ASSEMBLÉE NATIONALE

XIVe LÉGISLATURE

Compte rendu

Commission des affaires sociales

Mission d'évaluation et de contrôle des lois de financement de la sécurité sociale

Audition, ouverte à la presse, sur les données médicales personnelles inter-régimes détenues par l'assurance maladie, versées au SNIIRAM puis au Système national des données de santé (SNDS) *(M. Pierre Morange, rapporteur)* :

Mardi
10 janvier 2017
Séance de 10 heures 30

Compte rendu n° 04

SESSION ORDINAIRE 2016-2017

Présidence de Mme Gisèle Biémouret et de M. Pierre Morange, ***coprésidents***

ASSEMBLÉE NATIONALE

COMMISSION DES AFFAIRES SOCIALES

MISSION D'ÉVALUATION ET DE CONTRÔLE DES LOIS DE FINANCEMENT DE LA SÉCURITÉ SOCIALE

Mardi 10 janvier 2017

La séance est ouverte à dix heures trente-cinq.

(Présidence de Mme Gisèle Biémouret et de M. Pierre Morange, coprésidents de la mission)

La Mission d'évaluation et de contrôle des lois de financement de la sécurité sociale (MECSS) procède à l'audition, ouverte à la presse, de M. Kamel Gadouche, directeur du Centre d'accès sécurisé aux données (CASD), et de M. Philippe Cunéo, directeur général du Groupe des écoles nationales d'économie et statistique (GENES), sur les données médicales personnelles inter-régimes détenues par l'assurance maladie, versées au SNIIRAM puis au Système national des données de santé (SNDS) (M. Pierre Morange, rapporteur).

M. le coprésident Pierre Morange, rapporteur. Messieurs, je vous souhaite la bienvenue. Dans le rapport sur les données personnelles de santé gérées par l'assurance maladie communiqué en mars 2016, la Cour des comptes a étudié la sécurité informatique du dispositif, la confidentialité des données et les conditions de leur exploitation. Elle s'est aussi penchée sur le modèle économique. L'accroissement considérable de la masse de données collectées par l'assurance maladie et l'extrême rapidité de l'évolution technologique confèrent à ces questions une importance stratégique. Alors que les données personnelles de plusieurs centaines de millions d'utilisateurs de la société *Yahoo !* ont été piratées et que, selon les indications du ministre de la défense, ce seul ministère a bloqué 24 000 attaques informatiques en 2016, on ne saurait mésestimer le danger que représente la multiplication des portes d'entrée dans le stock des données personnelles de santé détenues par l'assurance maladie. Le risque de captations dévoyées ne peut être ignoré et l'on sait la difficulté, en de tels cas, de réprimer et de faire cesser ce type de piraterie informatique criminelle.

La Cour a formulé diverses observations relatives à la lisibilité d'une gouvernance parfois redondante. Elle s'est inquiétée de la qualité du « coffre-fort informatique » de la Caisse nationale de l'assurance maladie des travailleurs salariés (CNAMTS) et a jugé « obsolescent » son algorithme de cryptage, utilisé pour occulter les identifiants nominatifs des données versées au système national d'information inter-régimes de l'assurance maladie (SNIIRAM). La Cour a enfin pointé la longueur de la procédure d'instruction des demandes d'accès aux données. Nous aimerions connaître votre sentiment sur ces différents sujets et vous entendre dire quel est, selon vous, le modèle économique pertinent pour développer l'accès à ces données, qui ne fait pour l'heure l'objet d'aucun encadrement législatif.

M. Philippe Cunéo, directeur général du groupe des écoles nationales d'économie et statistiques (GENES). Le groupe des écoles nationales d'économie et statistiques, constitué il y a cinq ans en établissement public, rassemble les anciennes écoles de l'Institut national de la statistique et des études économiques (*INSEE)* à savoir l'École nationale de la statistique et de l'administration économique, installée à Malakoff et qui va rejoindre dans deux mois le plateau de Saclay, et l'École nationale de la statistique et de l'analyse de l'information de Rennes. Nous gérons également le Centre d'accès sécurisé aux données (CASD), qui met à la disposition des chercheurs les données de la statistique publique et des données administratives, ainsi que, en commun avec le CNRS et le

département d'économie de l'École polytechnique, le Centre de recherche en économie et statistique.

Nous avons pour l'instant le monopole de la diffusion des données confidentielles de la statistique publique – les données fiscales en particulier. Qu'à l'avenir nous ne soyons plus les seuls opérateurs en ce domaine est une bonne chose, mais la question de la compatibilité des deux systèmes nous préoccupe. Nous exerçons le contrôle *a priori* de la diffusion des données, ainsi conçu qu'il est impossible à ceux qui ont accès aux données mises à disposition de les sortir de la « bulle sécurisée » dans laquelle elles sont confinées. Il est prévu de passer à un contrôle *a posteriori*. Soit, mais sous peine de ne pas être effectif, un contrôle de cette sorte, qui coûte cher, doit être systématique. Le sera-t-il ? D'autre part, si nous accueillons la concurrence avec satisfaction, nous souhaitons qu'elle soit réelle. À ce sujet, ce qui nous inquiète dans le dispositif en phase d'installation est que le producteur de données a le monopole de la gouvernance pour la diffusion de ses données. Il n'est pas certain que cela soit une bonne chose, car tout producteur faisant passer au premier rang ses contraintes de production, le risque existe que les besoins des utilisateurs passent au second plan. Le fait que l'activité du CASD soit entièrement tournée vers la diffusion des données et la satisfaction des utilisateurs est un bon modèle.

Pour en venir au modèle économique, nous demandons aux chercheurs une contribution que certains estiment trop élevée mais qui, rapportée aux sommes demandées par nos homologues au Royaume-Uni ou dans des pays comparables, est très faible. Cela est dû à ce que nous bénéficions d'un financement de projet « Équipement d'excellence », obtenu dans le cadre du programme d'investissements d'avenir (PIA). Ce financement viendra à échéance en 2019. Nous avons indiqué que, s'il n'est pas renouvelé, nous y suppléerons en augmentant le montant de la contribution. C'est le contrôle *a priori* des sorties de données qui coûte cher car cela demande du temps. Cela étant, il ne faut pas penser que le dispositif présenté pour l'accès aux données de santé n'aura pas ce coût puisqu'il prévoit un contrôle *a posteriori* indispensable. Il peut également coûter assez cher.

M. le coprésident Pierre Morange, rapporteur. Si je vous entends bien, le consensus existe sur les objectifs visés mais des interrogations demeurent quant aux procédures à suivre et, singulièrement, aux moyens à mettre en œuvre pour sécuriser et exploiter les données, parce que l'on passe d'un contrôle *a priori* à un contrôle *a posteriori* et que les deux systèmes ne sont pas encore parfaitement compatibles. Cela demande de définir les liens contractuels ou conventionnels du CASD avec la CNAMTS, l'Agence technique de l'information sur l'hospitalisation (ATIH) et le Système national des données de santé (SNDS). Pouvez-vous préciser quel devrait être le financement des moyens humains nécessaires au contrôle effectif de la sécurité des données détenues par l'assurance maladie ? Par exemple, que versent au CASD les chercheurs autorisés à consulter les données fiscales individuelles ?

M. Kamel Gadouche, directeur du Centre d'accès sécurisé aux données (CASD). Le CASD est né de la volonté de l'INSEE de mettre à la disposition des chercheurs les données individuelles détaillées couvertes par le secret statistique sans faire aucun compromis sur la sécurité. Le projet a été engagé en 2008 en étudiant les méthodes utilisées à l'étranger. Ce recensement a montré deux types d'accès à ces données : accès physique d'une part, accès par la voie informatique d'autre part.

Dans le premier cas, l'utilisateur se rend dans les locaux du diffuseur de données. Il est fouillé à l'entrée et son identité est vérifiée avant qu'il n'accède à un ordinateur, non relié

au réseau, où ont été enregistrées les données mises à disposition, et il peut y travailler. Lorsqu'il souhaite récupérer un résultat, il le demande spécifiquement ; à ce moment, un opérateur étudie ce résultat pour déterminer s'il ne contient pas de données confidentielles dissimulées volontairement ou involontairement. Cette vérification *a priori* étant faite, le fichier demandé est transmis à l'utilisateur, pas nécessairement le jour même. Un dispositif de ce type existe depuis plusieurs décennies en Amérique du Nord et depuis un peu moins longtemps en Allemagne et au Royaume-Uni et dans pratiquement tous les pays européens.

La seconde méthode utilisée est la transposition informatique du dispositif « physique ». Une infrastructure de traitement sécurisée garantit que les données restent confinées dans une bulle sécurisée sur un serveur auquel l'utilisateur peut se connecter à distance pour travailler sur des bases de données individuelles détaillées, sans qu'à aucun moment il puisse lui-même récupérer un fichier de données.

M. le coprésident Pierre Morange, rapporteur. Où est placé le boîtier SD qui garantit cette sécurité et comment en prend-on livraison ?

M. Kamel Gadouche. Le boîtier est placé sur l'ordinateur de l'utilisateur habilité, dans l'établissement où il travaille, après qu'il soit venu le prendre dans nos locaux ou que nous le lui ayons expédié par colis postal. Il peut alors accéder à distance à la bulle sécurisée.

M. Philippe Cunéo. Je précise que la connexion ne peut être établie qu'après l'insertion d'une carte sur laquelle figure une empreinte biométrique de l'utilisateur.

M. Kamel Gadouche. La transposition informatique du système d'identification physique se fait par le biais d'une carte à puce contenant l'empreinte digitale de l'utilisateur. Nous délivrons cette carte au terme d'une des séances dites « d'enrôlement » que nous organisons régulièrement. Elles ont lieu dans nos locaux parisiens et durent une demi-journée. Elles commencent par le rappel de dispositions législatives qui s'appliquent à l'utilisation des données, se poursuivent par une sensibilisation aux bonnes pratiques informatiques qui trouve sa traduction dans la charte d'utilisation du dispositif et finissent par la description des règles destinées à assurer le respect de la confidentialité des données lors de l'extraction des résultats. Ainsi ne peut-il y avoir moins de trois unités par tableau sur une sortie tabulaire et, en termes de pondération, une unité ne doit pas représenter plus de 85 %. On apprend en réalité à l'utilisateur à dépendre du producteur de données. Au terme de la séance d'enrôlement, l'utilisateur enregistre son empreinte digitale dans la carte à puce qui lui est remise, et nous lui confions un boîtier SD, ou nous le lui envoyons par la poste.

M. le coprésident Pierre Morange, rapporteur. Cette première étape est-elle payante ?

M. Kamel Gadouche. Son coût est inclus dans le forfait d'utilisation demandé. Notre grille tarifaire est affichée sur notre site. La contribution moyenne demandée est de 800 euros par an et par utilisateur – mais, comme je vous l'ai indiqué, ce montant a été établi en tenant compte de la subvention que nous recevons pour financer ce projet dit équipement d'excellence.

M. le coprésident Pierre Morange, rapporteur. Autrement dit, une contribution de 800 euros par utilisateur ne suffit pas à assurer à elle seule l'équilibre de votre bilan ?

M. Philippe Cunéo. Sans subvention, le montant demandé devrait être de 1 200 euros par utilisateur. Nous avons fait savoir aux chercheurs que c'est ce que nous leur

demanderons à partir de 2019 si la subvention au titre de l'équipement d'excellence n'est pas renouvelée.

M. le coprésident Pierre Morange, rapporteur. Cette contribution permet-elle le libre accès aux données ?

M. Kamel Gadouche. Oui, quel que soit le nombre de bases de données utilisées.

En matière de sécurité, le principe du dispositif est d'assurer un très fort confinement des données à caractère personnel pour permettre l'authentification des utilisateurs ainsi que la traçabilité des traitements et des sorties de résultats : bien qu'ayant été vérifiées *a priori*, les sorties sont conservées. Le CASD a été conçu pour mettre à la disposition des chercheurs les données de l'INSEE mais, peu à peu, d'autres producteurs de données individuelles détaillées se sont intéressés au dispositif : le ministère de l'agriculture puis, après que la loi l'a permis en 2013, le ministère des finances pour les données tirées des déclarations fiscales, les ministères de la justice – pour les seules données civiles –, de l'éducation nationale, du travail, et aussi la Banque publique d'investissement, ou encore l'Agence centrale des organismes de sécurité sociale. La liste complète des sources de données disponibles au CASD figure sur notre site.

M. le coprésident Pierre Morange, rapporteur. Combien d'utilisateurs servez-vous, et combien êtes-vous pour cela ?

M. Kamel Gadouche. Nous avons plus d'un millier d'utilisateurs. Notre effectif est de 25 personnes.

M. le coprésident Pierre Morange, rapporteur. Avez-vous une idée de l'impact qu'aura sur l'activité du CASD l'augmentation de la masse des données recueillie par le nouveau système national des données de santé ? Avez-vous engagé une réflexion à ce sujet avec la CNAMTS, l'ATIH et le SNDS ?

M. Kamel Gadouche. Il y a quatre ans, nous étions trois. Notre croissance a donc été assez rapide. Sur le plan technique, un développement futur ne présente pas de problème particulier, l'équipement que nous avons conçu pour répondre aux besoins de diffusion de données sécurisées fonctionnant comme un jeu de construction : pour tout nouveau projet, il suffit d'ajouter une brique et il peut y en avoir autant que nécessaire, sans limites.

M. le coprésident Pierre Morange, rapporteur. Le coût unitaire de 1 200 euros ne sera donc pas modifié ?

M. Kamel Gadouche. Non, sauf en cas de demandes de travaux particuliers sur une base exhaustive comme celle du SNIIRAM qui contient des téraoctets de données, car cela supposerait la création d'une infrastructure spécifique ; mais le surcoût ne serait pas considérable.

M. Philippe Cunéo. Nous diffusons déjà à une centaine d'opérateurs, pour le compte de l'ATIH, les données du programme de médicalisation des systèmes d'information (PMSI), qui forment une base assez considérable.

M. Kamel Gadouche. Effectivement, nous avons travaillé avec des cohortes mais nous diffusons depuis 2015 les données du PMSI sur les séjours hospitaliers à une centaine d'utilisateurs répartis dans vingt-cinq sociétés privées.

M. le coprésident Pierre Morange, rapporteur. Quel est l'état d'avancement du projet Teralab ?

M. Kamel Gadouche. Ce projet, défini dans le cadre du PIA est mené en partenariat par le GENES et l'Institut Mines-Télécom, comporte deux volets. L'un, coordonné par l'Institut Mines-Télécom, est une plateforme de technologie *Big data* sécurisée et accessible par internet ; elle permet à des laboratoires de recherche de travailler sur de vastes quantités de données et à des entreprises de faire des démonstrations de faisabilité. L'autre volet consiste en l'hébergement dans la « bulle » du CASD des compartiments ultra-sécurisés incluant des plateformes de *Big data*, permettant ainsi à des chercheurs ou à des sociétés privées de travailler sur de gros volumes de données dans un environnement hautement sécurisé. Les deux volets du projet sont déjà mis en œuvre et de nombreux utilisateurs tels BNP-Paribas ou le Réseau de transport d'électricité (RTE) travaillent sur des données confidentielles au sein de la bulle sécurisée du CASD avec des technologies *Big data*.

M. le coprésident Pierre Morange, rapporteur. Quelles relations entretenez-vous avec la Commission nationale de l'informatique et des libertés (CNIL) ? Avez-vous formalisé des relations contractuelles avec le SNDS ?

M. Kamel Gadouche. Nous entretenons évidemment des relations étroites avec la CNIL, auprès de laquelle nous avons dû, comme la loi l'exige, solliciter une autorisation pour le traitement des données biométriques. À cette occasion la CNIL a examiné en détail toute l'infrastructure du CASD avant de délivrer son autorisation. Elle est par ailleurs fréquemment consultée sur des projets de recherche ou des demandes particulières, sachant que les chercheurs, lorsqu'ils font une demande d'accès aux données personnelles, doivent préciser que le traitement de ces données s'effectuera par le CASD, dont la CNIL connaît maintenant les procédures.

M. Philippe Cunéo. Il faut ici évoquer le comité du secret statistique : lorsqu'un chercheur prétend accéder à des données confidentielles au travers du CASD, il doit présenter un dossier devant le comité du secret.

M. le coprésident Pierre Morange, rapporteur. Sans doute pourriez-vous nous fournir un schéma retraçant les différentes étapes par lesquelles passe un chercheur et nous donner une idée du temps qui s'écoule entre le dépôt d'une demande et l'exploitation effective des données.

Cette question des délais est en effet l'un des points faibles du dispositif d'exploitation des données de santé, qui incite à vouloir le réformer. Fruit d'une construction empirique, guidée avant tout par le pragmatisme, ce système est en effet organisé autour de procédures complexes, qui entraînent des embouteillages et des délais de traitement des demandes oscillant entre douze et dix-huit mois.

M. Philippe Cunéo. Pour ce qui nous concerne, le délai moyen de traitement des demandes est de six mois. Un chercheur qui veut accéder à des données confidentielles dépose un dossier auprès du comité du secret, lequel se réunit tous les trois mois. L'accord du comité doit ensuite être confirmé par la Direction des archives, avant que le dossier nous soit transmis et que le demandeur assiste à l'une de nos séances d'enrôlement, qui ont lieu une fois par mois. Nous travaillons actuellement à une informatisation des dossiers avec le comité du secret, ce qui devrait permettre de raccourcir encore ces délais.

M. le coprésident Pierre Morange, rapporteur. Qui compose ce fameux comité du secret statistique ?

M. Philippe Cunéo. Il est présidé par un conseiller d'État, en l'occurrence M. Jean Gaeremynck, et se compose notamment de représentants des producteurs de données et de la CNIL.

M. Kamel Gadouche. Le comité du secret statistique se compose au total d'une vingtaine de personnes, parmi lesquelles des représentants des syndicats, des chercheurs, des représentants du service interministériel des Archives de France, ainsi qu'un député et un sénateur.

Pour en revenir à la CNIL, nous réfléchissons actuellement à une demande d'autorisation unique qui simplifierait le dispositif : il s'agirait de définir un cadre précis, l'utilisateur n'ayant plus qu'à spécifier qu'il se conforme à ce cadre pour obtenir en retour son autorisation.

M. le coprésident Pierre Morange, rapporteur. C'est une procédure assez proche de ce que propose l'Institut des données de santé, bien qu'il existe, selon vous, des différences de procédure entre l'IDS et le CASD. Dans quelle mesure les deux modèles peuvent-ils converger afin que l'interface entre les données gagne en fluidité ?

M. Philippe Cunéo. Nous sommes très favorables à la convergence, en particulier pour permettre l'appariement des données c'est-à-dire la possibilité pour les chercheurs d'apparier des bases de données différentes, en l'occurrence les bases de données du SNDS et les statistiques publiques dont nous disposons. C'est essentiel. Nous nous heurtons néanmoins au fait que, dans un cas, le contrôle est fait *ex ante* et, dans l'autre, *ex post*, ce qui constitue une difficulté majeure.

M. le coprésident Pierre Morange, rapporteur. La question des appariements est en effet tout à fait essentielle, notamment dans le cadre de la lutte contre la fraude. Il me semble indispensable d'arriver à une coïncidence entre le numéro de sécurité sociale et le numéro d'identification fiscale : ce n'est qu'ainsi que nous disposerons d'une « machine de guerre » parfaitement opérationnelle.

En matière de contrôle, la Cour des comptes prône le contrôle *a posteriori*, qui permettrait de limiter l'embolie des circuits. Quelles conséquences cette solution emporte-t-elle en termes de moyens et quelle est votre position sur la question ?

M. Kamel Gadouche. Si nous avons conçu le dispositif avec un contrôle *ex ante*, c'est qu'il s'agissait de procédures nouvelles avec lesquelles les chercheurs, qui n'étaient pas habitués au respect de la confidentialité des données, n'étaient pas encore familiarisés. Pour limiter cependant les risques d'engorgement que risquait de provoquer ce contrôle *a priori*, nous avons mis en place une bulle sécurisée, qui offre aux chercheurs un environnement complet, « meublé » de tous les outils nécessaires pour exploiter les données, afin qu'ils puissent effectuer l'essentiel de leurs recherches et de leurs travaux au sein même de la bulle, sans avoir à en sortir les données. Avec six ans de recul, on peut considérer que le pari est gagné, puisque nous n'avons chaque jour que six à dix demandes de sortie de données, ce qui est très peu rapporté aux mille utilisateurs et aux quatre cents projets de recherche. Cela montre que, si l'environnement est complet et permet aux utilisateurs de travailler dans des conditions confortables, le contrôle *a priori* est un modèle qui fonctionne bien.

M. Philippe Cunéo. Certains chercheurs vont même jusqu'à rédiger leur article à l'intérieur de la bulle. C'est une manière de travailler bien différente de celle d'il y a trente ans, lorsqu'il fallait analyser des mètres de listings comme je l'ai fait pour rédiger ma thèse.

M. Kamel Gadouche. La seule limite que je vois au contrôle *ex ante*, c'est la question du délai, qui pousse certains utilisateurs à demander que le contrôle se fasse *a posteriori*. Dans ce cas, la sortie de données demandée est immédiate. Une copie en est conservée afin d'assurer la traçabilité. Les sorties sont ensuite vérifiées de manière aléatoire, par sondages. L'inconvénient de ce dispositif, notamment mis en œuvre au Danemark, c'est que, lorsqu'un fichier de données confidentielles est sorti, il se retrouve exposé et qu'il faut réagir rapidement pour contacter l'utilisateur et prendre les mesures qui s'imposent. Cela étant, lorsque le contrôle s'opère *a posteriori*, nous limitons, à l'instar des Danois, à la fois la taille et la fréquence des sorties, en fonction de la taille de la base de données. Il est impossible de sortir la totalité d'une base, dont ne pourront être extraits qu'un à deux mégaoctets de données – entre cinq et dix, si la base est vraiment importante –, le nombre de sorties étant limité, en fonction des utilisateurs, à quatre par jour ou à une sortie toutes les deux heures.

En ce qui concerne les relations contractuelles avec les différents acteurs, il faut garder à l'esprit que le CASD ne réalise aucune étude et que sa mission quasi unique est de mettre à disposition des utilisateurs les données qui lui ont été confiées. Nous sommes, en quelque sorte, un tiers de confiance, ce qui signifie que nous sommes liés contractuellement avec les détenteurs des données, par l'intermédiaire d'une convention qui spécifie les modalités de mise à disposition de ces données ainsi que les modalités d'habilitation. En tant que tel, le CASD ne procède à aucune habilitation, lesquelles dépendent soit du producteur de données soit de la loi, par exemple de la loi de 1951.

En aval, nous sommes liés aux utilisateurs par un triple contrat : en premier lieu, un contrat d'hébergement, qui stipule que le boîtier d'accès ne peut être connecté n'importe où et que l'accès à la bulle doit se faire à partir de l'établissement signataire du contrat, qui garantit des conditions d'hébergement très strictement définies – boîtier installé dans un bureau fermé, écran visible par le seul utilisateur… En second lieu, est signé un contrat de financement, que nous avons déjà évoqué, et, enfin, un contrat d'utilisation, qui spécifie les conditions de consultation auxquelles doit se conformer l'utilisateur.

M. le coprésident Pierre Morange, rapporteur. Des acteurs des médias font-ils également partie de votre clientèle d'utilisateurs, et que peuvent vous inspirer des demandes émanant, par exemple, de journalistes ? Avez-vous une doctrine particulière sur l'utilisation de certaines données à des fins médiatiques ?

M. Kamel Gadouche. Il est assez rare que des journalistes s'adressent à nous.

M. le coprésident Pierre Morange, rapporteur. Pourtant, compte tenu de la richesse de vos données et par souci de transparence, on pourrait imaginer qu'ils utilisent vos bases de données, dans un strict respect de la confidentialité, pour l'information de la population.

M. Philippe Cunéo. La longueur et la complexité des procédures dissuadent le plus souvent les journalistes. Il faut en effet du temps pour apprendre à utiliser les données. Par ailleurs, se pose le problème du coût.

M. Kamel Gadouche. Nous avons néanmoins déjà eu – mais très rarement – des demandes émanant de journalistes. Encore doivent-ils être accrédités en amont et garantir que le boîtier d'accès sera utilisé dans un lieu sûr et identifié. S'il s'agit du domicile du demandeur, son installation doit être conforme aux normes de sécurité en vigueur.

M. le coprésident Pierre Morange, rapporteur. Que pensez-vous des réserves formulées par la Cour des comptes sur les algorithmes de la CNAMTS qui, selon elle, seraient obsolescents ?

M. Kamel Gadouche. Les données mises à disposition par le CASD sont des données très détaillées, indirectement nominatives mais pour lesquelles il est assez simple de retrouver à quelle personne ou à quelle entreprise elles renvoient. La Cour des comptes se préoccupe de l'obsolescence de l'algorithme de pseudonymisation et, en effet, il est recommandé de ne plus utiliser l'algorithme SHA-1, considéré comme dépassé, mais de lui préférer SHA-256 voire SHA-3. Cela étant, la procédure mise en place par la CNAMTS est globalement robuste et, même si l'on gagnerait à une mise à jour de l'algorithme, le dispositif présente un bon niveau de sécurité.

M. le coprésident Pierre Morange, rapporteur. La Cour des comptes préconise de reconnaître à la CNAMTS le statut d'opérateur d'importance vitale, ce qui implique entre autres dispositions, le renforcement des algorithmes de protection des données. Cela est d'autant plus nécessaire que nous sommes face à une masse de données considérable, vouée à s'enrichir de l'ajout de nouvelles bases, comme celle des données médico-sociales des MPDH ou celles du Centre d'épidémiologie sur les causes médicales de décès (CépiDc).

Dans la mesure cependant où ce vaste ensemble de donnée est par ailleurs alimenté quotidiennement par des sources extérieures dont l'hétérogénéité informatique n'est plus à souligner, on imagine mal que le système puisse être parfaitement sécurisé. Qu'en pensez-vous et quelles seraient, selon vous, les dispositions à prendre pour garantir au maximum cette sécurité, sachant que, selon le ministère de la défense, il a lui-même subi 24 000 attaques informatiques cette année, ce chiffre étant voué à doubler chaque année.

M. Kamel Gadouche. La sécurisation des données est en effet un vrai sujet de préoccupation.

M. le coprésident Pierre Morange, rapporteur. C'est la raison pour laquelle la MECSS s'en préoccupe. Nous sommes censés remettre un pré-rapport dans la première quinzaine de février, avant la suspension de nos travaux. La MECSS issue de la nouvelle assemblée reprendra le flambeau et s'appuiera sur les premiers décrets d'application de la loi, qui doivent paraître dans le courant du printemps, pour rendre un rapport définitif fin 2017 ou début 2018.

M. Philippe Cunéo. Pour les questions vitales comme les épidémies, il me semble que seul un petit nombre de personnes – de hauts responsables – parfaitement identifiées, doivent avoir accès à un maximum d'informations, car il faut pouvoir agir avec rapidité. Pour le reste, il serait dangereux, selon moi, de multiplier les niveaux de confidentialité et d'accréditation, car cela ne peut qu'être compliqué à mettre en place. La meilleure solution, à mes yeux, serait donc un dispositif à deux niveaux, le premier niveau facilement accessible à une poignée de personnes, le second accessible selon le même protocole de sécurité par tous les utilisateurs.

M. le coprésident Pierre Morange, rapporteur. Les responsables de l'IDS – futur INDS – que nous avons auditionnés estiment que la montée en puissance de l'établissement devrait s'asseoir sur un effectif de douze personnes, indispensable pour fluidifier les procédures et assurer l'instruction des dossiers. De son côté, la directrice du service santé de la CNIL s'est interrogée sur la capacité des six personnes en charge des dossiers à faire face à la multiplicité des demandes et à les traiter dans des délais satisfaisants. Pourriez-vous évaluer ce que peut coûter la sécurisation du système, tel qu'il est organisé par la loi de modernisation de notre système de santé et quels sont les moyens requis pour que le SNDS fonctionne correctement ?

M. Kamel Gadouche. C'est difficile à évaluer, car nous ne sommes pas encore assez avancés dans les discussions…

M. le coprésident Pierre Morange, rapporteur. Cela signifie donc que des discussions sont en cours ? Nous avions plutôt le sentiment que les différentes structures se regardaient en chiens de faïence. Vous semblez évoquer une dynamique commune : un groupe de travail a-t-il été constitué ?

M. Kamel Gadouche. Il s'agit plutôt d'échanges informels, qui portent essentiellement sur la mise en place d'un référentiel de sécurité applicable aux données de sécurité.

Pour en revenir à la question du coût et de son estimation, le rapport de la Cour des comptes distingue, d'une part, le volet gestion et alimentation du SNDS, d'autre part, le volet diffusion. En ce qui nous concerne, la production et l'alimentation sont gérées par les détenteurs des données. Ce qui nous est transmis, ce sont des copies des bases, à charge pour nous de les installer dans une bulle sécurisée pour les mettre à disposition. Le coût de cette opération croît de façon linéaire avec le nombre d'utilisateurs, dont nous n'avons pas encore, en l'occurrence, d'idée précise, pas davantage que nous ne savons quels seront les usages qui seront faits du système. Cela étant, pour vous donner un ordre de grandeur, l'accès d'un millier d'utilisateurs coûte aujourd'hui 2 millions d'euros, sachant qu'une part de cette somme correspond aux importants investissements de départ et qu'elle devrait décroître après 2020.

M. Philippe Cunéo. J'ajoute que garantir une bonne traçabilité des données – ce qui implique des captures d'écran et des capacités de calcul et de stockage significatives – a un coût, tout comme la mise en place de contrôles *a posteriori*, qui réclament nécessairement du personnel.

M. le coprésident Pierre Morange, rapporteur. Avez-vous des informations sur les hébergeurs et les fournisseurs d'accès au SNDS actuels et futurs ?

M. Kamel Gadouche. Non.

M. le coprésident Pierre Morange, rapporteur. Avez-vous une méthodologie particulière pour d'autres hébergeurs avec lesquels vous seriez en éventuellement en contact ?

M. Philippe Cunéo. Nous n'avons pas de contact avec d'autres hébergeurs. L'important pour nous est de pouvoir diffuser nos données et de permettre à nos utilisateurs de les apparier avec les données de santé. Cela signifie qu'il va falloir réfléchir avec le SNDS aux mesures de sécurité à adopter, car ça ne peut pas être « ceinture et bretelles », c'est-à-dire un double contrôle, *ex ante* et *ex post*. Il va donc falloir nous accorder sur un dispositif

commun, qui devra être agréé par la CNIL, le comité du secret scientifique et la statistique publique européenne.

M. le coprésident Pierre Morange, rapporteur. La MECSS recommandera évidemment que tout soit fait pour que vous puissiez travailler dans des conditions satisfaisantes. Il serait dommage en effet de se priver de vos compétences et de votre expérience, notamment pour ce qui concerne l'harmonisation des procédures.

Cela étant, comment le nouveau dispositif peut-il, selon vous, s'adapter à la réglementation européenne ?

M. Philippe Cunéo. J'ai négocié, lorsque j'étais à l'INSEE, le règlement de la statistique européenne. Or nos partenaires européens, en particulier les Allemands, sont beaucoup plus frileux que nous en la matière. Ils refusent notamment de remettre en cause le principe du centre d'accès physique. Dans ces conditions, il sera difficile de leur faire accepter des dispositifs qui peuvent leur sembler moins sécurisés.

M. Kamel Gadouche. Lors de la mise en place du CASD, nous avons effectué préalablement une analyse de risque et une étude d'impact, d'où il est ressorti que le risque majeur était que les données de la statistique publique ou les données fiscales se retrouvent diffusées sur un site internet, avec tout l'impact négatif que cela aurait en termes d'image pour les institutions qui mettent à disposition ces données. À partir de là, nous avons pris les dispositions qui s'imposaient pour sécuriser le système et limiter les risques. Il me semble donc, dans la mesure où les usages, les utilisateurs, la nature des données et les risques ont été précisément définis et identifiés, que le dispositif répond aux normes de sécurité imposées par la réglementation européenne et que, s'il doit y avoir des adaptations, elles se feront à la marge.

Par ailleurs, nos analyses ont intégré les données de la statistique publique comme l'ensemble des données administratives, de façon à ce que le dispositif puisse être mutualisé. Cette cohérence globale, notamment en termes d'accès aux données et d'habilitations, était indispensable non seulement pour l'appariement mais également pour garantir un niveau optimum de sécurité, ce qui est conforme au règlement européen qui entrera en vigueur en mai 2018.

M. le coprésident Pierre Morange, rapporteur. Messieurs, nous vous remercions d'avoir ainsi enrichi notre réflexion.

La séance est levée à onze heures quarante.