

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 M. Claude Gissot, inspecteur général de l'INSEE, directeur de la stratégie, des études et des statistiques (DSES), et de Mme Stéphanie Naux, directrice de mission au cabinet du directeur de la stratégie, des études et des statistiques, de la caisse nationale d'assurance maladie (CNAM) 2

Mardi

9 mars 2021

Séance de 11 heures

Compte rendu n° 37

SESSION ORDINAIRE DE 2020-2021

**Présidence de
M. Philippe Latombe,
*rapporteur***



Audition, ouverte à la presse, de M. Claude Gissot, inspecteur général de l'INSEE, directeur de la stratégie, des études et des statistiques (DSES), et de Mme Stéphanie Naux, directrice de mission au cabinet du directeur de la stratégie, des études et des statistiques, de la caisse nationale d'assurance maladie (CNAM)

La séance est ouverte à 11 heures 05.

Présidence de M. Philippe Latombe, rapporteur.

M. Philippe Latombe, rapporteur. Nous poursuivons notre cycle d'auditions consacrées au numérique en santé, en auditionnant aujourd'hui M. Claude Gissot, directeur de la stratégie, des études et des statistiques (DSES), et de Mme Stéphanie Naux, directrice de mission au cabinet du directeur de la stratégie, des études et des statistiques de la caisse nationale d'assurance maladie (CNAM).

Au cours des dernières semaines, nous avons déjà entendu à ce sujet la délégation ministérielle du numérique en santé, l'Assistance publique – Hôpitaux de Paris (AP-HP), ainsi que les représentants du Health Data Hub.

La CNAM est un établissement public national à caractère administratif (EPA), qui constitue la tête de réseau opérationnelle du régime d'assurance maladie obligatoire en France. Elle est placée sous la double tutelle du ministère des Solidarités et de la santé et du ministère de l'Économie, des finances et de la relance. Elle est en charge du système national des données de santé (SNDS), qui rassemble un certain nombre de bases de données de santé à des fins de recherche. La CNAM est donc particulièrement concernée par les enjeux de protection des données de santé, et de soutien à la recherche et à l'innovation. À ce sujet, nous aurons l'occasion de revenir sur le refus du conseil de la CNAM de transférer une copie des données du SNDS au Health Data Hub, dès lors qu'il utilisait la solution de *cloud* Azure de Microsoft.

Pourriez-vous compléter cette présentation en expliquant notamment le rôle de la CNAM en matière de numérisation de notre système de santé ? Comme cela nous a été rappelé la semaine dernière, de nombreux acteurs interviennent dans ce domaine, qu'il s'agisse des agences régionales de santé (ARS) ou des centres hospitaliers, la stratégie « Ma santé 2022 » étant pilotée par la délégation ministérielle du numérique en santé. Comment la CNAM participe-t-elle à cette dynamique ? La notion de souveraineté numérique est-elle prise en compte dans vos actions ?

En second lieu, pourriez-vous présenter la base de données que constitue le SNDS, en précisant ses différentes composantes et en détaillant son fonctionnement ? Quels choix techniques ont été effectués et quels types de données sont rassemblées en son sein ? Cet échange nous permettra d'évoquer ensuite les raisons du refus récent du conseil de la CNAM de transférer auprès du Health Data Hub une copie des données du SNDS, et les évolutions envisageables à terme sur cette question.

Enfin, je souhaiterais évoquer avec vous l'enjeu de la protection des données de santé et des systèmes d'information de l'assurance maladie. L'actualité récente a été marquée par des attaques cyber, très médiatisées, contre des établissements de santé, avec une fuite de données record qui concernerait environ 500 000 patients. Comment appréhendez-vous cet enjeu ? Échangez-vous avec l'agence nationale de la sécurité des systèmes d'information

(ANSSI) sur cette question ? La sécurité constitue en effet une condition indispensable pour rassurer les citoyens et leur montrer l'apport du numérique en santé. À ce sujet, nous sommes naturellement intéressés par votre regard sur les meilleurs voies et moyens permettant d'inclure les citoyens dans cette transformation.

M. Claude Gissot, directeur de la stratégie, des études et des statistiques (DSES) de la caisse nationale d'assurance maladie. La CNAM est un EPA, mais surtout le principal assureur obligatoire en santé. Il couvre maintenant presque la totalité de la population résidant en France, ou ayant du moins ouvert des droits en France auprès de l'assurance maladie. Son activité historique de paiement des prestations de santé aux assurés et de remboursement des soins aux professionnels ou aux assurés s'est élargie fortement ces dernières années à l'accompagnement des patients pour la gestion de leurs droits ou de leurs risques en matière de santé. Nous avons également développé une activité très forte au titre des relations conventionnelles avec les professionnels, notamment libéraux, puisque leur activité est régie par une convention négociée entre les représentations professionnelles de libéraux et l'assurance maladie, soit principalement la CNAM. Bien entendu, nous travaillons aussi à l'amélioration de la qualité des soins et surtout de l'efficacité des processus de soin, puisque l'objectif général de la CNAM est de garantir que le système de solidarité en santé existant aujourd'hui en France soit pérenne et résiste aux évolutions de la santé, dont nous pourrions si vous le voulez préciser les facteurs très puissants existant aujourd'hui.

Dans son activité, la CNAM a développé beaucoup de services numériques auprès des professionnels de santé ou des assurés, principalement avec le compte AMELI de l'assurance maladie que les assurés peuvent ouvrir, mais qui a été complété, il y a quelques années, par le dossier médical partagé (DMP), que la CNAM a repris pour assurer son développement. Presque 10 millions de DMP ont ainsi été ouverts aujourd'hui et complétés par l'assurance maladie elle-même, ce qui a permis de lever les difficultés des versions précédentes du DMP. La CNAM a également développé des services auprès des professionnels de santé, en lien avec la digitalisation de la pratique professionnelle, en termes de remboursement, avec les feuilles de soin électroniques, qui existent depuis longtemps, mais aussi des téléservices pour les prescriptions d'arrêt de travail, les déclarations de médecin traitant, etc. Tous ces outils sont développés depuis plusieurs années par la CNAM dans un objectif de numérisation de l'activité de gestion de la santé, porteuse d'efficacité pour les professionnels de santé et les assurés, mais aussi pour la CNAM.

Par ailleurs, nous collectons depuis longtemps des données « médico-administratives » (correspondant en réalité aux remboursements des soins de ville et hospitaliers) dans une base nationale décisionnelle qui s'appelait initialement, et s'appelle toujours, le Système national d'information interrégime de l'assurance maladie (SNIIRAM), qui constitue probablement la base principale du SNDS actuellement. Cette base consolide ainsi les remboursements des soins de ville et hospitaliers de tous les assurés, quel que soit donc leur régime, et sur un historique long, puisque le SNIIRAM a été créé à la fin des années 1990, c'est-à-dire en 2000 sur une enveloppe de financement de 1998-1999. Le temps que la quantité d'information considérable ainsi réunie soit organisée, le SNIIRAM est opérationnel de manière totalement efficace et exhaustive depuis les années 2005-2006. Les premières exploitations réellement fortes qui en ont été faites datent de 2007. Dès le départ, ce système d'information a été ouvert à l'ensemble des acteurs de la santé, d'abord aux acteurs publics, et, évidemment, à tous les acteurs de la recherche, même s'il a fallu un certain temps pour que les chercheurs s'intéressent à ces données administratives, qui ne comportent pas de données médicales au sens de diagnostics ou de résultats en termes de santé, ce qui peut parfois en détourner les chercheurs cliniciens. Néanmoins, nous avons pu démontrer, et l'assurance maladie

notamment en les utilisant, que ces données étaient très utiles pour analyser le système de santé et le système de soins, afin d'en déduire des informations sur la qualité des soins, l'efficacité, etc. À partir de 2019, le SNIIRAM a été élargi pour être transformé en SNDS en incluant d'autres sources de données, c'est-à-dire principalement les causes de décès et prochainement les données des maisons départementales des personnes handicapées.

La CNAM a donc évidemment été un membre historique de l'Institut des données de santé (IDS), créé en 2002 pour regrouper les acteurs de la santé autour de la recherche d'une meilleure utilisation des données de la santé et de leur ouverture à l'ensemble des acteurs de la santé. Les seules bases de données médico-administratives existantes à l'époque étaient le SNIIRAM et le programme de médicalisation des systèmes d'information (PMSI), qui relève tous les séjours hospitaliers réalisés dans les établissements de santé publics comme privés, et dans tous les secteurs. La CNAM a ainsi poursuivi son rôle de recueil et de mise à disposition des données à l'ensemble des acteurs de santé dans le cadre de l'IDS, devenu Institut national des données de santé (INDS) en 2016, et maintenant avec le Health Data Hub (HDH) depuis la loi de 2019. La CNAM conserve aujourd'hui ce rôle à travers son portail. Elle est un membre du groupement d'intérêt public (GIP) HDH. À ce titre, elle participe aux différentes instances de régulation du pilotage du HDH, comme tous les acteurs de santé, qui sont tous présents dans l'assemblée générale du HDH. Le vice-président du HDH notamment est un représentant de France Assos Santé, et il y représente donc les patients.

Historiquement, la CNAM a mis en place le SNIIRAM, qui constitue encore aujourd'hui la base de données principale du système de soins français. Nous sommes également le partenaire principal du HDH, aujourd'hui, pour sa constitution, sa montée en charge et sa montée en compétences, notamment s'agissant de l'utilisation de ses données dans sa « base centrale » qui rassemble le SNIIRAM, la base des causes de décès et le PMSI. En effet, les projets d'études déposés aujourd'hui sont principalement réalisés sur le portail de la CNAM, et principalement accompagnés par les équipes de la CNAM. Nous travaillons de manière très rapprochée avec le HDH pour partager notre expérience sur les données, sur la réalisation des projets, les appariements, etc. De manière générale, la CNAM apporte donc son soutien, et toute son expertise, au projet du Health Data Hub, car il doit permettre une plus grande utilisation des données, un enrichissement des données médico-administratives par des données plus médicalisées, incluant notamment des données cliniques, qui pourront décupler les possibilités d'analyse et d'étude afin de contribuer, au final, à l'amélioration des prises en charge des patients.

Mme Stéphanie Naux, directrice de mission au cabinet du directeur de la stratégie, des études et des statistiques (DSES) de la caisse nationale d'assurance maladie. Au moment de sa création en 2016, le système national des données de santé (SNDS) s'inscrivait dans la continuité du SNIIRAM, qu'il élargissait.

Créé au début des années 2000, le SNIIRAM visait à constituer une base d'information exhaustive sur le recours aux soins, avec un historique long (jusqu'à vingt ans), pour suivre non seulement les dépenses, mais aussi les parcours de soin, les effets à long terme des prises en charge et des pathologies traitées, ce qui a constitué son premier intérêt pour la recherche. Les données de cette base sont principalement issues de systèmes de gestion initialement utilisés à d'autres finalités : en particulier le remboursement des soins, la rémunération des producteurs de soins, les référentiels associés. Le SNIIRAM comprend les données de ville produites par l'assurance maladie (c'est-à-dire l'ensemble des informations issues des feuilles de soins), auxquelles sont associées les données des séjours hospitaliers issues du PMSI. Toutes ces données sont « pseudonymisées », ce qui signifie que toutes les

données directement identifiantes sont supprimées : aucun nom, aucun prénom, aucune adresse précise ne sont conservés, et le numéro de sécurité sociale ou NIR (numéro d'inscription au répertoire de l'INSEE) est remplacé par un pseudonyme irréversible, qui est constitué par un hachage, mais qui est unique pour une même personne, ce qui permet peu à peu d'articuler les données entre elles. La base est ainsi construite par l'attribution dans le temps des mêmes données aux mêmes personnes.

En 2016, la première version du SNDS a donc ajouté au périmètre initial du SNIIRAM les données des maisons départementales des personnes handicapées (MDPH), qui sont gérées et produites par la Caisse nationale de solidarité pour l'autonomie (CNSA), et les données des causes de décès incluses à la base du centre d'épidémiologie sur les causes médicales de décès (CépiDC) produite par l'Inserm. Ces données ont été ajoutées à la base du SNIIRAM avec un chaînage direct et une association physique des données sous le même pseudonyme, pour être réunies au sein du portail de la CNAM.

Une étape supplémentaire a été franchie en 2019 à l'occasion de la loi relative à l'organisation et à la transformation du système de santé (OTSS), qui a complété largement le cadre du SNDS en ouvrant, à côté de cette base, désormais nommée « principale » ou « historique », un catalogue de bases de données ayant vocation à être appariées ou appariables avec les données de la base principale, et qui sont pour beaucoup des données destinées aux professionnels de santé : données de résultats et d'imagerie, mais aussi d'autres sources.

La loi de 2019 a également créé le Health Data Hub, ou Plateforme des données de santé. Il s'agissait d'abord d'élargir les missions auparavant confiées à l'INDS :

– d'une part, des missions liées à la gestion du guichet unique de dépôt des dossiers de demande d'accès aux données, donc à l'initiation du circuit d'accès aux données par le comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (CESREES) et la Commission nationale de l'informatique et des libertés (CNIL) ;

– d'autre part, une mission de constitution de lieux d'échanges entre les acteurs, producteurs et utilisateurs des données. Depuis 2019, la plateforme de données de santé Health Data Hub a, comme la CNAM, une mission de mise à disposition des données aux acteurs souhaitant les utiliser, donc la constitution d'une plateforme pour la mise à disposition de ces données. HDH est également en charge de la constitution du catalogue des données et doit promouvoir le partage et l'usage de ces données, ainsi que la recherche et l'innovation en santé et l'information des usagers.

Deux voies d'accès aux données du SNDS existent.

La première voie est celle des « accès permanents », qui existent depuis 2016 pour les institutions remplissant une mission de service public (directions centrales des ministères, agences de santé, instituts de recherche comme l'Inserm ou l'INDS). Celles-ci disposent d'un accès direct aux données sur le portail de la CNAM, selon un périmètre défini par décret pour chaque institution, en fonction de ses missions, de zones géographiques, d'historiques de données et de son accès plus ou moins détaillé à certaines bases. Cet accès permanent est ouvert depuis la première version du SNDS, sur le portail de la CNAM. Chaque institution disposant d'un accès permanent peut également habiliter des acteurs individuels à travailler sur ces données.

La deuxième voie est celle des « accès sur projet », qui existent pour toutes les institutions ou entreprises qui, soit ne disposent pas d'un accès permanent et souhaitent développer un projet spécifique, soit disposent d'un accès permanent, mais souhaitent en élargir le périmètre pour un projet ou une étude donnée. Les porteurs de projet doivent alors déposer les demandes d'autorisation selon un protocole détaillant leur étude auprès du guichet unique ou du secrétariat du HDH. Ces protocoles sont examinés par le CESREES, qui rend un avis avant transmission pour autorisation ou non par la CNIL. Celle-ci peut aussi autoriser une mise en œuvre sur des « bulles sécurisées », prévues spécifiquement par la CNIL, pour accueillir certains projets.

M. Claude Gissot. Pour la CNAM, en tant que producteur et utilisateur de données jouant un rôle de régulateur du système de santé, la souveraineté numérique permet surtout une exploitation la plus efficace possible des données pour alimenter la recherche, la santé publique et l'innovation, sans que les innovations en santé en France dépendent de pays qui construiraient des bases de données à même de porter des projets scientifiques et technologiques, mais aussi de les assurer. Il n'est pas certain en effet que le cadre organisationnel propre à la France permettrait de produire les mêmes résultats en utilisant des données produites dans d'autres pays et d'autres contextes.

S'agissant du partage des données de santé et des freins éventuels existant à cet égard, il faut d'abord souligner que le SNIIRAM et le SNDS ont déjà été très utilisés. Ils permettent la réalisation de 150 projets d'études par an par des opérateurs publics ou privés, et le *benchmark* présenté par le Health Data Hub, il y a quelques semaines, lors de son assemblée générale montre que le nombre de projets ainsi conduits n'est pas tellement supérieur, dans d'autres pays, même si les contenus des bases de données exploitées ne sont pas nécessairement comparables. Certaines des bases de données étrangères sont moins larges en nombre, mais contiennent davantage de données médicalisées, ce qui leur donne des potentiels différents. L'avantage du SNDS est toutefois d'être exhaustif quant à la population, ce qui lui confère une puissance statistique sans égal pour les études qu'il est censé rendre possibles.

Ces projets peuvent être menés par des opérateurs publics ou privés, mais ils ne peuvent évidemment pas utiliser ces données pour des finalités interdites par la loi, comme la modification des contrats d'assurance individuels et la promotion des produits de santé auprès des professionnels. Les projets qu'ils présentent doivent donc pouvoir exclure ces finalités.

Par ailleurs, le SNIIRAM-SNDS contient des données administratives, recueillies à des fins premières de gestion, et non d'étude. L'exhaustivité et la précision de ces données (qui incluent le code CIP des médicaments et le codage précis des actes réalisés) permettent d'analyser les systèmes de recours aux soins avec une grande efficacité. Néanmoins, ces données ne sont pas construites initialement pour des finalités d'étude, et leur réutilisation dans ce cadre requiert une expérience et une formation importantes. C'est pourquoi la CNAM accompagne, depuis dix ans maintenant, des centaines de chercheurs afin qu'ils puissent réaliser leurs études. Elle a récemment publié un article dans la Revue française de santé publique dressant le bilan de l'utilisation de ces bases de données, et montrant notamment qu'un nombre extrêmement important d'études recourant à ces données ont été référencées dans des revues à comité de lecture, ce qui en montre le succès.

Le SNIIRAM-SNDS contient toutefois assez peu de données médicales : les résultats des tests, les stades des cancers, etc. n'y sont pas fournis. Son potentiel vient donc essentiellement des registres et des cohortes qui sont organisés par ailleurs par de nombreuses

unités de recherche, et qui sont malheureusement encore assez insuffisamment utilisés, notamment en appariement avec le SNIIRAM, alors que ces données sont extrêmement complémentaires. En effet, les entrées des cohortes et des registres passent par des pathologies très précises, auxquelles le SNIIRAM-SNDS pourra associer très précisément l'ensemble des comorbidités auxquelles les patients sont soumis également, en permettant ainsi de comprendre leur parcours de soins au-delà de la seule pathologie concernée par la cohorte ou le registre.

M. Philippe Latombe, rapporteur. La semaine dernière, nous avons interrogé l'Assistance publique- Hôpitaux de Paris (AP-HP). Elle s'est dite très intéressée par les données du SNIIRAM-SNDS, mais a confié avoir beaucoup de mal à y accéder, du fait notamment de la complexité de la procédure de dépôt des projets auprès de la CNIL. Tout le monde est donc aujourd'hui d'accord avec vous pour dire qu'il s'agit de données très intéressantes. Avec l'appariement des cohortes, elles permettraient en effet de disposer d'une vision très large de la situation. Toutefois, il semble difficile d'y accéder. Existe-t-il un moyen de fluidifier ou de simplifier l'accès aux données du SNDS ?

M. Claude Gissot. Nous partageons naturellement cette problématique du fait de notre expérience de l'accès aux données depuis de nombreuses années, et notamment au sein du HDH.

Le parcours d'accès à ces données a été défini par la loi. Il encadre la manière dont les dossiers doivent être déposés auprès du CESREES, qui les examine, au regard de leur intérêt public, de leur méthodologie et de la pertinence de la recherche envisagée, etc. Il émet un avis, qui est transmis à la CNIL, laquelle fournit sa réponse.

Plusieurs questions se posent toutefois à cet égard. En premier lieu, les données médico-administratives du SNIIRAM-SNDS sont complexes, et les chercheurs ne les connaissent pas nécessairement. Pour déterminer quelle extraction de données issues du SNDS est requise par le projet déposé par un chercheur, de nombreux allers-retours sont souvent nécessaires entre le chercheur et les responsables de données travaillant à la CNAM. Soit, par exemple, un chercheur souhaitant enquêter sur la cohorte des diabétiques en 2018 : il existe en réalité plusieurs types de diabétiques, répondant dans le SNDS à différentes définitions. Ainsi, le chercheur devra notamment préciser si les patients qui l'intéressent sont fortement traités (donc sous insuline), ou s'ils reçoivent trois, ou six, traitements antidiabétiques par trimestre, ce qui renvoie chaque fois à des catégories différentes. Le chercheur devra ainsi définir la notion de « diabétique » qui l'intéresse à travers des données administratives, et non médicales : il ne suffira pas de demander l'ensemble des personnes présentant tel ou tel niveau de glycémie lors de leurs tests. Il n'est donc pas simple de passer de l'idée d'un projet à la caractérisation des données à exploiter. Ce point a toujours été sous-estimé par les chercheurs. Bien sûr, nous travaillons avec le HDH pour délivrer une formation à tout chercheur qui dépose un projet, pour lui expliquer comment sont constituées les données, ce qu'il est possible d'en tirer et la manière de les traiter. Elles font également l'objet d'une documentation publique et ouverte à tous, de plus en plus volumineuse. La nécessité d'obtenir l'autorisation de la CNIL ne constitue donc pas nécessairement le principal problème dans la complexité actuelle du processus d'accès aux données.

Par ailleurs, les projets déposés sont nombreux, ce qui constitue une charge importante pour le CESREES comme pour la CNIL. Des travaux sont donc en cours entre la CNIL, le Health Data Hub et le ministère, pour encadrer le principe général des études susceptibles d'être autorisées, et les méthodologies de référence dans lesquelles elles pourront s'inscrire

pour accéder plus rapidement aux données. Un échange restera néanmoins nécessaire au terme de ce parcours pour déterminer quelles données le chercheur souhaite exactement extraire du SNDS. Les travaux de fluidification en cours sont donc nécessaires, et ils pourraient simplifier les parcours d'accès aux données pour un certain nombre d'études, mais il faut bien comprendre que ces données ne sont pas nativement construites pour la recherche, et qu'elles nécessitent un travail d'appropriation, généralement complexe, par les chercheurs. Naturellement, de plus en plus d'unités de recherche et de cabinets d'étude ont cependant déjà eu accès aux données du SNDS, et savent donc exprimer leurs besoins de manière pertinente et efficace. Tous les chercheurs « naïfs SNDS » (pour reprendre une expression courante en médecine) ont quant à eux besoin d'un certain temps pour comprendre quelles données sont présentes dans le SNDS et pouvoir les utiliser.

Or, comme vous le savez sans doute, et comme cela a dû être signalé par de nombreuses personnes au cours de vos auditions, le milieu de la recherche est en réalité très concurrentiel, et assez peu coopératif. Les producteurs d'une cohorte souhaitent donc d'abord valoriser leurs propres travaux sur cette cohorte avant de laisser les autres l'exploiter. Un vrai accompagnement des producteurs de données est donc nécessaire, pour qu'ils passent d'une attitude de propriétaire, à une volonté de partage de l'ensemble de leurs données. L'un des rôles du Health Data Hub est ainsi de réunir les « propriétaires » ou producteurs de données pour définir avec eux les conditions d'un partage qui reconnaisse aussi la tâche de production des données et la nécessité d'un retour sur investissement pour ceux qui s'attachent à produire des données, qui sont parfois exploitées par d'autres. Ce retour sur investissement est naturellement extrêmement important pour que les chercheurs continuent à construire des données et à les partager.

De la loi de 2019, résultent le SNDS élargi et un cadre réglementaire pour faciliter les appariements. Le HDH a aussi repris les missions de l'INDS pour les étendre à l'accompagnement de projets, au partage et à l'amélioration de l'usage des données pour l'ensemble des porteurs. Cette mission de fédération des acteurs est extrêmement importante pour le HDH.

S'agissant du choix de Microsoft par le Health Data Hub pour héberger ses données, je laisserai Mme Stéphanie Naux s'exprimer dans un premier temps.

Mme Stéphanie Naux. Lorsque vous avez reçu la directrice du Health-Data Hub, elle vous a expliqué les raisons de ses choix opérationnels. Il nous semble essentiel de trouver un équilibre entre, d'une part, la nécessité (totalement reconnue par la CNAM) de pouvoir utiliser plus largement et plus rapidement les données, dans l'intérêt de la santé de la population et de la recherche, et, d'autre part, les impératifs de sécurité des données (que la CNAM garantit depuis longtemps) et de maîtrise des usages. Cette maîtrise constitue en effet une priorité de très haut niveau pour la CNAM, en tant que responsable du traitement du SNDS et de la constitution de la base principale.

Lorsque la CNAM a construit le SNIIRAM et établi ses propres choix d'organisation à cette fin, la question du *cloud* ne se posait pas. Dès lors qu'elle a retenu le principe d'une architecture propriétaire, elle n'a pas été conduite à s'interroger, comme le Health Data Hub, sur la question du *cloud*, d'autant plus qu'il paraissait naturel de conserver cette architecture pour des données issues très largement des processus de gestion interne de l'assurance maladie.

M. Claude Gissot. Le conseil de la CNAM a ensuite été saisi par le ministère pour rendre un avis sur le décret dit « SNDS » (qui vise à organiser l'application de la loi de 2019), et, bien que cela ne fasse pas l'objet du décret, il a cherché à apprécier l'ensemble des conditions du projet HDH, et notamment les conditions d'hébergement de la plateforme. Compte tenu de la décision du Conseil d'État et de l'engagement du ministre sur le futur hébergement, cette question de l'hébergement doit être traitée le plus rapidement possible. Toutefois, comme cela a été signalé, la question du *cloud* souverain ne porte pas seulement sur les données de santé, même si celles-ci, de par leur sensibilité, et particulièrement celles du HDH, peuvent naturellement figurer parmi les premières concernées.

Le conseil de la CNAM a donc estimé nécessaire de trouver une solution à court terme à ce problème, et exprimé son opposition au transfert de la base de données centrale avant que cette solution soit trouvée. Son communiqué a néanmoins rappelé également qu'une plus grande utilisation des données était indispensable pour la médecine et la santé des assurés.

Par ailleurs, la CNIL a malgré tout autorisé la poursuite de projets sur la plateforme du HDH, s'agissant notamment de projets liés à la crise du Covid-19.

Il convient cependant de rappeler que le cadre fixé pour la transmission de ces données relève du pouvoir législatif et réglementaire sous le contrôle de la CNIL. Plusieurs étapes doivent encore être franchies avant d'en arriver là : la parution du décret SNDS, et la demande à la CNIL par la plateforme HDH, d'une autorisation pour traiter l'ensemble du SNDS. Mais, dès lors qu'un cadre juridique aura été fixé par ces textes et par ces autorités, la CNAM l'appliquera.

Mme Stéphanie Naux. La sécurité du SNDS fait l'objet d'un dispositif d'amélioration continue, qui date du SNIIRAM et qui a été contrôlé par la CNIL. Il ne s'agit pas d'un sujet récent pour la CNAM, même si l'actualité en rappelle souvent le caractère essentiel. Pour la CNAM, son enjeu dépasse également le seul SNDS, puisqu'elle gère des données de santé selon un périmètre plus large, et que des démarches de sécurité les concernent toutes.

S'agissant spécifiquement du SNDS, sa sécurité est encadrée par différents référentiels, dont un qui lui est spécifique, et qui est porté par un arrêté dédié, avec un niveau d'exigence de sécurité élevé. Il comprend un certain nombre de mesures spécifiques, comprenant un système d'authentification forte pour l'accès aux données et une vérification importante de la minimisation et du périmètre d'accès aux données personnelles, afin de s'assurer que les personnes autorisées accèdent uniquement aux données conformes aux autorisations délivrées par la CNIL. Tout ce processus s'inscrit dans des garanties contractuelles importantes. Les personnes physiques accédant aux données sont identifiées sous l'autorité du responsable de l'institution dont elles dépendent, qu'il s'agisse des accès permanents ou des accès sur projet. Les exportations de données non strictement anonymes sont interdites. Bien que pseudonymisées, les données du SNDS ne peuvent ainsi être utilisées et étudiées que dans des univers conformes au référentiel de sécurité. Seules peuvent être exportées des données strictement anonymes, donc fortement agrégées, la doctrine de la CNIL étant assez stricte en matière d'anonymat. Un système de traçabilité de l'activité des utilisateurs, et tout un ensemble de mesures techniques et organisationnelles faisant l'objet de revues régulières et de plans d'action ont été mis en place pour assurer cette sécurité de manière continue. Tout ce système de la CNAM est homologué, comme toutes les autres bulles susceptibles de recevoir des données du SNDS, ce qui permet de réajuster les risques à prendre en compte, et de mettre en place les mesures correctives requises pour les éliminer ou les diminuer.

M. Philippe Latombe, rapporteur. Depuis deux jours, une consultation a été lancée par la CNIL à propos des entrepôts de données de santé, afin d'aboutir à un cadre harmonisé entre tous ces entrepôts et à une doctrine en France plus cohérente. Constatez-vous aujourd'hui une distorsion entre les données de santé recueillies à différents endroits du territoire, lorsqu'il est ensuite demandé de les corréliser ou de les mettre en cohorte avec des données du SNDS ? Les pratiques, les modes de fonctionnement et l'attachement aux données de santé varient-ils en fonction des territoires ?

M. Claude Gissot. Je ne suis pas en mesure de vous répondre. Parmi les données de santé, il faut distinguer, en premier lieu, les données individuelles nominatives destinées à la gestion, et qu'on trouve dans les établissements de santé, etc. : pour ces données aussi, des questions de sécurité se posent, mais chacun de ces systèmes d'information se déclare à la CNIL, en précisant ses conditions de sécurité, et il revient à la CNIL d'autoriser ou non le traitement de ces données. En deuxième lieu, il existe des données pseudonymisées telles que nous vous les avons présentées, et qu'on trouve également dans des entrepôts hospitaliers, comme celui de l'AP-HP, même si elles n'utilisent pas nécessairement le même pseudonyme, et ne sont donc pas interconnectables immédiatement. Pour ces entrepôts aussi, une autorisation de la CNIL est nécessaire, sur la base d'une description du traitement et de l'ensemble des sécurités prévus pour ces données.

La CNIL a donc tout intérêt à mettre ces systèmes en cohérence, à défaut de les uniformiser, car chacun d'eux présente ses propres spécificités.

M. Philippe Latombe, rapporteur. Parler d'une mise en cohérence semble en effet adéquat.

M. Claude Gissot. Il s'agit de mettre ces systèmes en cohérence et de supprimer les écarts évoqués, s'ils existent, mais je ne dispose d'aucune documentation sur ces écarts. Je ne connais pas tous les systèmes d'information.

Mme Stéphanie Naux. Ces dernières années, la CNIL a été plusieurs fois sollicitée pour autoriser la mise en place d'entrepôts par différents acteurs, mais elle manquait souvent de bases juridiques à cette fin. Les dispositions de la loi Informatique et libertés ne lui laissaient pas nécessairement la latitude d'autoriser des entrepôts. Elle avait la possibilité d'autoriser des traitements, assortis de finalités, mais il lui était plus difficile juridiquement d'autoriser un réservoir de données susceptible de différentes utilisations. Des évolutions du texte ont eu lieu en ce sens, et il s'agit maintenant d'harmoniser les contraintes juridiques diversifiées dans lesquelles les entrepôts ont dû se constituer.

M. Philippe Latombe, rapporteur. Nombre des intervenants que nous avons auditionnés à propos des données de santé ont évoqué la question de leur valorisation. En quoi consisterait selon vous la valorisation des données du SNDS ?

M. Claude Gissot. Dans certains pays, les données s'échangent monétairement.

M. Philippe Latombe, rapporteur. Ce n'est pas le cas chez nous.

M. Claude Gissot. Ce n'est pas le cas chez nous, où les données administratives, notamment, ne sont finalement que réutilisées, puisqu'elles sont collectées pour nos propres usages de gestion, de sorte que la collecte est déjà assumée dans nos missions de remboursement de soins, etc. Je ne sais pas si vous avez interrogé la direction de la recherche, des études, de l'évaluation et des statistiques (DREES) ou le ministère sur cette question, mais aujourd'hui, la donnée SNDS n'est pas valorisable.

La question des services créés autour de la donnée peut toutefois se poser, car leurs utilisateurs peuvent être amenés à rémunérer les services créés par exemple par un groupement hospitalier (GH). Toutefois, l'objectif de ces services est plutôt de faciliter l'usage des données, et de les rendre les plus accessibles possible. Aucune valorisation de la donnée elle-même n'existe donc en soi.

La valorisation pour les producteurs se conçoit plutôt en termes de rôle ou d'image. Pour certains instituts de recherche, elle peut consister à gagner des points dans le système d'interrogation, de gestion, d'analyse des publications scientifiques (SIGAPS), permettant de valoriser l'activité de recherche. L'activité de constitution de bases de données par la recherche constituerait évidemment une forme de valorisation, au même titre que le fait d'être nommé dans les articles utilisant ces données, puisque ce type de références sont très importantes pour la recherche.

Les données de santé du SNDS ou du HDH sont de toute manière d'intérêt public. Or, l'intérêt public ne se monnaie pas. Leur valorisation vient du fait qu'elles sont partagées dans la collectivité, qu'elles servent à réaliser de bonnes études et à améliorer la prise en charge et la santé des patients.

M. Philippe Latombe, rapporteur. Les systèmes de santé sont différents entre tous les pays, mais des données de santé sont quand même collectées dans nos pays voisins. Échangez-vous avec vos « homologues » européens ? Examinez-vous leurs pratiques et certains de leurs modes de fonctionnement mériteraient-ils d'être importés ? À l'inverse, certaines de leurs pratiques doivent-elles absolument être évitées ?

M. Claude Gissot. Nous suivons en effet nos pratiques respectives de manière rapprochée. Même si nos échanges ne visent pas nécessairement à harmoniser les pratiques entre tous les pays, plusieurs initiatives européennes sont en cours pour mettre en place un espace européen des données de santé et faciliter l'utilisation des données de santé dans le cadre européen.

Les autres systèmes de santé ont souvent été créés de manière différente, mais aussi avec des pratiques de gestion des données assez différentes. Il est courant d'opposer les pays du Nord et du Sud : ils s'opposent sur ce point également. La pratique de la connexion des fichiers est ainsi souvent plus répandue dans les pays du Nord que dans les pays du Sud comme la France. L'appariement et l'utilisation croisée des sources sont beaucoup plus facilement acceptés dans les pays du Nord (le Danemark, la Suède, etc.). Ce n'est pas seulement une question de droit : la manière de gérer les données et l'acceptation du partage des données, même fines, sont aussi des faits de société.

En Angleterre, la *Clinical Practice Research Database (CPRD)* est intéressante, dans la mesure précisément où elle combine des données médicales et des données médico-administratives, mais elle ne l'a pas fait de manière exhaustive jusqu'à présent, ce qui lui confère une moindre puissance statistique.

Aux États-Unis, les données sont beaucoup plus riches, mais elles sont encore plus « silotées » qu'ici, puisqu'elles sont collectées par assureur. Même *Medicare* et *Medicaid* ne permettent donc pas de disposer d'une vision d'ensemble du système de santé, puisque toute une partie de la population n'y sera pas prise en compte.

En Allemagne, les systèmes d'assurance sont régionalisés également, même s'il existe certainement des bases de données communes. En Espagne et en Italie aussi, la collecte est relativement régionalisée.

Ce qui nous intéresse est d'examiner de quelles pratiques nous pourrions nous inspirer. Le HDH porte également cette démarche. La base de données médico-administratives centrale SNDS-SNIIRAM-PMSI est très utile, très utilisée et totalement exhaustive, avec une puissance statistique sans égal, mais il y manque un certain nombre d'informations médicales, ou médicalisées, qui constitueraient le complément idéal pour favoriser les études, les recherches, et finalement les innovations en santé. Nous pourrions nous inspirer d'autres pays à ce sujet, mais nous sommes parfaitement conscients de cet écart.

M. Philippe Latombe, rapporteur. Les citoyens français sont-ils aujourd'hui très attachés à leurs données de santé ? Leur accordent-ils une plus grande importance qu'auparavant, et sont-ils plus vigilants ? Constatez-vous un changement d'état d'esprit de nos concitoyens vis-à-vis des données de santé ? De ce point de vue, le compte AMELI a-t-il été reçu en tant que marque bénéficiant de la confiance des citoyens quant à la protection des données de santé qu'elle apporte ?

M. Claude Gissot. L'assurance maladie d'une manière générale est reconnue par ses assurés comme gérant leurs données de santé avec beaucoup d'attention et de précautions. Même si les enquêtes que nous réalisons ne portent que sur la perception qu'ont nos assurés de nos actions, elles font en tout cas ressortir que nos assurés nous font confiance,

Il faut toutefois développer la compréhension des données de santé et de leurs usages par le citoyen. Je ne sais pas si une évolution peut être mesurée à cet égard, car nous ne disposons pas d'un point zéro, ou d'une mesure de l'état de cette compréhension, il y a dix ans. Les débats afférents sont de plus en plus actifs dans la presse et dans la sphère publique ouverte. Il existe un réel besoin, pour les citoyens, de parvenir à maîtriser les données de santé et leurs usages, ce qui n'est évidemment pas le cas aujourd'hui. Lorsque nous offrons des services, que ce soit le compte AMELI ou le DMP, nous essayons d'être les plus clairs possible concernant l'usage qui y est fait des données. S'agissant du SNDS, la situation est différente, puisque les données sont pseudonymisées et ne donnent pas lieu à des usages individuels, mais collectifs de santé publique. Le degré de connaissance de ces enjeux est certainement encore très hétérogène selon les citoyens, et il faut d'ailleurs en reconnaître la complexité. Notre discussion et les auditions que vous conduisez aujourd'hui en attestent.

Il est important en tout cas que les associations de patients, et France Assos Santé en particulier, soient bien représentées dans le Health Data Hub, comme au conseil de la CNAM, etc., car elles constituent le bon intermédiaire avec les patients isolés, qui ne savent pas nécessairement quoi faire de leurs données de santé, ou ne voient pas l'intérêt de les partager. Elles constituent donc un vecteur très important pour faire progresser la conscience de ces enjeux.

Mme Stéphanie Naux. L'un des indicateurs les plus factuels dont nous disposons à cet égard tient aux remontées d'exercice de leurs droits par les personnes ou au nombre de questions qu'elles posent à ce sujet. Or, nous n'avons pas observé d'évolution majeure de ce point de vue.

La séance est levée à 12 heures.



Membres présents ou excusés

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

Réunion du mardi 9 mars 2021 à 10 heures

Présents. – MM. Éric Bothorel, Philippe Latombe, Jean-Luc Warsmann

Excusées. – Mme Frédérique Dumas, Mme Nathalie Serre