

A S S E M B L É E      N A T I O N A L E

X V <sup>e</sup>      L É G I S L A T U R E

# Compte rendu

## Commission spéciale chargée d'examiner le projet de loi relatif à la bioéthique

Mercredi  
4 septembre 2019  
Séance de 12 heures

Compte rendu n° 26

– Audition de Mme Marie-Laure Denis, présidente de la Commission nationale de l'informatique et des libertés (CNIL), Mme Hélène Guimiot-Breud, cheffe du service santé, M. Erik Boucher de Crèvecoeur, ingénieur expert à la direction des technologies et de l'innovation, Mme Tiphaine Havel, conseillère pour les questions institutionnelles et parlementaires ..... 2  
– Présences en réunion ..... 13

SESSION EXTRAORDINAIRE DE 2018-2019

**Présidence de  
Mme Monique Limon,  
vice-présidente**



## COMMISSION SPÉCIALE CHARGÉE D'EXAMINER LE PROJET DE LOI RELATIF À LA BIOÉTHIQUE

**Mercredi 4 septembre 2019**

*L'audition débute à douze heures quarante-cinq.*

*(Présidence de Mme Monique Limon, vice-présidente)*



*La commission spéciale procède à l'audition de Mme Marie-Laure Denis, présidente de la Commission nationale de l'informatique et des libertés (CNIL), Mme Hélène Guimiot-Breaud, cheffe du service santé, M. Erik Boucher de Crèvecoeur, ingénieur expert à la direction des technologies et de l'innovation, Mme Tiphaine Havel, conseillère pour les questions institutionnelles et parlementaires.*

**Mme la présidente Monique Limon.** Notre matinée s'achève par l'audition d'une institution chargée de protéger une part de plus en plus importante de nos vies, celle qui est gouvernée par les sciences et techniques de l'information. Je souhaite donc la bienvenue à Mme Marie-Laure Denis, présidente de la Commission nationale de l'informatique et des libertés (CNIL), Mme Hélène Guimiot-Bréaud, cheffe du service santé, M. Erik Boucher de Crèvecoeur, ingénieur expert à la direction des technologies et de l'innovation, et Mme Tiphaine Havel, conseillère pour les questions institutionnelles et parlementaires.

Les techniques médicales contemporaines conduisent à recueillir, conserver, traiter et utiliser un nombre croissant d'informations. Certaines, comme les informations génétiques, ont une sensibilité particulière. Le projet de loi introduit des évolutions souvent substantielles en la matière. Je pense notamment aux articles 3, 9, 11, 18 et 24. La commission spéciale est donc très désireuse de recueillir les observations de la CNIL sur des sujets où elle devra intervenir en application de la loi.

**Mme Marie-Laure Denis, présidente de la CNIL.** Je suis très honorée de pouvoir m'exprimer devant votre commission spéciale, dans le cadre de la préparation de la révision de la loi de bioéthique. Je vous remercie de donner à la CNIL l'occasion de venir présenter les points de vigilance qu'elle a exprimés dans son avis rendu par son collègue, le 11 juillet dernier. Nos réflexions se sont nourries de différents travaux, notamment ceux du Comité consultatif national d'éthique. Je tiens également à souligner ici l'investissement du collègue de la CNIL et celui de notre rapporteur, Mme Valérie Peugeot, avec l'éclairage des services de la CNIL. Nous nous sommes prononcés dans des délais extrêmement contraints, comme l'a également regretté le Conseil d'État.

Je commencerai par quelques remarques de contexte sur l'action de la CNIL dans le domaine de la bioéthique. C'est la deuxième fois que le collègue de la CNIL est appelé à se prononcer formellement sur un projet de loi de bioéthique, mais elle n'avait plus eu l'occasion de le faire depuis les toutes premières lois de 1994. L'avis de la CNIL marque une nouvelle dimension dans ces débats situés aux confins du rapport entre l'identité humaine, le numérique et les nouvelles technologies. De manière plus générale, la CNIL intervient depuis longtemps sur des sujets d'ordre bioéthique, depuis la loi sur la République numérique et aussi, en octobre 2016, au titre de la mission qui lui a été confiée de conduire une réflexion sur les

problèmes éthiques et les questions de société soulevées par l'évolution des technologies numériques.

J'illustrerai mon propos par trois enjeux d'ordre bioéthique, dans la mesure où ils trouvent un écho dans ce projet de loi. Premièrement, les possibilités d'exploitation des données génétiques sont en plein essor. Nous regardons de près cette évolution, car ce sont des données très sensibles, pluripersonnelles, à fort potentiel prédictif et indélébiles. Jusqu'à présent, leur exploitation était très bornée et bénéficiait d'un cadre juridique protecteur. Aujourd'hui, en raison des évolutions technologiques (potentiels de calcul, possibilités de conservation sans altération), leur usage est bien plus aisé, comme en témoignent les nombreux projets de recherche en santé reposant sur la constitution de grandes bases de données génétiques, ainsi que d'autres usages qui se développent, comme les tests dits « récréatifs » accessibles sur Internet.

La deuxième évolution est la montée du recours à l'intelligence artificielle en santé. L'intelligence artificielle soulève en tant que telle des enjeux considérables, *a fortiori* appliqués au domaine de la santé. La CNIL a rendu public en 2017 un rapport intitulé *Comment permettre à l'homme de garder la main ?* ; appliqué au domaine de la santé, il s'agit de garder la main sur notre intimité. Ce rapport a mis en évidence deux principes fondateurs. Le premier est un principe de loyauté, selon lequel un algorithme ne doit pas entrer en opposition avec les intérêts des utilisateurs et des citoyens et doit offrir une certaine transparence. Le second principe cardinal est l'obligation de vigilance, qui entend permettre l'organisation d'une forme de questionnement régulier, méthodique et délibératif à l'égard des algorithmes.

Enfin, le dernier exemple d'enjeu bioéthique est la massification des données de santé et les potentialités de la recherche sur ces données. Les données de santé traitées dans le cadre des projets de recherche proviennent de différentes sources : les dossiers patients, les entrepôts de données, etc. Ces données peuvent également provenir d'échantillons biologiques et en cela, les collections d'échantillons représentent des sources de données presque illimitées, dans la mesure où les techniques de conservation et d'analyse sont en constante évolution et se perfectionnent. C'est un formidable potentiel, mais il soulève de multiples questions « informatique et libertés », que nous abordons à la CNIL avec la constante préoccupation de faire émerger des pratiques éthiquement responsables, tout en favorisant l'innovation et en garantissant la qualité de la recherche en France. Pour vous donner une indication, dans le domaine de la santé, la CNIL a délivré 360 autorisations de recherche. C'est à peu près une par jour, avec des dossiers de plus en plus complexes.

Sur le projet de loi, Madame la vice-présidente, vous avez précisé le champ de notre saisine, à la suite d'une consultation qui était obligatoire. La CNIL n'a pas eu à se prononcer sur ce qui n'était pas dans son champ de compétences, notamment sur les dispositions relatives à l'ouverture de l'assistance médicale à la procréation. Notre intervention a été centrée sur la vérification des conditions de la bonne application du cadre juridique « informatique et liberté » et a porté sur les articles que vous avez cités.

L'article 3 sur l'accès à leurs origines des enfants nés de l'assistance médicale à la procréation (AMP) est un article très important du projet de loi qui concilie le principe d'anonymat du don et un droit nouveau et très large de l'enfant d'accéder à des données non nominatives sur le donneur, dites « données non identifiantes », ainsi qu'à son identité. Dans son avis, la CNIL n'a pas arrêté de préférence entre les deux versions du texte qui lui avaient été initialement soumises. En revanche, elle a mis en valeur des points de vigilance. Premièrement, il faut être très clair sur les termes employés, parce que derrière eux se joue le

degré d'exposition réelle de la vie privée des donneurs. Le projet de loi utilise la notion de données non identifiantes, pour décrire une série d'informations accessibles, selon les cas, par le médecin ou par l'enfant. Ces termes sont ambigus. Ils laissent entendre qu'il s'agira toujours de données anonymes, c'est-à-dire de données qui ne se rattachent pas ou plus, de manière irréversible, à une personne physique. Or, à ce stade et dans l'attente des textes d'application, nous ne pouvons exclure que selon leur granularité et leurs possibilités de recoupement, certaines de ces données permettent de remonter à une personne physique identifiable. Nous ne pouvons donc exclure qu'elles ne restent pas des données non identifiantes, fût-ce indirectement. C'est un point de vocabulaire, mais c'est également une question de fond qui doit, selon nous, être clarifiée. L'enfant peut vouloir accéder aux données dites non identifiantes, mais qui sont en réalité potentiellement identifiantes, mais pas à l'identité du donneur. Ce choix doit être préservé.

Le deuxième enjeu de l'article 3 est l'information des personnes, notamment du donneur, sur les conséquences de son choix de faire un don. La transparence et la loyauté sont des principes généraux de la loi Informatique et Libertés et du Règlement général sur la protection des données (RGPD). Les personnes doivent être informées des traitements qui seront faits de leurs données de manière intelligible, accessible, claire et non ambiguë. Appliqués à ce contexte et compte tenu de la portée des traitements qui seront mis en œuvre, ces principes prennent une coloration particulière. L'information devra être particulièrement approfondie et mise à la portée des donneurs. Ceux-ci devront être conscients non seulement du couplage entre don et consentement à la transmission des données, mais aussi du fait qu'ils ne pourront pas s'y opposer. En l'état du texte, ils ne pourront pas s'opposer à ce que leurs données soient transmises par la suite. Cette pédagogie particulière est le corollaire de la portée du choix politique qui est fait.

Le dernier enjeu que je souhaiterais souligner sur l'article 3 est celui des futurs textes d'application qui détermineront le calibrage fin du dispositif et la portée réelle des risques pour la vie privée des personnes. Il est essentiel que la CNIL soit consultée sur le futur décret en Conseil d'État.

L'article 11 est quant à lui emblématique des enjeux inédits que dessinent les nouveaux usages des technologies, en particulier les traitements algorithmiques appliqués aux données de santé. Par ces objectifs généraux, cet article rejoint d'ailleurs les préoccupations soulevées par la CNIL dans son rapport de décembre 2017, que j'ai déjà évoqué. Cependant, il nous semble que les dispositions du projet de loi restent au milieu du gué, si je peux me permettre. Je ferai d'abord une remarque liminaire sur la structure de l'article, qui soulève d'importantes difficultés de compréhension. Il est certes question de traitements algorithmiques de données massives, mais à quel stade ? Parle-t-on du stade amont, de la mise au point des algorithmes dans une phase expérimentale ? Autrement dit, ce sont les données utilisées pour effectuer l'apprentissage et le paramétrage de l'algorithme. Parle-t-on, en aval, du stade de leur utilisation au bénéfice d'un patient en particulier ? Nous comprenons que le premier alinéa de l'article 11 vise la phase d'utilisation, mais la lecture du second alinéa est plus incertaine. Cette incertitude doit être levée clairement, parce qu'elle conditionne le sens même de ces dispositions, ainsi que le rôle donné aux médecins.

Cette observation liminaire étant faite, je soulignerai rapidement trois enjeux. Premièrement, le premier alinéa dispose que le recours à un traitement algorithmique de données massives doit être porté à la connaissance de la personne. C'est une contribution importante au principe de loyauté et d'information des personnes, mais le projet de loi ne précise pas la temporalité de cette information. Avant l'usage ou après, au stade de la

communication des résultats aux patients ? En toute logique, le principe de transparence devrait conduire à une information préalable. Cela rejoint l'esprit du code de la santé publique prévoyant que les décisions concernant la santé d'un patient sont prises par celui-ci, en ayant en main toutes les informations utiles.

Le deuxième enjeu en matière d'information des personnes se situe au moment de la communication des résultats. Il nous paraît souhaitable de bien distinguer deux niveaux d'information à délivrer aux patients. Il y a d'une part les informations à donner sur le résultat brut du traitement algorithmique, autrement dit ce que donne la machine et d'autre part, il y a les conclusions qu'en tire le professionnel de santé, à savoir son appréciation personnelle sur le résultat. Cette distinction est utile, notamment dans l'optique de préserver l'espace de la décision médicale, dont l'objectif souligné par la CNIL dans son rapport de 2017 de « garder la main », lorsque l'on a recours à des traitements algorithmiques.

Le troisième enjeu – qui dépasse le cadre du présent article – est celui de l'évaluation de la qualité et de l'efficacité des traitements algorithmiques. Certains peuvent être regardés comme des dispositifs médicaux. Je ne détaille pas, mais ils sont alors mis sur le marché et utilisés dans le cadre d'une législation spécifique. Il n'est pas évident que tous les traitements visés par cet article seront des dispositifs médicaux. Pour ceux qui ne seraient pas susceptibles de recevoir cette qualification, une réflexion devrait être conduite à la lumière des principes éthiques rappelés précédemment sur leur statut et l'existence d'une évaluation scientifique encadrée.

J'en viens à l'article 18 sur les échantillons biologiques et les données génétiques qui sont d'une particulière sensibilité. Ils nécessitent un régime de protection renforcée, car ces échantillons sont des sources de données intarissables, dans un contexte de multiplication des bases de données et de banalisation de l'analyse génétique. Nous sentons bien que, dans ce contexte, le terrain est encore plus propice au risque de réidentification. C'est pourquoi il est là aussi essentiel que la CNIL soit saisie pour avis du décret d'application qui définira les modalités d'information et d'opposition des personnes concernées.

De mon point de vue, il y a deux enjeux de fond en termes « informatique et libertés ». Le premier est l'information des personnes sur l'utilisation de leurs données à des fins de recherche. Le principe dans nos textes est l'information individuelle des personnes. Or, nous constatons que souvent, l'information individuelle n'est pas effective. D'ailleurs, les textes eux-mêmes consacrent des dérogations à ce principe. Dans le projet de loi, cette tendance s'observe également, puisqu'il envisage que les personnes soient informées des programmes de recherche – j'insiste sur le mot « programmes » – qui englobent par principe plusieurs projets de recherche, et non de chaque projet. Dans ce contexte, pour concilier les principes « informatique et libertés » et les enjeux opérationnels, nous recommandons que soient trouvées des solutions adaptées, par exemple en ayant recours à des dispositifs innovants, que nous pourrions détailler le cas échéant. Ce serait par exemple l'information individuelle initiale qui renvoie à un site internet, etc.

S'agissant des droits individuels, après l'information des personnes, le second enjeu est le droit d'opposition. Il est capital d'aménager la possibilité pour la personne d'exprimer son opposition, non seulement à la réalisation de l'examen, comme le précise le projet de loi, mais également à l'utilisation des données obtenues à la suite de l'analyse. Ainsi, il ne doit y avoir aucune ambiguïté à la lecture de l'alinéa 3 de l'article 18. Celui-ci ne saurait tenir en échec le droit d'opposition au traitement des données personnelles, même après la réalisation de l'examen. Ce droit d'opposition, les personnes le tirent directement du RGPD. Il est

indispensable que l'exercice de ce droit soit aménagé, en portant clairement à la connaissance des personnes les conditions dans lesquelles il pourra être exercé.

Enfin, la CNIL ne s'est pas prononcée sur un éventuel encadrement des tests génétiques récréatifs et pour cause, puisque le projet de loi ne prévoit pas de disposition en ce sens, comme l'a d'ailleurs rappelé devant vous le président du Comité consultatif national d'éthique. Nous sommes préoccupés par le décalage entre l'interdiction juridique de commercialiser en France des tests ADN grand public et dans la pratique, un recours croissant à ces tests par des dizaines de milliers de Français qui confient leurs données génétiques à des sociétés privées, essentiellement basées hors de l'Union européenne. Ces données génétiques sont les leurs, mais par ricochet, c'est également une partie de celles des membres de leur famille, ascendants ou descendants. En matière de protection de la vie privée, cela pose de nombreuses questions pratiques quant à l'exercice des droits des personnes concernées, mais également des questions éthiques quant à l'exploitation de ces données par des industriels, ainsi que leur conservation et leur sécurité, d'autant que ces données sont particulièrement identifiantes, puisque par nature, elles ne sont pas totalement anonymisées. Le choix appartient au législateur : soit, à droit constant, rappeler vigoureusement les interdictions prévues par la loi et peut-être les accompagner d'une campagne d'information ayant pour objectif d'expliquer clairement au public les risques encourus en recourant à de tels procédés en termes de protection de la vie privée ; soit modifier les textes applicables afin de créer un cadre protecteur dans lequel ces tests pourraient être effectués.

En conclusion, je rappelle que la CNIL a pour ambition de sensibiliser davantage la société à l'enjeu de la protection des données personnelles, de contribuer à faire des citoyens des utilisateurs éclairés et critiques des technologies, sans pour autant nier leur intérêt, de contribuer à nourrir la réflexion des pouvoirs publics dans un contexte d'innovation permanente. L'année dernière nous avons rendu plus de 120 avis sur des projets de textes et avons participé à plus d'une trentaine d'auditions parlementaires. Cette ligne de conduite me paraît particulièrement nécessaire dans le domaine de la bioéthique, car la protection des données personnelles est le moyen d'assurer la défense d'un modèle de développement humaniste, français et européen qui soit garant des droits et des libertés des personnes.

**M. Philippe Berta, rapporteur.** Ma première question porte sur la protection des informations dans les programmes de recherche. Qu'en est-il de l'effacement des données, après que la recherche entreprise a trouvé son issue à travers une publication ou un autre débouché ?

La publication des résultats pose un problème particulier. Vous pouvez faire s'engager la personne concernée par la recherche ou les personnes engagées par la recherche à ce principe de confidentialité, mais certaines pathologies sont dites rares. Il en existe environ 8 000, dont certaines concernent très peu de patients. Parfois, il n'est pas bien difficile de savoir de qui on parle, en lisant la publication. Que pouvons-nous faire dans ces cas-là ?

**Mme Marie-Laure Denis.** Le droit à l'effacement est prévu par le Règlement général sur la protection des données en matière de recherche, notamment dans les conditions détaillées à l'article 17. Il faut noter qu'il est possible d'écarter l'exercice du droit d'opposition lorsque le traitement est nécessaire à l'exécution d'une mission de service public. L'articulation est complexe avec des dispositions du code de la santé publique et du règlement européen sur les essais cliniques. En tout cas, il y a une vigilance accrue de la CNIL sur les durées de conservation des données, jusqu'à la publication du rapport, puis l'anonymisation, la suppression des tables de correspondance, l'archivage, etc.

**Mme Hélène Guimiot-Bréaud, cheffe du service santé de la CNIL.** Vous avez parlé du droit à l'effacement. Nous sommes également très vigilants quant aux modalités d'application du droit d'opposition. Très souvent, dans l'instruction des demandes d'autorisation, il est fait part du souhait d'écarter la possibilité d'exercer ce droit d'opposition, notamment pour les raisons que vous avez évoquées en termes de publication, etc. Les textes prévoient effectivement la possibilité d'aménager l'exercice du droit d'opposition voire de ne pas pouvoir exercer ce droit dans des cas très précis. En général, nous demandons d'informer les personnes en amont, c'est-à-dire que dès le départ, elles doivent savoir qu'il est possible qu'elles ne puissent pas exercer leur droit d'opposition. Vous avez évoqué des études fondées sur peu de cas, parce qu'elles concernent des pathologies rares. Effectivement, s'il y a dix cas en France et que cinq patients exercent leur droit d'opposition, il sera compliqué d'établir des résultats et de construire une publication dans laquelle il n'y ait pas de biais.

Quand il y a très peu de cas et que la publication pourrait permettre de reconnaître les patients, là encore, nous recommandons d'être extrêmement vigilant et de limiter la publication aux informations qui sont absolument nécessaires pour la compréhension scientifique de l'article. Par exemple, il n'est pas nécessaire d'utiliser le prénom des patients. Ce sont des choses qui paraissent assez évidentes, c'est du bon sens, mais cela fait partie des éléments recommandés. Quand les cas sont vraiment exceptionnels, uniques, là encore, le bon sens veut que le patient soit informé qu'une publication scientifique serait susceptible d'aboutir à ce qu'on le reconnaisse. C'est tout de même un cas extrêmement rare.

**M. Philippe Berta, rapporteur.** Vous avez parlé du prénom qui apparaîtrait. À mon sens, c'est l'une des ambiguïtés de l'article 18. Pour moi, il y a une vraie séparation entre le médecin et le responsable du programme ou du projet de recherche. Le plus souvent, ce n'est pas du tout la même personne. Il est clair que le transfert de prélèvements ou d'éléments relatifs au patient doit être anonymisé vis-à-vis du porteur du programme de recherche. Le scientifique n'a pas à connaître quelque nom que ce soit. Il doit éventuellement voir apparaître un numéro de code, mais rien de plus.

**Mme Hélène Guimiot-Bréaud.** Nous allons plutôt utiliser le terme de « pseudonymisation » ou de « données codées ». C'est inscrit dans la loi Informatique et Libertés. En recherche, le principe est que le promoteur de l'étude, le responsable du programme de recherche, même l'investigateur principal, c'est-à-dire celui qui va récupérer l'ensemble des informations, ne sont pas censés avoir accès à l'identité des personnes, puisque les noms et prénoms ne sont pas nécessaires pour obtenir des résultats scientifiques à partir des données collectées.

**M. Thibault Bazin.** Madame la présidente de la CNIL, vos propos confirment les propos que nous avons pu entendre sur les prélèvements biologiques de la part d'un certain nombre d'intervenants, notamment de chercheurs qui se posaient eux-mêmes des questions sur leur propre respect de la protection des données.

Dans quelle mesure les deux principes que vous avez mentionnés – loyauté et vigilance – peuvent-ils être respectés, lorsque l'intelligence artificielle se développe sans algorithmes ? D'autre part, dans quelle mesure la restriction de collecte de données, notamment due à la législation, peut-elle générer des biais dans les résultats à visée diagnostique ou thérapeutique à la suite du traitement ainsi tronqué de données massives ?

**Mme Marie-Laure Denis.** La CNIL a beaucoup travaillé sur l'intelligence artificielle et les algorithmes. Nous avons fait une soixantaine de débats et un rapport dont

nous avons essayé de tirer des principes, dont ceux de la loyauté et de la vigilance, avec l'idée qu'il faut qu'à un moment ou à un autre, il y ait une intervention humaine et un questionnement des process en cours. D'ailleurs, le RGPD exclut tout traitement de données totalement automatisé. M. Boucher de Crèvecœur, qui est expert à la direction des technologies et de l'information, va répondre plus précisément à vos deux questions.

Dans nos travaux, nous avons également mis en lumière les biais que vous évoquez dans votre deuxième question, mais je comprends bien l'idée de l'intelligence artificielle sans algorithmes.

**M. Erik Boucher de Crèvecœur, ingénieur expert à la direction des technologies et de l'innovation de la CNIL.** Que ce soit par apprentissage ou par des algorithmes plus manuels, la mise au point va de toute façon nécessiter d'être totalement conscient des biais qui peuvent être introduits par le choix des données massives sélectionnées pour alimenter le système, qu'il soit avec ou sans intelligence artificielle et avec ou sans algorithmes. Nous allons essayer d'isoler des modèles par rapport à des données massives que nous avons choisies. Ces modèles vont forcément hériter de biais, conscients ou inconscients, dans les données qui auront été présélectionnées pour ce travail.

L'un des points soulevés dans le rapport de la CNIL est qu'il faut rester durablement vigilant, c'est-à-dire au-delà de la mise au point de l'algorithme. Au-delà même de sa mise en routine, par exemple dans un parcours de soins, l'idée est d'avoir, par exemple, par une instance collégiale capable de porter un regard sur les bénéfices de cet algorithme pour les patients, de vérifier s'il y a des dérives et s'il rend bien le service attendu. Cela doit s'inscrire dans la durée, au-delà des tests préalables liés à la mise au point.

Sur les données tronquées, les données massives et les biais qui peuvent être introduits, il y a une tension entre le respect de la vie privée et le fait que les algorithmes ont besoin de beaucoup de données afin d'être mis au point. Il est donc important de maintenir cet équilibre entre les données strictement nécessaires à l'usage du système, à sa mise au point et le respect de la vie privée. Effectivement, dans certains cas, nous avons été sollicités sur les conséquences possibles de l'anonymisation ou de la pseudonymisation sur la qualité des informations recueillies. Ces problèmes sont souvent visibles sur des cas rares, puisque par nature, les cas rares sont identifiables assez facilement. Or, parfois, l'algorithme a justement pour fonction de retrouver des cas rares. Là, nous arrivons vraiment à la limite de l'exercice.

**M. Pierre Dharréville.** Le traitement des données massives et leur utilisation posent des questions de respect des libertés individuelles, mais également de respect des libertés collectives, puisque les données collectées servent à établir un certain nombre de profils. À la suite de la réponse que vous venez de faire, il ne s'agit pas tant de savoir comment vous appréhendez ces enjeux, mais de prendre conscience qu'en termes de données de santé, ce n'est tout de même pas neutre.

Ma deuxième préoccupation concerne la propriété des données et des logiciels propriétaires, qui risquent d'être de plus en plus utilisés par les différents acteurs. Cela touche également le stockage des données. Une autorisation ne devrait-elle pas être requise pour utiliser ces données dans le cadre du *big data* et de logiciels de traitement massif ? J'essaie d'exprimer ces interrogations le plus clairement possible, mais elles me semblent assez vertigineuses.

**Mme Marie-Laure Denis.** Il est vrai que la protection des données personnelles prises dans leur ensemble a structurellement une incidence sociale allant au-delà de la

protection de la vie privée d'un individu. C'est l'une des problématiques soulevées par ce texte dans le domaine de la santé, comme dans d'autres.

**M. Pierre Dharréville.** Avez-vous des préconisations, notamment sur les logiciels à utiliser dans le cadre de ce qui nous occupe ?

**Mme Hélène Guimiot-Bréaud.** Sur le fait de savoir s'il faudrait une autorisation pour inclure des données dans un traitement de *big data* ou une base de données massives, tout dépend. Le terme « consentement » est peut-être un peu galvaudé, dans la mesure où le traitement de données de santé n'est pas nécessairement soumis au recueil du consentement des personnes concernées. C'est effectivement une condition au regard de laquelle la loi peut autoriser le traitement de données de santé, mais le consentement exprès n'est pas nécessairement requis. En revanche, est bien requise une information de la personne dans le cadre du RGPD comme de la loi Informatique et Libertés.

Nous pouvons prendre l'exemple du système national des données de santé (SNDS). C'est une base de données dont la CNAM est responsable. Toutes les données des assurés sociaux y figurent. Une information a été aménagée par la loi, mais le consentement de chaque assuré social n'a pas été demandé, afin que les données de tous figurent dans la base. En revanche, il y a théoriquement une information et chacun peut s'opposer à ce que ses données soient utilisées dans le cadre de projets de recherche. Ce mode de fonctionnement est mis en œuvre ailleurs, par exemple dans les grands entrepôts hospitaliers de données. Je pense notamment à l'entrepôt de données de santé qui a été constitué au sein de l'AP-HP et a fait l'objet d'une autorisation de la CNIL. C'est également un système d'information, avec une possibilité pour les personnes de s'opposer au traitement de leurs données. Théoriquement, pour tout traitement de données massif, pour toute base qui aspire des données – par exemple les entrepôts hospitaliers aspirent les données des dossiers médicaux – les personnes concernées devraient être informées.

**M. Thibault Bazin.** Pourquoi dites-vous « théoriquement » ?

**Mme Hélène Guimiot-Bréaud.** En pratique, il y a des difficultés, que la présidente a évoquées, notamment dans le cadre de projets de recherche – mais nous allons les rencontrer aussi avec les entrepôts hospitaliers. Le souci est la reprise de l'existant. Quand on crée un entrepôt à partir des nouveaux patients qui viennent consulter, il est possible de les informer par la remise de documents et autres actions, mais pour ceux qui sont venus consulter dans les années 1990, on n'est pas sûr d'avoir une adresse valable à laquelle envoyer un document.

**M. Thibault Bazin.** Pour rebondir sur cet élément, l'un des articles du projet de loi porte sur les prélèvements biologiques qui ont été faits sur des patients qui ne sont pas venus en vue de recherche, mais pour être soignés. Le texte prévoit qu'à partir du moment où il y a eu une intervention pour conserver ces prélèvements, il suffit d'informer les patients. L'intervenant nous a dit que vous préconisez un mois pour exprimer cette opposition, mais nous ne savons pas s'il faut l'envoyer par mail, par courrier et s'il faut vérifier qu'elle est bien arrivée. L'idée d'une autorisation tacite vous semble-t-elle suffisante pour exprimer un consentement libre et éclairé sur des recherches qui au départ n'étaient pas prévues sur ces échantillons biologiques ?

**Mme Hélène Guimiot-Bréaud.** Le délai d'un mois est quelque chose que nous avons pu évoquer avec des responsables de traitement pour gérer concrètement certaines situations. Par exemple, ils nous disent : « Nous avons une adresse assez ancienne. Si nous

envoyons une information, que nous n'avons pas de réponse au bout d'un mois, pouvons-nous mettre en œuvre le traitement ? » Cela relève vraiment de la mise en œuvre pratique de ce qui est prévu dans l'article 18 – qui préexistait, puisque ce sont des dispositions incluses dans le code de la santé publique depuis la loi Jardé. Le principe est l'information de la personne, qui doit pouvoir s'opposer. Encore une fois, nous ne sommes pas dans un consentement exprès : le responsable de traitement n'attend pas un acte positif du patient pour mettre en œuvre le traitement. C'est plutôt l'inverse : si le patient indique qu'il ne souhaite pas que l'on traite ses données, on ne va pas mettre en œuvre le traitement sur ces données ou on va l'interrompre.

**M. Pierre Dharréville.** Je reviens sur ma question concernant la propriété des données et des logiciels. Comment faire pour qu'il y ait réellement une protection publique et pas d'utilisation privée abusive d'un certain nombre de données liées soit à la propriété des logiciels, soit au stockage, à la propriété des données elles-mêmes, etc. ?

**M. Erik Boucher de Crèvecœur.** S'agissant de la propriété des algorithmes, nous essayons de développer avec les différents industriels et les chercheurs le fait qu'une fois que les règles de traitement auront pu être déduites du *big data* et que des données massives auront été traitées, ces règles soient anonymes. Il s'agit donc de faire en sorte qu'elles n'embarquent pas par erreur ou parce que le système n'aurait pas été bien « nettoyé », des éléments identifiables liés à un patient un peu spécifique ou ayant une maladie rare et qui serait facile à repérer. Dès lors que cette réflexion sur l'algorithme est correctement menée au moment de sa mise au point, que nous avons bien vérifié que ne restent dans les règles que des choses neutres et pas liées à une personne en particulier – par exemple, nous ne voulons pas trouver une règle qui dise : « Si c'est un centenaire, qu'il s'appelle Maurice, alors il a un cancer » –, alors les règles définies n'ont plus de lien avec des données personnelles et sont industrialisables et vendables sans être soumises à des contraintes d'information ou de consentement.

**Mme Hélène Guimiot-Bréaud.** Pour l'instant, il n'y a pas de droit de propriété sur les données. À ma connaissance, ce n'est pas la solution qui a été retenue prioritairement par le collège des commissaires de la CNIL. En tout cas, la CNIL n'a jamais poussé pour que soit établi un droit de propriété sur les données.

Pour ce qui est des utilisations de données de santé en général, notamment par des sociétés privées, et du risque d'un usage abusif, le principe prévu dans la loi Informatique et Libertés est l'utilisation dans une finalité d'intérêt public. Tous les projets de recherche sont soumis à autorisation de la CNIL, sauf lorsqu'ils sont conformes à des référentiels que la CNIL a elle-même établis. Par le passé, nous avons pu autoriser des entrepôts de données de santé massives mis en œuvre par des sociétés privées, mais en vérifiant que le principe de l'intérêt public était respecté dans la finalité des traitements mis en œuvre. Cela fait partie des missions de la Commission.

**M. Pierre-Alain Raphan.** J'aurais aimé avoir votre point de vue sur l'extraterritorialité des lois. Nous nous inquiétons notamment du *Cloud Act* américain et de ses conséquences en matière de stockage de données, certains pays pouvant accéder à nos données sans forcément avoir le consentement d'un patient, d'un citoyen et, sans même parler d'un consentement éclairé, un consentement tout court. Il y a des enjeux très forts, au regard de la protection de ces données et d'une forme de souveraineté numérique, voire génomique. Vous l'avez dit tout à l'heure à propos des tests génétiques : nous ne savons pas vraiment comment sont exploitées ces données à l'étranger.

Quel est votre point de vue sur les données de santé récupérées, notamment, par les montres ou les enceintes connectées au regard de l'exigence de consentement éclairé ? Ces dernières semaines ont vu les GAFAs avouer clairement : « Nous écoutons tout, nous prenons toutes les données, que vous soyez d'accord ou non. » Ces données biométriques ne sont pas utilisées contre le citoyen, mais elles le sont sans son autorisation. Avez-vous des éclairages ou des préconisations à nous apporter ?

**Mme Marie-Laure Denis.** Vous abordez des sujets qui dépassent de beaucoup le cadre du projet de loi. Certains députés ont eu l'occasion de m'entendre sur le *Cloud Act*, à la fois devant la commission des Affaires européennes de l'Assemblée nationale et devant la commission d'enquête sur la souveraineté numérique au Sénat. Il est clair qu'il y a un conflit de lois avec l'article 48 du RGPD, sur lequel j'avais beaucoup insisté. Nous avons été très actifs sur ce sujet au sein du Comité européen de la protection des données, qui réunit les CNIL européennes. Tout récemment, au mois de juillet, ce sujet a fait l'objet d'une publication de ce Comité, à laquelle je me permets de renvoyer. Elle explicite clairement les choses.

La CNIL a un laboratoire d'innovation qui décortique les enceintes connectées. Nous voyons que ces enceintes peuvent notamment écouter des informations quand nous pensons qu'elles ne les écoutent pas. Quand une personne dit : « Hey, Alexa ! » parce qu'une personne présente dans la pièce s'appelle aussi Alexa, cela peut déclencher un mécanisme d'écoute. Les utilisateurs de ces enceintes ne sont pas conscients des risques encourus. Des tiers arrivent dans des foyers où il y a ces enceintes et ne sont pas au courant qu'il y en a. C'est un enjeu d'autant plus important que la diffusion de ces objets s'accélère. Nous parlons des enceintes connectées, mais la question peut se poser pour un grand nombre d'autres objets. Les industriels qui fabriquent et exploitent ces enceintes disent parfois recueillir des données pour améliorer leurs appareils et pas pour exploiter les enregistrements. C'est ce qu'ils disent, mais leurs pratiques interrogent...

Pour nous, c'est un vrai sujet de préoccupation. C'est d'ailleurs l'un des trois axes de travail de la CNIL dans le domaine de l'innovation, de la technologie et de la prospective pour l'année à venir.

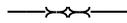
**M. Erik Boucher de Crèvecœur.** Sur les objets connectés et les données de santé, nous avons l'habitude de faire un petit propos liminaire. Nous identifions différents degrés de risque, d'identification et de même de classification de ces données. Je prends l'exemple d'un bracelet servant à suivre votre jogging. Si vous avez juste un parcours sur une journée, cela reste une donnée de bien-être assez classique. À partir du moment où vous aurez l'historique des courses sur une année, on pourra voir les moments où vous n'êtes pas allé courir, peut-être parce que vous étiez malade, et inférer éventuellement des données de santé. Si le dispositif fait en plus le suivi du poids ou d'autres éléments diététiques liés au parcours sportif, ce sont clairement des données de santé au regard du RGPD. Un gradient supplémentaire est la notion de donnée médicale, par exemple si l'objet connecté est prescrit ou recommandé par un médecin afin de suivre son patient.

Autour de ces objets connectés, il y a donc tout un gradient de données plus ou moins sensibles, plus ou moins identifiables qui révèlent des choses plus ou moins intimes sur la personne. Notre laboratoire effectue un suivi assez régulier des « bonnes pratiques », en décortiquant quelques objets. Régulièrement, à la Station F, nous faisons également des interventions auprès des start-up qui nous posent les mêmes questions que vous, afin de savoir dans quelle classe de données ils vont se loger, s'ils ont des obligations par rapport à l'hébergement des données de santé, etc.

Le principe cardinal est qu'il ne faut stocker que des données strictement pertinentes. La pertinence des données est l'une des bases de notre loi Informatique et Libertés. Toutes ces données sont-elles nécessaires pour rendre le service promis à l'utilisateur ? L'utilisateur est-il conscient que ses données vont être collectées afin de lui rendre ce service ? La pertinence des données et l'information de l'utilisateur sont souvent un peu malmenées par les objets connectés, parce qu'il n'y a pas d'interface ou parce que l'on doit juste ouvrir et installer la boîte. Avec les industriels, nous réfléchissons à trouver les moyens d'informer les personnes *via* l'application, des indicateurs, des voyants qui s'allument, au moment où les données sont collectées.

**Mme Monique Limon, présidente.** Merci beaucoup pour vos interventions et vos réponses, qui vont au-delà du projet de loi.

*L'audition s'achève à treize heures trente.*



## **Membres présents ou excusés**

### **Commission spéciale chargée d'examiner le projet de loi relatif à la bioéthique**

#### **Réunion du mercredi 4 septembre à 12 heures**

*Présents.* – M. Thibault Bazin, Mme Valérie Beauvais, M. Philippe Berta, Mme Marine Brenier, M. Pierre Dharréville, M. Pierre-Henri Dumont, Mme Agnès Firmin Le Bodo, Mme Emmanuelle Fontaine-Domeizel, M. Brahim Hammouche, Mme Monique Limon, M. Maxime Minot, M. Jean-Pierre Pont, M. Alain Ramadier, M. Pierre-Alain Raphan, Mme Laëtitia Romeiro Dias, M. Pierre Vatin

*Assistait également à la réunion.* - Mme Christine Hennion