préfiguration de cette structure, par des travaux législatifs et d'infrastructures technologiques. La structure a été créée à la fin de l'année 2019 et fonctionne maintenant depuis un peu plus d'un an.

Ce projet s'inscrit dans une démarche large portée par le ministère de la santé. La feuille de route du numérique en santé est particulièrement ambitieuse. Le Health Data Hub est l'une des trois plateformes numériques qui doivent être bien articulées : une plateforme pour les citoyens, proposant des applications pour les aider dans leurs soins ; une plateforme pour les professionnels de santé, prévoyant des outils pour les accompagner dans leurs activités ; et, enfin, le Health Data Hub qui s'apparente à une plateforme de recherche et développement.

La structure Health Data Hub regroupe 56 parties prenantes, réparties en neuf collèges – ce hub doit représenter l'ensemble de l'écosystème des données de santé, qui est extrêmement vaste. L'État, bien sûr, y est présent : onze directions de l'État sont parties prenantes ; parmi elles, une direction du ministère de la santé et une direction du ministère de la recherche siègent au conseil d'administration. La Caisse nationale de l'assurance maladie (CNAM) est également un partenaire clé du projet. Les autres collèges sont formés par les organismes d'assurance maladie complémentaire ; les établissements de recherche et d'enseignement ; les établissements de santé ; les représentants des professionnels de santé et des usagers ; les agences, opérateurs et autorités publiques indépendantes ; et, enfin, les industriels. L'assemblée générale réunit l'ensemble de ces parties prenantes ; le conseil d'administration, quant à lui, réunit un représentant de chacun des collèges, à l'exception de l'État qui en a deux. Le financement du Health Data Hub est majoritairement public : la structure bénéficie de vingt millions d'euros par an, dont une moitié provient de l'objectif national de dépenses d'assurance maladie (ONDAM) et l'autre moitié du fonds de transformation de l'action publique, un outil de financement de l'innovation.

Le Health Data Hub doit fournir un accès simplifié aux données de santé en France pour améliorer la qualité des soins et l'accompagnement des patients. La plateforme s'adresse aux acteurs qui animent des projets de recherche et poursuivent une finalité d'intérêt public ; ceux-ci doivent soumettre un dossier pour accéder aux données de santé. La mission du Health Data Hub s'articule autour de quatre missions principales. Tout d'abord, la plateforme est un guichet unique. L'institut national des données de santé (INDS), créé par la loi de 2016, jouait déjà ce rôle : un porteur de projet devait monter un dossier de demande pour accéder aux données de santé ; celui-ci était soumis à la Commission nationale de l'informatique et des libertés (CNIL), seule autorité compétente pour autoriser un traitement de données ; enfin, le porteur de projet déposait son dossier à l'INDS.

Nous avons cherché à élargir les missions de ce guichet. En effet, une fois que la CNIL avait autorisé le traitement de données, le porteur de projet pouvait rencontrer des difficultés d'accès aux données car celles-ci pouvaient être éparpillées ou bien impossibles à traiter en raison de problèmes technologiques. Par le Health Data Hub, nous souhaitons apporter un service jusqu'à la réalisation de l'étude. La plateforme sécurisée donne accès à un espace projet, où nous versons les données autorisées dans le cadre des projets et les utilisateurs pourront les traiter avec des logiciels de programmation à l'état de l'art. L'utilisateur n'a accès ni aux données des autres utilisateurs, ni à davantage de données que nécessaire pour son projet. De la même manière, il ne peut pas sortir les données de l'espace projet et toutes ses activités sont tracées. Mais cela n'est pas une obligation : le Health Data Hub n'est pas une plateforme unique. Si certains acteurs disposent déjà de plateformes technologiques au bon niveau de sécurité, alors il n'est pas besoin de passer par le Health Data Hub. Cependant, la mission de préfiguration a montré que beaucoup d'acteurs ne disposent pas de telles plateformes et il est très coûteux de mettre sur pied une infrastructure dotée des bons niveaux de sécurité et des fonctionnalités adéquates.

La plateforme technologique met à la disposition de ses utilisateurs un catalogue de données. La loi a élargi le système national des données de santé (SNDS) à l'ensemble des données associées à un remboursement de l'assurance maladie. Il est évident que toutes ces données ne seront pas versées à la plateforme technologique du hub – cela n'est ni viable ni souhaitable techniquement, financièrement et scientifiquement. Nous nous posons plutôt la question suivante : quelles données du patrimoine de données de santé français sont intéressantes pour la communauté scientifique et de l'innovation ? Un comité stratégique, piloté par l'État, élaborera une priorisation de ces données. Suite à cette priorisation, une liste des bases du catalogue sera établie : elle sera prise par arrêté après avis de la CNIL et sera mise à jour régulièrement.

La dernière mission du Health Data Hub est une mission d'animation. Beaucoup d'initiatives existent sur le territoire et il est important de les agréger afin de générer un impact à l'échelle européenne et internationale. Il faut, pour y arriver, connecter entre elles les initiatives existant dans les différents établissements de santé – cela permettra d'atteindre une masse critique. Le Health Data Hub poursuit cet objectif.

Le SNDS, créé en 2016, est la base des feuilles de soin pseudonymisées pour servir à des fins de recherche. Il constitue une base unique en son genre au niveau international, car la centralisation du système de santé français permet d'obtenir la consommation de soins de l'ensemble de la population. Cela est extrêmement intéressant pour la recherche. Néanmoins, il s'agit d'une base médico-administrative : elle ne comprend, par exemple, ni résultats d'analyses, ni scanners. Il est donc intéressant d'associer cette base avec d'autres.

C'est ce qu'a fait la loi sur l'organisation et la transformation du système de santé en 2019 : cette loi a créé le hub et a élargi le SNDS. Le SNDS constitue un système de base, qui regroupe des données de registre, des cohortes de recherche, des entrepôts de données hospitalières et la base de l'Assurance maladie. Le catalogue du hub présentera une sous-catégorie, constituée par les données du SNDS, dont les bases seront chaînées avec celles de l'Assurance maladie pour élargir la capacité à les réutiliser. Cela est fait en partenariat avec les gestionnaires de bases de données : une convention est signée qui fixe les modalités et les règles de partage. Ainsi, nous discutons aujourd'hui avec un grand nombre d'acteurs pour concevoir ce catalogue – il ne s'agit pas du tout d'un mécanisme d'aspiration des données, comme cela a pu être dit.

Une première version du catalogue a pris forme dans l'entrepôt COVID, autorisé par l'arrêté du 21 avril 2020. Il rend disponible la base du SNDS *fast-track* avec les données de l'Assurance maladie issues du programme de médicalisation des systèmes d'information (PMSI) et les données de Santé publique France. Les données du système de vaccins y seront également bientôt disponibles.

Que nous manque-t-il pour mettre en œuvre ce catalogue, qui constitue l'ambition clé du Health Data Hub ? Un décret est toujours manquant – sa publication est attendue au mois de février ou de mars 2021. Le comité stratégique doit également se réunir, afin de définir la liste des bases du catalogue, qui sera prise par arrêté après avis de la CNIL. Il est essentiel de pouvoir poser tous ces jalons avant le milieu de l'année 2021. Le Health Data Hub a été lancé en 2019 ; nous ne pouvons pas nous permettre d'accumuler du retard dans la mise en place du catalogue.

L'accès aux données du catalogue ne se fait pas du tout en *open data* – les règles habituelles s'y appliquent. En revanche, le caractère centralisé des données facilite les temps d'accès, l'accès à des logiciels métier adaptés – notamment en cas d'usage pour l'intelligence artificielle – et les chaînages. Ainsi, le Health Data Hub ne va pas remplacer les autres initiatives, notamment les plateformes locales déjà existantes. Par exemple, l'entrepôt de l'Assistance publique-Hôpitaux de Paris (AP-HP) n'a pas besoin du Health Data Hub pour conduire la plupart de ses projets. Mais le Health Data Hub pourrait l'intéresser pour certains d'entre eux – ainsi, nous travaillons actuellement avec l'AP-HP sur une dizaine de projets, ce qui prouve bien nos intérêts communs et complémentaires.

Le Health Data Hub permettra en premier lieu le croisement des sources. Un projet s'appuyant sur les données d'un établissement de santé est intéressant, mais il lui manque les données de santé de ville et peut-être les données des autres établissements de santé. Le chaînage entre les données de l'Assurance maladie et celles de l'établissement de santé permettra d'adopter une approche en parcours de soins et de conduire des projets qu'il n'est pas possible de concevoir autrement. Aujourd'hui, un chaînage, c'est-à-dire un croisement de sources de données, peut durer trois à quatre ans en France. L'ambition du Health Data Hub est de réduire ce délai à six ou neuf mois.

L'autre intérêt du hub réside dans ses capacités élastiques de calcul et de stockage – c'est la raison pour laquelle nous avons choisi une infrastructure *cloud*. À titre d'exemple, une clinicienne de l'AP-HP développe un projet d'aide au dépistage du cancer de la prostate qui mobilise le traitement de 10 000 imageries par résonance magnétique (IRM) par l'utilisation du *deep learning*. Elle travaille actuellement avec plusieurs établissements de santé mais aucun d'entre eux ne peut réunir au même endroit les 10 000 IRM et ne dispose des capacités de calcul et de stockage suffisantes pour appliquer les algorithmes – c'est pourquoi elle a eu recours aux services du hub.

Nous avons eu l'occasion de poser quelques jalons en 2020. Nous avons tout d'abord mis en production une plateforme pour des projets liés au COVID et avons procédé à sa mise à jour en fin d'année. Nous avons également conduit un second appel à projets au terme duquel nous avons sélectionné dix nouveaux projets sur le thème de l'intelligence artificielle et de la santé. Nous accompagnons aujourd'hui quarante projets : parmi eux, trente sont des projets pilotes et dix sont des projets liés au COVID – huit d'entre eux ont déjà reçu une autorisation de la CNIL et les deux derniers sont en attente de la recevoir. Nous animons des discussions avec les partenaires du catalogue ainsi que des actions de fédération de l'écosystème : un colloque, un *data challenge*, une *winter school* à laquelle 400 personnes se sont inscrites. Le hub rassemble aujourd'hui une cinquantaine de collaborateurs et nous

formons l'ambition d'être soixante-dix en 2021. Enfin, nous sommes impliqués dans les travaux européens – j'y reviendrai.

Nous avons cinq priorités en 2021 :

- continuer la mise en place de la structure créée il y a un an ;
- industrialiser l'accompagnement des projets de bout en bout ;
- mettre en place des partenariats stratégiques et les développer, par exemple avec la CNAM, l'Institut national de la santé et de la recherche médicale (Inserm) et les établissements de santé ;
- associer le grand public au dispositif, en particulier compte tenu de la sensibilité des données ;
- enfin, développer l'infrastructure technologique.

Comment accéder aux données présentes dans le hub ? La réglementation française est très précise à ce sujet. Les porteurs de projets doivent constituer un dossier de demande d'autorisation d'accès aux données auprès de la CNIL. En la matière, le hub fait office de simple guichet administratif : le porteur de projet dépose son dossier au hub, qui le transmet à un comité éthique et scientifique national, qui vérifiera la solidité du projet tant du point de vue de la finalité poursuivie que de la méthodologie. Ce comité sera composé d'experts de très haut niveau, reliés à des experts extérieurs spécialistes de sujets très pointus. Si le comité accorde un avis favorable au dossier, le hub le transmettra à la CNIL. La CNIL est la seule autorité en mesure de délivrer ou non une autorisation de traitement de données. Le porteur de projet pourra ensuite s'adresser au hub en sa qualité de gestionnaire d'infrastructure afin qu'il mette à disposition les données, mais cela n'est pas obligatoire. Le porteur de projet doit être transparent sur l'objet de son étude, qui doit être décrite sur le site Internet du Health Data Hub et figurer au répertoire public disponible en ligne, ainsi que partager une partie des résultats de sa recherche.

Vous m'avez interrogée sur les articulations du Health Data Hub avec les initiatives locales, et en particulier avec le Ouest Data Hub. Nous conduisons un projet pilote avec le groupement de coopération sanitaire des hôpitaux universitaires du Grand Ouest (HUGO) qui porte le projet du Ouest Data Hub. Le Pr Marc Cuggia, impliqué dans ce projet, a participé à la mission de préfiguration du Health Data Hub. La convergence entre les initiatives a donc été imaginée dès le départ. Il n'est absolument pas dans l'intérêt du Health Data Hub de remplacer les initiatives locales : au contraire, il faut bien plus d'entrepôts de données hospitalières qu'il n'en existe aujourd'hui. Nous souhaitons permettre le passage à l'échelle par le développement de projets d'envergure nationale et le croisement des données.

Le projet que nous conduisons avec HUGO s'appelle Hugo-Share. Il vise à analyser les trajectoires médicamenteuses de 420 000 patients afin de comprendre et d'éviter les accidents iatrogéniques en ville et à l'hôpital et afin d'améliorer les parcours de soins des patients les plus fragiles, notamment les personnes âgées. Le Health Data Hub cofinance le projet et accueille la base chaînée au sein de la plateforme. La base clinique, elle, est fournie par le Ouest Data Hub et ses six établissements partenaires. Nous menons d'autres partenariats de ce type avec d'autres acteurs : par exemple, avec la Fédération nationale des centres de lutte contre le cancer (Unicancer) afin de mutualiser les données et de construire une base oncologique d'envergure, dans le respect de la réglementation et du droit des patients.

Vous m'avez également interrogée sur les raisons d'être et le rôle de la direction citoyenne. Son rôle s'articule autour de l'obligation légale du hub : informer les patients, promouvoir et faciliter leurs droits. La direction citoyenne ne remplace pas la direction juridique. Nous identifions un enjeu à « embarquer » avec nous la société civile : il ne s'agit pas seulement d'informer les citoyens, mais de faire d'eux des partenaires du dispositif.

Cette direction anime quatre actions. La première consiste à étudier, consulter et concerter avec la société civile : nous constituons des groupes de travail afin de comprendre les attentes de la société civile et de recueillir sa perception du partage des données de santé. Nous sommes très impliqués dans l'action conjointe de la Commission européenne afin de mettre en place un espace commun de données de santé : le hub est l'autorité française compétente pour coordonner le travail des acteurs français sur le sujet. Ainsi, nous sommes chargés d'un lot de travaux relatif à l'infrastructure (cela concerne GAIA-X) ainsi que d'un lot de travaux sur l'engagement. Dans ce dernier lot de travaux, nous mettons en place une e-consultation pour nous adresser très largement à la société civile.

Nous cherchons également à mettre en place des partenariats concrets : par exemple, les associations de patients peuvent être autorisées à traiter des données. Nous avons ainsi conduit un premier partenariat avec France Asso Santé autour de l'étude intitulée « Vivre COVID », afin d'étudier comment les patients chroniques vivaient le premier confinement. Dans ce cas de figure, le hub apporte son appui à l'association de patients qui réalise l'étude et apporte ses données. Nous étudions actuellement comment le hub pourrait faire des requêtes au nom des citoyens ou des associations, quand ceux-ci ne sont pas en mesure de le faire eux-mêmes.

Il est également essentiel d'informer et de vulgariser, car la donnée de santé est particulièrement abstraite et complexe. Nous mettons en place des baromètres pour évaluer la connaissance ainsi que des outils de communication les plus vulgarisés possibles. Nous avons déjà produit deux vidéos et rédigé une note d'engagement avec un groupe de travail de patients. Nous souhaitons construire un contenu facile à lire et à comprendre sur le site Internet, afin de le rendre accessible au plus grand nombre de personnes. Cela pose évidemment de nombreuses questions sur l'exercice des droits, et ouvre également un chantier technologique sur la mise en œuvre des droits.

La direction citoyenne recouvre enfin une dimension de formation. Nous sommes actuellement en discussion avec l'Éducation nationale afin de mettre en place des outils de vulgarisation.

La fin de ma présentation porte sur le sujet qui, je pense, vous intéresse principalement : la sécurité et la souveraineté de la plateforme technologique. Je souhaite revenir sur les étapes qui nous ont amenées à choisir Microsoft et vous détailler où nous nous situons aujourd'hui au regard des décisions prises.

À l'été 2018, la ministre de la santé confie à la direction de la recherche, des études, de l'évaluation et des statistiques (DRESS) la feuille de route élaborée lors de la mission de préfiguration. À cette occasion, nous avons rencontré énormément d'acteurs : principalement des industriels français et des acteurs du monde de la recherche. Nous avons ensuite élargi notre champ, puisque nous nous sommes rendus compte que nos exigences étaient assez élevées. La première de nos exigences était la sécurité – cela n'a pas été tout de suite compris dans les débats qui ont eu lieu l'année dernière. D'aucuns pensent que nous avons choisi Microsoft en raison de ses capacités de *machine learning* ; en réalité, nous avons choisi Microsoft pour les services managés de sécurité. Il est extrêmement important de

comprendre cela. Il n'existe aucun niveau équivalent dans l'industrie française en matière de services managés de cybersécurité, de ségrégation des droits, de gestion des droits, de traçabilité totale des activités de la plateforme. Or c'est cela que nous recherchions spécifiquement. Nous n'avons pas décidé nous-même de ces exigences de sécurité. Celles-ci sont réglementaires : elles sont issues du référentiel de sécurité du Système national des données de santé, qui est très peu connu du grand public. La plupart des gens connaissent la certification des hébergeurs de données de santé (HDS), qui n'est pas obligatoire dans notre cas, mais bienvenue. C'est bien le référentiel de sécurité du SNDS qui est « incontournable » et qui constitue une exigence légale.

Nous avons également des exigences de performance : comme il ressort des exemples que j'ai précédemment donnés, nous souhaitons pouvoir croiser beaucoup de données et faire tourner du *deep learning* sur des IRM en masse, par exemple. Nous devons donc avoir une capacité à *scaler*, c'est-à-dire disposer d'une infrastructure capable de changer de taille en fonction des projets. C'est ce que le *cloud* nous permet de faire.

Enfin, nous posons une exigence de délai. Ce projet ne peut pas attendre : nous devons développer les usages numériques en santé et la crise a conduit, je crois, à une prise de conscience générale à ce sujet.

Nous avons ainsi choisi la solution de Microsoft, qui était la seule à répondre à toutes ces exigences. Il faut bien comprendre que notre plateforme n'est pas confiée à Microsoft : nous avons choisi le logiciel Azure de Microsoft, et Microsoft est, à ce titre, l'un de nos partenaires technologiques. Nous travaillons avec une dizaine de partenaires technologiques : la start-up française de cybersécurité Wallix, par exemple, nous fournit le bastion. Le Health Data Hub est souvent résumé à Microsoft ; mais nous n'avons pas demandé à Microsoft de construire une plateforme pour répondre à nos besoins. Microsoft est l'un de nos partenaires technologiques. L'un de nos plus gros partenaires technologiques est, d'ailleurs, la société française Open, qui est notre intégrateur.

Nous avons, dès le départ, posé la réversibilité de la plateforme comme l'une de nos exigences. Cela aurait été le cas même si nous n'avions pas eu recours à Microsoft. La réversibilité est indispensable. Dans tous les cas, il ne faut pas se retrouver piéger dans une solution technologique. Nous développons la plateforme en infrastructure *as Code*, c'est-à-dire programmable : nous essayons de disposer de scripts et de faire le moins d'actions manuelles possible. Nous pourrons alors, le moment venu, réutiliser les programmes informatiques : nous devons alors les reparamétrer en partie, évidemment, mais une grande partie du travail sera réutilisable – c'est ce que l'on appelle la réversibilité.

Qu'avons-nous réalisé en faveur de la réversibilité, hormis cette automatisation ? Nous avons conduit deux études de réversibilité à ce jour. Cette exigence de réversibilité est inscrite dans la feuille de route et partagée par tous les acteurs partenaires du groupement d'intérêt public. À la fin de l'année 2019, nous avons publié notre première étude qui comparait l'acteur français OVH avec Microsoft et identifiait un écart important entre les deux. En juin 2020, nous avons mis à jour cette étude avec la direction interministérielle du numérique (DINUM). Nous avons alors identifié les quatorze besoins indispensables de la plateforme. Pour le moment, OVH n'en couvre que cinq. Ceci étant dit, nous n'avons aucun doute sur le fait que les acteurs français ont mis au point des feuilles de route très ambitieuses et vont progressivement réduire cet écart.

La souveraineté est un objectif pour chacun d'entre nous. Nous devons, nous, combiner cet objectif avec d'autres : ainsi, notre objectif prioritaire est de servir les patients.

Les mesures de sécurité que nous avons mises en place ont été approuvées. Le débat ouvert aujourd'hui sur la souveraineté de la plateforme ne concerne donc pas la sécurité. Les données sont pseudonymisées et chiffrées. Nous avons recours à énormément de services de cybersécurité. Nous avons fait réaliser plusieurs audits par des prestataires d'audit de la sécurité des systèmes d'information (PASSI) qualifiés par l'ANSSI. Un audit de l'ANSSI est actuellement en cours. Nous avons reçu un avis de la DINUM, un considérant du Conseil d'État dans son ordonnance a souligné le haut niveau de sécurité de la plateforme. La CNIL a déjà autorisé huit projets. Il est maintenant clair que la plateforme a atteint un très haut niveau de sécurité. Cela n'empêche cependant pas de poursuivre un objectif de souveraineté ; la difficulté est de savoir quels objectifs se cachent derrière la notion de souveraineté. Cela n'est pas clair pour l'instant.

Nous identifions actuellement des financements dans le cadre de France Relance, et menons des discussions avancées avec l'ANSSI pour construire une plateforme souveraine qui pourrait être la cible de la migration que vous évoquiez. Nous sommes très impliqués dans l'action conjointe de la Commission européenne, qui a été lancée le 1^{er} février et devrait durer deux ans. Le lot de travaux concernant l'infrastructure pourrait faire le lien avec GAIA-X. L'initiative French Gaia-X Hub, quant à elle, a été lancée il y a quelques semaines : elle met en place des groupes de travail thématiques et nous participons notamment au groupe de travail des « utilisateurs santé ». Nous souhaitons travailler collectivement à construire une solution plus souveraine, tout en gardant en tête que le Health Data Hub n'est pas nécessairement la seule dimension d'un espace commun de données de santé à l'échelle européenne.

Par ailleurs, et pendant ce temps, nous continuons nos échanges très réguliers avec les différents acteurs de la filière. Nous avons récemment conduit un premier échange au sujet de la solution Anthos avec OVH et Google. Nous continuons donc évidemment à suivre l'évolution du marché.

M. Philippe Latombe, rapporteur. Vous avez souligné, à la fin de votre intervention, le haut niveau de sécurité du dispositif et indiqué que la souveraineté était une autre question. Vous avez alors fait remarquer que la définition de la souveraineté n'était pas claire. Que signifie, selon vous, la souveraineté ?

Mme Stéphanie Combes. Je ne suis pas compétente moi-même pour la définir, mais les discussions que nous avons pu avoir au sujet du Health Data Hub font ressortir un enjeu d'autonomie stratégique. Il s'agit de savoir si l'on est en situation de dépendance et si nous pouvons nous assurer que les données puissent être à tout moment récupérées. Cette notion recouvre également un enjeu lié aux lois extraterritoriales, s'agissant des hébergeurs de *cloud* américains. La question des lois extraterritoriales est assez technique – plusieurs sont mises en avant, dont le *Cloud Act* américain – et il convient d'étudier le sujet de manière assez fine. Ces lois s'appliquent dans certains contextes, qui ne sont pas forcément valables pour tous les hébergements de toutes les données personnelles. Dans le cas du hub, les données sont pseudonymisées.

Se posent également des questions de filières. Nous pourrions souhaiter soutenir le plus possible les acteurs de la filière française – d'autant plus que le projet est porté par l'État. Cela crée des injonctions parfois contradictoires : l'État porte un projet, donc il ne souhaite travailler qu'avec des acteurs français ; en même temps, l'État nous demande d'être rapides et d'avoir des résultats concrets.

Il faut vraiment traiter cette question de la souveraineté – mais je ne pense pas être la personne pour le faire. Ma crainte serait qu'elle ne soit pas complètement traitée, c'est-à-dire que l'on n'atteigne pas une définition conceptuelle claire, et que cela ait un impact non maîtrisé sur l'écosystème. On parle de loi extraterritoriale – mais parle-t-on des actionnaires étrangers ? Tous ces critères doivent être élaborés et partagés, et nous devons nous mettre d'accord sur la cible. Sinon, nous allons mettre en place des critères extrêmement restrictifs et nous allons supprimer des usages.

J'ai une seconde crainte. J'entends beaucoup parler du *cloud* et de Microsoft, mais je n'entends pas beaucoup parler de la souveraineté des usages numériques de santé. Dans certains autres pays, et notamment aux États-Unis, ces questions avancent très vite. En mai 2018, le dispositif médical pour les examens de fond d'œil était le premier dispositif médical intégrant l'intelligence artificielle à être autorisé par la *Food and drug administration* (FDA). Cela a constitué une très belle avancée et depuis, une trentaine d'autres dispositifs médicaux ont été autorisés par la FDA. Nous téléchargerons bientôt toutes ces applications sur nos téléphones, car elles proposeront des usages de santé extrêmement intéressants et performants ; mais elles n'auront pas été construites grâce à des données de patients français et l'on ne saura même pas si elles ont été développées dans le respect du Règlement général sur la protection des données (RGPD). Il faut donc garder en tête les questions sur la souveraineté des usages, afin de ne pas nous retrouver dans cinq ans à discuter de ces mêmes sujets car nous aurons pris du retard par rapport à d'autres acteurs. Il faut donc procéder à des arbitrages en ayant bien en tête tous les enjeux ayant cours au même moment. Cela n'est pas simple.

M. Philippe Latombe, rapporteur. Les critiques ne portent pas sur le fond du projet du Health Data Hub ni sur les objectifs qu'il poursuit. Elles visent, de façon très claire, l'hébergement dans le *cloud* d'Azure de Microsoft. Ces critiques sont présentes quasiment depuis le lancement du projet. Comment les appréhendez-vous et comment les intégrez-vous à votre démarche ? Vous avez insisté sur le fait que Health Data Hub devait être développé rapidement. Ces critiques ne constituent-elles justement pas un frein à son développement ? Je pense notamment à la procédure lancée devant le Conseil d'État, à la réticence de la CNAM quant au transfert de ses données, à l'expression permanente de critiques de la part de l'écosystème, à l'intervention du secrétaire d'État chargé du numérique devant le Sénat pour évoquer la réversibilité. Quels sentiments ces critiques suscitent-elles chez vous ?

Mme Stéphanie Combes. Il faut prendre les critiques de manière précise. Quand j'ai commencé à conduire le projet, on a critiqué sa lenteur annoncée, en prédisant que le Health Data Hub connaîtrait le même destin que le dossier médical partagé (DMP). Si j'avais choisi une solution qui n'aurait pas aujourd'hui permis au projet d'aboutir, je serais également critiquée. En tant que chef de projet d'un projet d'État, je suis assez à l'aise avec l'idée d'être critiquée. On ne va pas encore assez vite, je vous l'accorde – mais nous n'aurions même pas encore esquissé le début d'une plateforme si nous avions opté pour une autre solution.

Il faut donc prendre les critiques, étudier où elles prennent leurs sources et les traiter. Certaines critiques relèvent d'un problème de compréhension technologique – il faut donc faire de la vulgarisation. Il est vrai que l'État et les activités publiques sont moins à l'aise avec les solutions de *cloud* que le secteur privé. La DINUM travaille sur ces sujets, notamment en élaborant une stratégie *cloud* souveraine. Nous nous inscrivons dans cette dynamique.

Le conseil de la CNAM ne représente pas à lui seul l'ensemble de la CNAM. Le conseil de la CNAM ne s'est pas exprimé seulement sur Microsoft : il a également critiqué le

décret, qu'il a jugé incompréhensible. Il faut étudier l'origine du problème et le traiter. La mise en place des dispositifs de *cloud* relève de la conduite du changement ; ce projet est très différent des choix opérés par l'Assurance maladie depuis des dizaines d'années. Nous devons donc accompagner ce mouvement.

Les critiques proviennent souvent de l'écosystème du numérique français. Les acteurs du numérique français trouvent injuste que nous ayons opté pour Microsoft comme opérateur d'un projet public. Je comprends cette critique, mais si nous avions opté pour un autre partenaire que Microsoft, certains autres acteurs encore auraient été mécontents. On ne pourra jamais satisfaire tous les prestataires industriels. Nous sommes confrontés au jeu de la construction et de la sélection des partenaires technologiques.

Une autre critique provient du monde hospitalier, qui me semble pouvoir être traitée par la preuve. Le secteur hospitalier craint que le Health Data Hub cherche à remplacer ses activités. Il est intéressant, à ce sujet, d'étudier l'historique des critiques formulées par l'association InterHop et de savoir qui a fondé cette association. Il s'agit de deux anciens personnels de l'entrepôt de données de santé de l'AP-HP : leur propos premier était d'affirmer qu'il n'était pas besoin de mettre au point le Health Data Hub car la solution de l'AP-HP existait déjà. Nous sommes parfaitement d'accord avec l'idée selon laquelle il ne faut pas centraliser et il ne faut pas tuer les activités locales : nous n'allons pas remplacer le travail de l'Assurance maladie, ni celui de l'entrepôt de données de l'AP-HP. Ils fournissent un travail énorme, que nous respectons et que nous voulons soutenir. L'association InterHop utilise aujourd'hui l'argument du logiciel libre. Nous n'avons pour le moment pas bien compris quels éléments ils ont versé en *open source*.

Il faut donc identifier chaque source de critique, étudier comment la traiter et apporter tous les éléments de preuves qui nous sont demandés. C'est de cette manière que nous pourrions conduire l'innovation : à ce titre, le Health Data Hub est loin d'être le premier projet à souffrir la critique.

M. Philippe Latombe, rapporteur. N'avez-vous pas l'impression que les critiques, qui sont apparues dès le début du projet, entravent vos activités – qu'elles sont en quelque sorte devenues un boulet ? J'en veux pour preuve la réversibilité, qui a été l'une des conséquences de ces critiques.

Mme Stéphanie Combes. Non, la réversibilité a été identifiée dès le premier jour.

M. Philippe Latombe, rapporteur. Je l'ai bien compris. Cette réversibilité, cependant, est aujourd'hui prévue à deux ans.

Mme Stéphanie Combes. Le courrier du ministre diffusé dans la presse ne mentionne pas la migration, mais l'annulation du risque extraterritorial. Nous travaillons, ensuite, dans un objectif de souveraineté, car cet objectif est poursuivi par l'ensemble des services de l'État. Il faut étudier les sujets précisément. Vous pouvez, si vous voulez, évoquer un boulet : je préfère parler d'agilité. Si l'on avait procédé différemment, aucun projet n'aurait été conduit sur le hub aujourd'hui et aucun d'entre eux ne produirait des résultats dès cette année. Disposer d'une cible industrielle conforme à d'autres ambitions que celles d'abord exprimées par la ministre, auxquelles je souscris entièrement, d'avoir des résultats rapides représente une progression et ouvre de nouveaux chantiers. Il s'agit d'un projet extrêmement ambitieux, il est donc normal que nous ayons plusieurs étapes de réalisation – cette étape d'infrastructure souveraine en fait partie.

M. Philippe Latombe, rapporteur. Pensez-vous que le projet aurait été mené plus rapidement sans ces critiques de départ ?

Mme Stéphanie Combes. Non. Cela n'a pas de lien avec la question de la parution du décret. L'impact de la crise sanitaire sur le ministère suffirait à l'expliquer. En revanche, le ministère a été très réactif sur l'arrêté qui nous a permis de préfigurer le catalogue avec les données du COVID. Je ne crois donc pas que nous aurions pu faire plus vite en procédant différemment.

L'étude de réversibilité conduite l'été dernier avec la DINUM montre qu'OVH, le leader français dans le domaine, ne satisfait toujours pas nos prérequis. Nous serions donc encore aujourd'hui en train de construire une solution.

M. Philippe Latombe, rapporteur. Je voudrais que les choses soient claires. Comment s'est déroulé le processus de sélection au tout départ ? Avez-vous lancé un appel d'offres ? Si oui, quels en étaient les critères ? Quelles entreprises ont été autorisées à y répondre ? Beaucoup de critiques portent également sur la phase de démarrage du projet.

Mme Stéphanie Combes. Il n'y a pas eu d'appel d'offres. Nous avons recours aux services de l'union des groupements d'achats publics (UGAP). Tout comme la centrale d'achat de l'informatique hospitalière (CAIH) ou le réseau des acheteurs hospitaliers (Resah), l'UGAP est une centrale d'achat mise en œuvre pour faciliter l'achat public. Ces structures mettent elles-mêmes en concurrence les acteurs, conçoivent un catalogue dans lequel nous pouvons choisir. Il ne s'agit donc pas d'un contournement du code des marchés publics. La mise en concurrence a été faite, mais non ciblée sur notre projet. Nous nous appuyons donc sur l'UGAP. Nous sommes aujourd'hui une petite structure de 50 personnes, qui n'est pas en capacité de porter un marché de 200 millions d'euros.

Le ministère de la santé a élaboré à l'époque les prérequis : ceux-ci relèvent à la fois du juridique et de la cybersécurité. Nous avons rencontré les industriels et les avons interrogés, à chaque fois, sur le référentiel de sécurité du SNDS, qui est différent de la certification HDS. Le prérequis HDS était également important pour nous : nous n'envisagions pas de sélectionner une structure qui ne soit pas certifiée HDS. Enfin, les prérequis demandés concernaient les fonctionnalités, puisque nous souhaitions mettre en place une plateforme dotée de capacités de calcul et de stockage élastiques. Un document, résumant tous ces prérequis, a été rendu public sur le site de la direction de la recherche, des études, de l'évaluation et des statistiques (DRESS).

Nous avons par ailleurs conduit des échanges bilatéraux avec tous les industriels de l'écosystème français déjà évoqués – Atos, Thales, OutScale, CASD, TeraLab – pour étudier leurs offres. La première étude de réversibilité cote la présence, ou non, des certifications ou des fonctionnalités chez chacun des acteurs. Ces résultats auraient pu être « challengés » par les acteurs, mais cela n'a pas été fait – et pour cause, seul Microsoft était habilité HDS à l'époque. Cela est vrai et vous pouvez le vérifier. Nous avons donc conduit la comparaison, puis nous avons vérifié qu'Open, notre intégrateur, et Microsoft étaient tous deux disponibles au catalogue de l'UGAP. Cette manière de procéder est légale, et plus encore, elle est recommandée par la DINUM. Le mécanisme de centrales d'achat a été mis en place pour faciliter l'achat public.

M. Philippe Latombe, rapporteur. Nous avons auditionné l'UGAP et la DINUM. La DINUM nous a expliqué que les achats étaient aussi guidés par le besoin de facilité : les administrations achètent des solutions totalement intégrées car elles n'ont pas les

moyens d'intégrer ensemble des blocs différents pour constituer une solution complète. Ce souci de facilité et de rapidité explique-t-il que vous ayez également choisi Microsoft ?

Mme Stéphanie Combes. Je ne parlerais pas de simplicité – je ne dirais pas que notre chantier d'infrastructure est simple. Nous avons choisi une solution qui répondait à notre demande, alors que les acteurs français ne proposaient pas les fonctionnalités dont nous avons besoin. Il aurait fallu construire ces fonctionnalités et cela aurait pris un certain temps. Encore aujourd'hui, OVH n'a pas développé toutes ces fonctionnalités.

M. Philippe Latombe, rapporteur. On aurait pu envisager un consortium : plusieurs acteurs auraient pu travailler ensemble pour apporter ces différentes fonctionnalités.

Mme Stéphanie Combes. Si nous avons publié un marché public, Microsoft y aurait répondu. On ne peut pas choisir de travailler avec un acteur français car on a envie de soutenir son développement industriel, si un autre acteur répond à notre demande, dispose de toutes les certifications requises et mène une activité sur le sol de l'Union européenne. On ne peut pas interdire à des acteurs internationaux de fournir des services. Si demain il existe une nouvelle certification, que Microsoft ne la vérifie pas et qu'OVH la vérifie, alors les choses seront différentes. À l'époque, si nous avons publié un marché, Microsoft l'aurait remporté. Il a même été proposé, pour apaiser le débat, de publier un marché *a posteriori* : c'est à nouveau Microsoft qui aurait été choisi et cela n'aurait fait que jeter de l'huile sur le feu. Il nous manque, collectivement, un *benchmark* du niveau de maturité des solutions de *cloud* françaises endossé par l'État. Il faudrait que ce *benchmark* soit conduit par le ministère de l'industrie ou bien par la DINUM. Cela nous permettrait d'attester collectivement du fait que des acteurs français sont très forts dans tel domaine, et plus faibles dans d'autres. Cela traduirait la réalité du terrain.

M. Philippe Latombe, rapporteur. Je peux le comprendre, et je n'ai pas de critique à y apporter. Je reviens sur ma question du boulet. Les critiques initiales ont marqué le Health Data Hub du sceau d'une réelle difficulté. On évoque aujourd'hui le hub bien davantage pour ces critiques que pour ses succès. Il a été annoncé qu'il faudra migrer vers une solution souveraine et que cela prendra deux ans. Cela donne l'impression qu'une décision prise nous lie pendant des années et que la réversibilité n'est pas si simple que cela à mettre en œuvre. Cette situation jette l'opprobre sur le projet.

Mme Stéphanie Combes. Pour mettre en œuvre la réversibilité, il faut que la cible soit prête. La cible n'est pas prête pour le moment, il faut donc prendre le temps de la construire. Cela n'est pas la faute du hub. De notre côté, la migration nous prendra quelques mois. La CNIL en avait parfaitement conscience. Tout le monde est d'accord sur le fait qu'il est souhaitable que l'hébergement soit souverain – il faut se donner les moyens d'y arriver, et pour cela, il faut donc le temps à l'industrie de développer la solution cible. Cela explique la durée de deux ans.

Nous avons, par ailleurs, en France, un problème avec le numérique. Par faute de compréhension de ses enjeux très techniques et de manque d'ingénieurs dans les administrations, tout le monde donne son avis sur des sujets d'une complexité très importante. Nous ne pouvons pas remettre en cause des choix technologiques : nous avons travaillé à l'infrastructure technologique sécurisée avec le haut fonctionnaire de défense et de sécurité du ministère de la santé et avec l'ANSSI, qui est la plus haute autorité en la matière en France. Nous avons collectivement besoin d'éduquer les Français sur cette composante numérique qui va être de plus en plus présente dans nos vies, afin de conduire des débats moins passionnels et plus objectifs. Cela rejoint la question de la souveraineté que vous me posiez plus tôt : tout

le monde parle de la souveraineté numérique, mais sa définition n'existe pas. Cela est un vrai problème.

Il est vrai que depuis sa création, la communication du hub tournait surtout autour de la plateforme technologique et de ses critiques. Nous commençons cependant maintenant à produire des résultats ; cela est de nature à davantage montrer la finalité du Health Data Hub. L'intérêt de notre action et ses vrais usages seront bientôt visibles : le développement des outils à destination des professionnels et des patients, la recherche médicale, les outils de prédiction des ré-hospitalisations. Mais notre action appartient à la recherche, cela prend donc du temps.

M. Philippe Latombe, rapporteur. S'agissant des coûts, il nous a été expliqué que des différences tarifaires assez fortes peuvent exister entre les différents fournisseurs. Comment se situe Microsoft en la matière ? Était-il significativement moins cher que les autres fournisseurs, de manière à « acheter le marché » ? Ou était-il au contraire plus cher car il se savait être le seul fournisseur à répondre au cahier des charges ?

Ensuite, Microsoft vous a-t-il imposé de négocier toutes les conditions générales d'utilisation et les conditions particulières au sein du contrat ?

Enfin, il nous a été dit que l'envoi des données sur le *cloud* de Microsoft était relativement peu cher, mais que la sortie était extraordinairement chère. Avez-vous pris en compte ce critère et si oui, avez-vous pu le négocier ?

Mme Stéphanie Combes. S'agissant des coûts, la centrale d'achat met en compétition les acteurs en intégrant un critère de coût. Nous avons comparé le coût entre OVH et Microsoft lors de notre première étude de réversibilité. La comparaison a montré qu'il était beaucoup moins cher de construire l'infrastructure Microsoft – car elle est tout intégrée – mais qu'à l'usage, elle est un peu plus chère.

Nous avons évidemment également étudié les conditions de sortie des données. Cet argument était très valide, il y a quelques années, mais il l'est beaucoup moins aujourd'hui. Les acteurs ne sont plus dans la logique de capter les utilisateurs par le coût de sortie des données, car les utilisateurs y sont désormais vigilants. Les acteurs américains vont plutôt développer des services de plus en plus intelligents, ergonomiques, de haut niveau, pour capter l'utilisateur car ils n'ont pas d'équivalents. Ainsi, nous n'utilisons ainsi pas d'outils intégrés de *machine learning*, nous mettons à disposition des outils très standard de *data science* en *open source* comme R et Python.

Nous avons en effet négocié plusieurs avenants aux clauses de contrat ; cela s'est fait au fur et à mesure de nos discussions notamment avec la CNIL. Nous en avons négocié trois au total et nous continuerons d'en négocier si cela est nécessaire.

M. Éric Bothorel. Notre mission sur la souveraineté numérique fonctionne sur Zoom et je constate que la totalité de nos données n'est pas chiffrée.

Je me souviens des débats ayant eu lieu au printemps sur l'application StopCovid, aujourd'hui Tous anti-Covid. Les choix opérés sont systématiquement critiqués. Cela est probablement le cas pour de bonnes raisons, car des subtilités technologiques vont parfois à l'encontre de notre autonomie stratégique.

Il était question tout à l'heure de chiffrement. Qui stocke les clés de chiffrement ? Cela constitue une garantie de la protection des données stockées sur le Health Data Hub.

Le Health Data Hub fait-il l'objet de cyberattaques actuellement ? Si oui, connaît-on la nature et l'origine de ces tentatives ? Je souhaiterais savoir comment le Health Data Hub s'organise pour faire face à la sphère cybercriminelle qui pourrait être tentée d'accéder à ses données et quelles sont les mesures déployées pour l'en empêcher.

Mme Stéphanie Combes. Les données sont toutes chiffrées, à la fois lorsqu'elles sont stockées et lorsqu'elles se déplacent. Les clés sont stockées dans des modules numériques appelés *hardware security modules* (HSM), qui constituent le plus haut niveau de sécurité internationale. Nous créons des clés dites maîtresses à l'extérieur de la plateforme dans un HSM maîtrisé par le ministère. Elles sont ensuite envoyées à l'intérieur de HSM de Microsoft, car la clé doit se trouver à l'intérieur de la plateforme pour y chiffrer et déchiffrer les données. Le HSM ne nécessite aucune intervention d'un administrateur de Microsoft ni même du hub : cela constitue une norme de sécurité internationale. Le fait que nous possédions les clés maîtresses nous permet, si nous le souhaitons, de révoquer les clés en cas d'accident majeur et donc de supprimer les données (qui ne sont que des copies, puisqu'il s'agit de données de recherche). Les avenants au contrat prévoient bien que Microsoft ne cherchera jamais à contourner ce chiffrement ou à récupérer des clés de chiffrement pour les confier à un tiers.

Le Health Data Hub n'a pas aujourd'hui fait l'objet de cyberattaques. Cela constitue d'ailleurs l'un des indicateurs de notre stratégie pluriannuelle de transparence envers notre assemblée générale. Si nous sommes informés d'une cyberattaque, les analystes de mon équipe RSSI examinent si nous sommes concernés et instruisent la cyberattaque. Nous organisons par ailleurs des « attaques » de manière volontaire pour tester la plateforme : nous en avons déjà mené deux et un troisième audit est actuellement en cours avec l'ANSSI.

M. Philippe Latombe, rapporteur. Vous avez expliqué tout à l'heure que le hub n'était pas exclusif d'initiatives locales. Ainsi l'AP-HP disposait déjà d'une plateforme technologique et d'un entrepôt de données. Pourquoi ne pas avoir utilisé l'architecture du dispositif de l'AP-HP pour le faire grossir ? Pourquoi avoir fait le choix d'une solution technologique différente ?

Mme Stéphanie Combes. Nous avons évidemment envisagé cette option et nous avons eu l'occasion d'échanger à plusieurs reprises avec le DSI de l'AP-HP à ce sujet. L'AP-HP est l'acteur français le plus avancé en matière d'entrepôt de données hospitalières. Leur travail est formidable. Pour le moment, leur solution n'est pas conforme au référentiel de sécurité du SNDS – or cela constituait une obligation légale pour le hub. Par ailleurs, leur infrastructure sur site ne leur permet pas de passer à l'échelle pour tout type de projet, comme nous pouvons le faire avec une infrastructure *cloud*. Nous menons actuellement des projets en partenariat, qui montrent bien la complémentarité de nos solutions technologiques : l'AP-HP commence le projet avec d'autres partenaires, et nous arrivons en bout de chaîne par le traitement algorithmique en apportant la capacité de calcul nécessaire.

M. Philippe Latombe, rapporteur. Y'a-t-il eu de l'entrisme de la part de Microsoft auprès du Health Data Hub ? Existe-t-il des liens, des connexions – y compris amicales – entre des personnes du Health Data Hub et Microsoft ? Ces propos sont revenus fortement sur les réseaux sociaux. L'on parle beaucoup de mobilités de personnels entre les prestataires et l'administration. Est-ce l'une de vos préoccupations ? Avez-vous pu vérifier ce type de mouvements qui pourraient poser, à terme, des questions d'éthique et de déontologie ?

Mme Stéphanie Combes. Nous n'avons aucun lien, à l'origine, avec Microsoft. Je suis ingénieure utilisatrice de données et j'ai eu l'occasion d'utiliser entre autres les

solutions de Microsoft Azure et d'Amazon Web Services (AWS) par le passé. Je ne connaissais absolument pas les personnes de Microsoft avec lesquelles je travaille aujourd'hui. L'équipe technique aujourd'hui présente au hub n'existait d'ailleurs pas lors du lancement du projet. Je suis administratrice de l'INSEE et s'agissant du choix des partenaires technologiques, j'ai d'abord pensé à TeraLab et au centre d'accès sécurisé aux données (CASD) qui sont les acteurs de la statistique publique. Nous nous sommes tournés vers les solutions américaines très tardivement. Nous avons élargi nos recherches lorsque nous nous sommes rendus compte que les acteurs n'étaient pas en mesure de répondre à notre cahier des charges. Nous avons un vrai problème si nous n'étions pas capables de mettre en place une plateforme de *data science* en santé avec des outils sur étagère. En fin de course seulement donc, nous avons commencé à interroger Microsoft, AWS et Google Platform. Je vous garantis que le choix de Microsoft n'a pas été un choix d'influence. Je suis ingénieur. Nous avons été pragmatiques et nous avons souhaité obtenir une solution rapide dans un contexte dans lequel la France n'est pas en avance. La législation française et européenne est très forte : cela est positif, mais cela constitue également une complexité – réaliser des projets sur des données de santé en France est complexe. Le choix de Microsoft n'a pas été un choix d'influence mais, la crise sanitaire l'a montré, l'écosystème français de la recherche en santé est confronté à plusieurs autres problèmes.

M. Philippe Latombe, rapporteur. Comment avez-vous analysé l'arrêt *Schrems II*, lorsqu'il a été rendu ? Maintenant que les conséquences de cet arrêt ont été formulées et que l'on vous a demandé de migrer vers un *cloud* souverain, où en êtes-vous de ce travail ?

Mme Stéphanie Combes. Il est intéressant que vous reliez ces deux questions car, à mes yeux, il n'existe pas tellement de rapport entre elles.

L'arrêt *Schrems II* porte sur le transfert de données personnelles. Pendant toute la première moitié de l'année 2020, notre effort de pédagogie lors des débats sur le Health Data Hub a consisté à expliquer qu'il n'y avait pas de transfert de données de santé. D'aucuns affirmaient alors que les données étaient hébergées aux Pays-Bas puis traitées aux États-Unis – cela est un non-sens du point de vue technologique. Nous avons mis un certain temps à convaincre et à montrer, avenant à l'appui, que les administrateurs de Microsoft n'accédaient jamais aux données et que les services d'utilisation des données étaient également régionalisés. Nous avons réussi à documenter cela et à en convaincre la CNIL. Le Conseil d'État en a ensuite attesté en affirmant qu'il n'y a pas de transfert de données personnelles. En revanche, *Schrems II* est un vrai problème pour les acteurs qui procèdent à des transferts de données, en raison des clauses contractuelles types et des mesures techniques et organisationnelles à mettre en place pour améliorer la sécurité des transferts de leurs données personnelles.

Les engagements pris en matière de souveraineté par M. Cédric O ou par le ministère de la santé sont fondés sur le risque extraterritorial, lequel n'a pas été reconnu par le Conseil d'État. La compréhension des lois extraterritoriales est donc problématique. Nous menons des travaux juridiques pour documenter l'impact des lois extraterritoriales ; nous identifions, avec la DINUM, nos services essentiels ; et enfin, nous recherchons des financements. Si nous créons demain cette plateforme, nous aurons besoin de financements. Nous avons identifié des financements dans le plan de relance – l'accélération de la stratégie du *cloud* souverain est portée par la direction générale des entreprises (DGE). Par ailleurs, nous participons aux projets européens et nous suivons toutes les activités de l'ANSSI autour de la cybersécurité dans le cadre du plan de relance. Je suis donc assez confiante dans notre capacité interministérielle à débloquer les 20 millions d'euros nécessaires au développement

d'une solution souveraine. Évidemment, cette solution souveraine ne servirait alors pas qu'au Health Data Hub – l'enjeu d'une plateforme souveraine va concerner de nombreux autres *data hubs* dans beaucoup d'autres secteurs.

M. Philippe Latombe, rapporteur. C'est justement pour cela qu'il était important de vous recevoir – vous avez été précurseur en la matière s'agissant des données de santé et vous avez par conséquent essuyé toutes les critiques. Ces solutions de *data science* vont intéresser demain beaucoup d'autres domaines. Ne faudrait-il pas, de votre côté, accélérer autant que possible le développement du projet ? Disposez-vous du soutien financier et technique de l'État pour mener ce projet le plus vite possible ?

Mme Stéphanie Combes. Nous tenons depuis longtemps des discussions avec la DINUM, l'ANSSI, le Premier ministre, le ministère de la santé. Nous constatons un alignement avec la stratégie *cloud* de l'État et cela va permettre de concrétiser ces ambitions. En tant que chef de projet, j'identifie les fonds, car sans fonds, nous ne pourrions rien faire de concret. Je suis aujourd'hui confiante en notre capacité à faire. En revanche, le délai est extrêmement ambitieux sur le plan industriel. Mais c'est en se fixant des objectifs ambitieux que nous réussirons à franchir des paliers.

M. Philippe Latombe, rapporteur. Comme vous l'avez dit, la solution souveraine intéressera de nombreux autres ministères et de nombreux autres sujets. Il serait sensé de construire une solution qui sera utilisable par l'ensemble des acteurs.

Que retenez-vous, depuis la création du Health Data Hub jusqu'à maintenant, des difficultés auxquelles vous avez été confrontée ? Comment voyez-vous les choses dans les mois et les années à venir ? Allez-vous réussir à faire la preuve d'une réversibilité qui fera oublier le démarrage du Health Data Hub ?

Mme Stéphanie Combes. Mener des projets et produire des résultats de recherche concrets va donner du sens à notre démarche. La critique adressée au Health Data Hub s'agissant de Microsoft intéresse l'écosystème du numérique français et les acteurs nourrissant des craintes sur la protection des données – ils peuvent être rassurés par les avis apportés par les autorités prescriptrices. Les gens comprendront pourquoi nous avons mis en place le Health Data Hub quand nous pourrons mettre en avant des résultats de recherche convaincants et concrets.

Les enjeux numériques, aujourd'hui extrêmement passionnels et mal compris, doivent être mis au second plan par rapport aux finalités que les plateformes poursuivent. Il en va exactement de même pour StopCovid : les gens se sont d'abord interrogés sur la manière dont cette application traiterait leurs données personnelles. Mais il faut surtout et avant tout se demander à quoi sert cette application. Nous devons, nous, tous les acteurs du numérique, travailler à rendre l'usage final très clair.

Nous continuerons donc à travailler sur le chantier de l'infrastructure et à produire des résultats. Un écosystème très enthousiaste nous suit, et regrette que notre développement soit trop lent ; nous recevons énormément de réponses à nos appels à projets ; nous enregistrons une très forte participation à nos événements fédérateurs comme le colloque, la *data challenge*, la *winter school*.

La vraie difficulté du Health Data Hub pour les prochaines années est ailleurs. Nous sommes face à un problème culturel de réflexe en ce qui concerne les données de santé. Les données de santé n'appartiennent à personne et personne ne le sait. Certains écosystèmes

de santé et établissements pensent que les données leur appartiennent parce qu'ils ont fait l'effort de collecte et qu'ils ont soigné les patients. Les acteurs ne savent pas quand ils ont le droit ou non de traiter les données. Cela est extrêmement complexe et constitue un vrai frein à l'innovation en santé. Ainsi, certaines personnes ne respectent pas le cadre applicable et certaines autres personnes sont frileuses, car elles ne maîtrisent pas le cadre. Si nous ne cassons pas cette spirale, le Health Data Hub sera un prestataire d'outils de *data science* pour quelques projets et nous passerons à côté de l'ambition du dispositif.

Nous avons donc besoin d'un écosystème institutionnel autour de la donnée de santé, avec une politique nationale et des financements dédiés. Je citerai l'exemple de la *UK Bio Bank*, une base de donnée anglaise reconnue au niveau international et dont les données sont utilisables par tous. Les contributeurs à cette base n'ont aucun réflexe de propriétaire. Pourquoi ? Le financement de cette base est pérenne et ne dépend pas des publications des acteurs qui sont à l'origine de la base. Ainsi, n'émergent pas chez les acteurs des comportements pervers qui impliquent de réserver les données pour eux-mêmes car ils ont besoin des financements associés aux données. Le Plan national pour la science ouverte est porté au plus haut niveau par le ministère de la recherche, mais il doit être associé à des dispositifs incitatifs : il n'est pas concevable, pour les chercheurs, d'ouvrir la science mais d'encore devoir lutter pour trouver des financements. Le hub peut porter mais ne peut pas être maître d'œuvre de cette politique de la science ouverte et de la santé – cela n'est pas notre rôle. À titre d'exemple, le Ségur du numérique ne prévoit pas d'enveloppe pour les données de santé de recherche – et ce, même à l'issue d'une crise sanitaire.

M. Philippe Latombe, rapporteur. Cela veut-il dire que le frein provient aussi de nos concitoyens, qui voient les données de santé comme une donnée sacrée, qu'on ne peut pas utiliser ni vendre ? Doit-on évoluer vers un modèle semblable à la *UK Bio Bank* ou bien au modèle israélien ? La nature même de la donnée de santé ne constitue-t-elle pas un frein ?

Mme Stéphanie Combes. C'est une question complexe à laquelle je ne saurais répondre de manière définitive, mais je n'en suis pas certaine. Nous avons mené plusieurs études de perception du partage des données de santé par les citoyens. Ils ne sont pas opposés à l'idée de partager leurs données de santé à des fins de recherche. Ils formulent seulement une crainte particulière quant à l'accès des assureurs aux données, et expriment une réserve supplémentaire au sujet de l'industrie pharmaceutique. La vente des données de santé est interdite par la loi française. La tarification de la mise à disposition des données, en revanche, est possible en France. Cela permettrait de tarifier la mise à disposition de données à un acteur privé afin d'investir dans la base de données. Cela est plutôt sain car investir dans une base de données est très coûteux.

À mes yeux, le sujet porte donc davantage sur l'écosystème : les acteurs qui collectent les données et se sentent insuffisamment valorisés pour ce travail. À ce titre, la politique pour la science ouverte vise à valoriser la production de données au même titre qu'une publication scientifique. Je viens du secteur de la statistique publique et je considère que la production de données est un métier essentiel. Dans l'écosystème de la recherche, au contraire, ce métier est connoté tout à fait négativement. Nous avons donc besoin de valoriser le *data sharing*, de financements pérennes pour les bases et que les personnes qui portent des cohortes ne soient pas inquiètes des financements pour les grandes cohortes essentielles.

M. Philippe Latombe, rapporteur. Comment les choses se passent-elles avec les bases locales, s'agissant notamment de la rémunération et du partage des coûts ?

Mme Stéphanie Combes. Nous avons mis en place une offre de services à destination des responsables de données, c'est-à-dire des acteurs à l'origine de la réunion des données. Nous avons travaillé avec des panels d'acteurs sur une convention-type qui encadre le partage. Deux articles ont nécessité un travail important.

Il s'agit tout d'abord d'un article sur la valorisation scientifique. Selon les recommandations internationales, une personne qui a contribué en apportant des données ne peut pas être signataire d'un article scientifique. Cela constitue un vrai problème, car être signataire permet de gagner des points qui donnent lieu à une rémunération. À ce sujet, nous avons réussi à trouver un compromis dont nous sommes assez satisfaits.

Le second article problématique portait sur la tarification : la question est de définir l'assiette possible de la tarification. Les bases publiques à finalité administrative ont, pour la plupart, toutes déjà été financées par des fonds publics. En revanche, les bases de recherche supposent la création d'une infrastructure pérenne, avec des équipes et des investissements technologiques importants, qui sont très coûteux. Les réflexions sont toujours en cours à ce sujet. La Commission européenne réfléchit en ce moment même, notamment avec la mise en place du *Data Governance Act* et la réflexion autour de l'espace commun des données de santé. Le *Data Governance Act* prévoit un article sur la tarification et les mêmes questions vont émerger sur le point de savoir ce que doit recouvrir ce tarif.

M. Philippe Latombe, rapporteur. Nous avons souhaité cette audition car vous êtes précurseur des enjeux de souveraineté en matière de données de santé avec la création du Health Data Hub. L'utilisation d'un écosystème afin de créer un support pour toutes les *data science* est une idée. Je vous remercie pour ces échanges.

Mme Stéphanie Combes. N'hésitez pas à me dire si vous souhaitez que je documente certains points auxquels j'ai fait référence pendant mon intervention.

Pour conclure, je pense qu'il est extrêmement important de ne pas réduire la question de la souveraineté à l'hébergement, mais d'y inclure également la question de l'usage. De cette manière, nous ne nous retrouverons pas dans cinq ans à tenir les mêmes échanges, cette fois sur l'intelligence artificielle. Les acteurs américains développent des services « managés » d'intelligence artificielle de façon très rapide. Cela est un sujet majeur : nous avons tous les atouts en France pour développer ces usages, mais nous risquons de nous retrouver dans quelques années dans la même situation qu'actuellement pour le *cloud*.

M. Philippe Latombe, rapporteur. Cela est tout à fait intégré. Je vous remercie.

La séance est levée à 11 heures 15.



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 9 heures 30

Présents. - MM. Éric Bothorel, Thomas Gassilloud, Mme Amélia Lakrafi, MM. Philippe Latombe, Jean-Michel Mis, Pierre-Alain Raphan, Jean-Luc Warsmann