

A S S E M B L É E N A T I O N A L E

X I V ^e L É G I S L A T U R E

Compte rendu

Commission des affaires sociales

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

Auditions, ouvertes 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*):

- M. Philippe Burnel, délégué à la stratégie des systèmes d'information de santé (DSSIS), secrétariat général du ministère des affaires sociales et de la santé 2
- Dr Dominique Blum, praticien hospitalier..... 12

Mardi

13 décembre 2016

Séance de 10 heures

Compte rendu n° 02

SESSION ORDINAIRE 2016-2017

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



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

Mardi 13 décembre 2016

La séance est ouverte à dix heures dix.

(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 d'abord à l'audition, ouverte à la presse, de M. Philippe Burnel, délégué à la stratégie des systèmes d'information de santé (DSSIS), secrétariat général du ministère des affaires sociales et de la santé, 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. Dans le cadre de son travail sur les données médicales personnelles inter-régimes détenues par l'assurance maladie et versées au système national d'information inter-régimes de l'assurance maladie (SNIIRAM) puis au système national des données de santé (SNDS), notre mission d'évaluation est heureuse d'accueillir M. Philippe Burnel, délégué à la stratégie des systèmes d'information de santé (DSSIS) au secrétariat général du ministère des affaires sociales et de la santé.

S'agissant d'une matière en plein changement, connaissant en particulier des évolutions législatives et technologiques, la Cour des comptes a, dans un rapport sur les données personnelles de santé gérées par l'assurance maladie, regroupé les principales questions dans quatre grands chapitres : la sécurité du système informatique, la confidentialité des données, leur utilisation et le modèle économique.

Pour ce qui est de l'utilisation, la loi de modernisation de notre système de santé de janvier 2016 a fait en sorte de préserver ces données de leur marchandisation par les assurances complémentaires et les laboratoires, alors que des organismes ayant toute légitimité à les exploiter les utilisent peu : est-ce le fait d'une inertie administrative ou bien le recours à ces données est-il peu entré dans les pratiques mises au service d'une stratégie nationale de la santé ?

Sur l'aspect économique, étant donné les moyens humains, techniques et financiers à développer pour lutter contre la piraterie informatique, et protéger et traiter la masse colossale des données collectées – nous avons affaire au premier fichier d'Europe dont le volume sera quasiment multiplié par cinquante dans les cinq prochaines années grâce aux objets connectés –, peut-être est-il temps de réfléchir à une redevance que pourraient acquitter certains partenaires économiques qui recourraient à ces données, en dehors de celles qui relèvent de la liberté d'information gratuite et qui sont anonymisées.

M. Philippe Burnel, délégué à la stratégie des systèmes d'information de santé (DSSIS), secrétariat général du ministère des affaires sociales et de la santé. J'ai coanimé avec Franck Von Lennep, directeur de la direction de la recherche, des études, de l'évaluation et des statistiques (DREES), la commission dite « *open data* en santé », au sein de laquelle nous avons mené une réflexion sur les conditions d'ouverture du SNIIRAM, que la loi du

26 janvier 2016 de modernisation de notre système de santé a intégré au SNDS. C'est principalement à ce titre que je répondrai à vos questions, car ma compétence est moindre concernant l'écriture de la loi et sa mise en œuvre, qui relèvent davantage de la DREES. Mes réponses sur ces sujets ne pourront donc qu'être approximatives.

Dans le questionnaire que vous m'avez transmis préalablement à cette audition, vous indiquez que, s'agissant de l'usage des données, trois destinations sont explicitement exclues. La commission *open data*, qui associait l'ensemble des parties prenantes – patients, professionnels de santé, industriels, administrations, chercheurs –, avait estimé que l'intérêt général justifiait l'ouverture de données au public, tout en reconnaissant les risques encourus en termes de confidentialité. Après de nombreux débats, nous avons considéré que cet intérêt public n'interdisait pas l'existence, en parallèle, d'un intérêt privé, mais que ce dernier ne pouvait constituer la seule finalité pour accéder aux données médicales. L'accès, refusé à ceux qui poursuivent un intérêt exclusivement privé, pourrait donc être accordé à ceux dont l'intérêt privé est porteur d'un intérêt public, et, bien sûr à ceux dont l'intérêt serait uniquement public. Nous avons, en conséquence, considéré que la collecte de données par un assureur privé ayant pour objectif de sélectionner les risques d'un assuré ou par un fabricant de produits de santé à des fins de repérage commercial ou de segmentation de la visite médicale n'allait pas dans le sens de l'intérêt public et devait donc être exclue par la loi.

Parmi les trois finalités exclues par la loi mentionnées dans le questionnaire, « *la quantification des risques en vue de leur couverture par une assurance complémentaire concurrentielle* » procède clairement d'une logique de sélection du risque et d'adaptation de la tarification. À ce titre, elle est prohibée. En revanche, pour « *la discipline des pratiques de soins par les assureurs concurrents qui les remboursent* », tout dépend de l'objectif poursuivi. Il ne s'agit pas ici de sélectionner les assurés, mais plutôt de réguler l'offre de soins. La loi ne me semble donc pas exclure cette utilisation. Ce n'est pas le cas du « *ciblage territorial ou professionnel du démarchage commercial des prescripteurs de médicaments et de dispositifs médicaux* », qui est exclu sans ambiguïté dans l'article 193 de la loi de 2016 précitée, car ce genre de démarche est contraire à l'intérêt public et à celui de l'assurance maladie.

M. le coprésident Pierre Morange, rapporteur. Je comprends votre prudence s'agissant de la diffusion de données destinées « *à la discipline des pratiques de soins par les assureurs concurrents qui les remboursent* », car, dans le cadre des réseaux de soins, elle pourrait porter atteinte à la liberté de la pratique médicale. Ce sujet a-t-il donné lieu à des approches différentes au sein de la commission ?

M. Philippe Burnel. Cela n'a pas vraiment été le cas. Il n'y a pas eu davantage de contestation des industriels présents s'agissant des deux finalités explicitement écartées. Il était clair que toute segmentation visant les professions de santé ou les assurés devait être exclue. Les industriels, qui avaient bien compris que ces aspects pouvaient motiver un refus total de l'accès aux données, ont préféré les écarter d'eux-mêmes pour bénéficier d'autres accès.

Ce qui n'a pas été exclu n'est donc pas interdit, sans pour autant être autorisé. Il ne peut en être décidé qu'après étude des dossiers et des finalités.

M. le coprésident Pierre Morange, rapporteur. L'idée d'une base de données publique favorisant une concurrence plus ouverte entre assureurs complémentaires est-elle écartée ?

M. Philippe Burnel. Le SNDS répond à des objectifs d'intérêt public, de recherche et de pilotage. Il est globalement au service de la régulation du système. L'utilité de renforcer la concurrence entre les assureurs complémentaires n'entraîne pas dans notre champ de réflexion. Mais si le régulateur, considérant que cet objectif est d'intérêt général, souhaitait mettre en place un mécanisme en ce sens, on peut imaginer que la base de données pourrait être utilisée. Pour l'heure, ce n'est pas le cas.

M. le coprésident Pierre Morange, rapporteur. La base de données est-elle adaptée aux besoins d'une recherche scientifique sur les soins ? La sous-utilisation des bases existantes par les institutions publiques s'explique-t-elle par une sorte d'inertie administrative ?

M. Philippe Burnel. La base SNIIRAM a été créée par l'assurance maladie, sous le pilotage de l'État et de la direction de la sécurité sociale, pour ses propres finalités de gestion du risque. À l'époque, les partenaires de l'assurance maladie ne manifestaient à l'égard d'une telle base de données qu'une assez grande indifférence, voire un certain scepticisme. Le programme de médicalisation des systèmes d'information (PMSI) a connu le même parcours. En charge de ce programme il y a fort longtemps, je me souviens avoir subi les quolibets d'enseignants en santé publique m'expliquant que ce programme n'avait aucun intérêt pour la recherche et qu'il n'en aurait jamais aucun. Avec le temps, les chercheurs ont pris conscience du grand intérêt de ces données pour la recherche médicale, parfois même pour la recherche clinique, et pour le pilotage de la santé publique.

Leur demande s'est d'abord exprimée à l'égard du PMSI, car il est d'un abord plus simple et son langage est plus médical que celui du SNIIRAM qui comporte essentiellement des données relatives aux remboursements. Avec le SNIIRAM, on peut toutefois obtenir des informations sur les pathologies à partir des données relatives aux médicaments prescrits, et la maladie des patients atteints d'une affection de longue durée (ALD) est mentionnée. Il donne donc accès à une information partiellement clinique, mais uniquement de façon indirecte.

Les données cliniques constituent aujourd'hui le gros manque du SNIIRAM. Si elles sont recueillies dans les établissements de santé dans le cadre des résumés de séjour, les pathologies ne sont pas codées en médecine de ville. Une information qui n'est pas recueillie n'a aucune chance de se retrouver dans une base de données. L'accès que permet le SNIIRAM à des données exhaustives sur l'ensemble de la population et à un chaînage sur toute la vie du patient en fait cependant une base extrêmement riche dont l'intérêt est majeur en matière de recherche. Le travail sur le Mediator, visant à montrer la relation entre un médicament et des complications hospitalières, a été effectué à partir de la base de données SNIIRAM.

M. le coprésident Pierre Morange, rapporteur. L'inertie administrative dont je parlais est-elle due, selon vous, à la lenteur et à la complexité des procédures d'accès aux données ?

M. Philippe Burnel. Plusieurs facteurs jouent. Déjà, l'accès aux données était limité aux administrations et n'était autorisée que de façon récente. Les premiers acteurs publics à avoir eu accès aux données sont les chercheurs, notamment ceux de l'Institut national de la santé et de la recherche médicale (INSERM).

En 2009, la loi portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires (HPST) prévoit que les agences régionales de santé (ARS) qu'elle créait pourraient

avoir accès aux données SNIIRAM « dans des conditions garantissant l'anonymat des personnes », et sous l'autorité d'un médecin. Cette dernière mention, assez traditionnelle, est fondée sur l'idée que seuls les médecins peuvent accéder à des données médicales. Il s'agit, à mon sens, d'une interprétation un peu abusive du secret médical : le médecin a accès aux données médicales du malade non parce qu'il est médecin, mais parce qu'il le soigne. Or cette exigence pose problème, car le SNIIRAM est d'une telle complexité qu'il est nécessaire d'être statisticien pour en comprendre les données – pour bien faire, il faut donc être à la fois médecin et statisticien.

La condition relative à l'anonymat était une deuxième difficulté importante, liée à l'évolution de la perception de ce qu'est une donnée anonyme. Lorsque j'étais en charge du PSMI, nous avons créé le résumé de sortie anonyme (RSA) de l'hôpital qui tronquait certaines informations : au lieu des dates d'entrée et de sortie, c'est la durée du séjour qui était mentionnée, l'âge remplaçait la date de naissance... À l'époque, nous étions persuadés d'avoir « anonymisé » le RSA. En 1994, la Commission nationale de l'informatique et des libertés (CNIL), saisie de l'arrêté de création de ce dispositif, avait même considéré qu'elle n'avait pas à se prononcer s'agissant de données anonymes. Cependant, progressivement, nous avons pris conscience, et la CNIL avec nous, que ces données n'étaient pas « anonymisées » mais seulement « pseudonymisées ». Dans certaines conditions, il était possible d'identifier le patient, et si l'on « repérait son voisin de palier », on pouvait accéder à l'ensemble de ses données. Par maladresse ou approximation, la loi HPST n'a pas pris en compte cette nouvelle approche et a repris l'exigence initiale d'anonymat. En conséquence, lorsque nous avons voulu organiser l'accès des ARS au SNIIRAM, la CNIL nous a clairement opposé la mention de l'anonymat, dont la persistance dans les textes a compliqué la vie de l'administration. L'article 193 de la loi de 2016 est revenu sur cette rédaction.

Outre les problèmes juridico-administratifs liés à la mauvaise écriture des textes, une autre difficulté tenait à la complexité de la base de données, qui présente quasiment que des informations de remboursement brutes avec des codes de modulations tarifaires. Bien l'appréhender demandait donc un investissement en temps non négligeable.

Ces obstacles à l'usage du SNIIRAM ne concernaient pas le PSMI, d'accès plus facile. Ce dernier était intensément utilisé par l'administration centrale comme par les ARS. Or la loi a redéfini le SNDS en réunissant dans un même ensemble les données PMSI et SNIIRAM et en en fixant des règles d'accès identiques. De ce fait, en même temps que l'accès devenait ouvert à plus d'acteurs, les conditions en étaient durcies, ce qui a compliqué les choses pour ceux qui utilisaient le PMSI avant la loi de janvier 2016. Reste que, dans la sphère administrative, les usages du PMSI sont très développés tant en matière de planification que de construction de projets d'établissement de santé ou d'étude de pertinence des soins, ceux du SNIIRAM le sont un peu moins – les ARS s'en servent notamment pour déterminer les territoires de permanence des soins.

M. le coprésident Pierre Morange, rapporteur. La loi incite les assureurs complémentaires et les industriels du médicament à déléguer leurs requêtes à des bureaux d'études conventionnés par la CNIL. Pourraient-ils être réunis aux tiers de confiance par lesquels ces assureurs alimentent les SNDS ?

M. Philippe Burnel. Ils exercent deux métiers différents, mais je n'ai pas vraiment de religion. Je note que la loi permet aux assureurs complémentaires et aux industriels du médicament de ne pas déléguer leurs requêtes à un bureau d'études s'ils démontrent que leur méthodologie garantit qu'ils n'enfreignent pas les interdictions relatives aux usages prohibés.

De nombreuses études ont prouvé que les travaux sur un médicament, s'ils sont menés par le laboratoire qui le produit, sont généralement plus positifs que ceux effectués par des tiers plus neutres. Si l'on veut mieux garantir l'objectivité des travaux conduits à l'initiative d'un acteur de l'industrie pharmaceutique, du dispositif médical ou de l'assurance, il semble judicieux de passer par un bureau d'études qui engage sa crédibilité et son image de marque.

Les bureaux d'études et les spécialistes du traitement de données font un métier différent de celui du tiers de confiance, même s'ils peuvent être tous les deux complémentaires. *A priori*, je ne vois pas de raison d'affirmer qu'il faudrait qu'ils soient strictement séparés, mais je ne vois pas davantage l'intérêt de les réunir.

M. le coprésident Pierre Morange, rapporteur. Pour en venir maintenant aux questions relatives au contrôle des usages et à la gouvernance du SNDS, que pensez-vous de la préconisation de la Cour des comptes relative à un contrôle *a posteriori* des usages du SNDS ?

M. Philippe Burnel. La réflexion de la Cour portait sur la question suivante : le SNIIRAM, compte tenu de son coût – dont l'évaluation est difficile –, est-il suffisamment utilisé ? En d'autres termes, est-il efficace ? Cette question rejoint celle que vous m'avez posée sur le caractère assez récent du développement des usages du SNIIRAM dans la sphère administrative à des fins de régulation.

Par la même occasion, la Cour s'est intéressée aux aspects de sécurité. Nous en avons débattu avec elle et le Gouvernement a apporté plusieurs éléments de réponse dont certains sont directement pris en considération dans le rapport tandis que d'autres figurent dans les observations.

Il y avait, entre nous et la Cour, deux divergences. L'une concernait la gestion du risque par la Caisse nationale de l'assurance maladie des travailleurs salariés (CNAMTS) qui est chargée de gérer le SNIIRAM. Nous rangeant à l'avis du responsable de la sécurité des systèmes d'information (SSI) du ministère, nous considérions cette gestion comme sérieuse, avec un plan d'action fondé sur une analyse du risque. Mise à disposition, celle-ci a permis de mettre en évidence les risques résiduels inévitables dans toute politique de sécurité, de les calibrer et de chercher à composer avec eux plutôt qu'à tenter de les supprimer, ce qui paraît plus raisonnable. Il s'agit souvent de risques humains, comme des malversations de la part d'une personne travaillant dans les services.

L'autre divergence portait sur l'obsolescence de l'algorithme de hachage, qui aurait pour conséquence de lever l'anonymat. Si le premier risque est celui d'une infraction, d'une entrée illicite dans le système, le second serait qu'à l'occasion d'une entrée tout à fait légitime, cette fois, l'on puisse obtenir la clé de hachage permettant d'identifier les individus. La Cour a donc pointé l'ancienneté de la version utilisée et préconisé le passage à l'algorithme de la génération suivante. Selon la CNAMTS et le responsable SSI du ministère, le problème ne se posera pas au cours des quatre ou cinq ans à venir, même s'il faudra bien planifier ce changement, dont les informaticiens nous annoncent d'ailleurs qu'il sera lourd de conséquences, notamment financières.

M. le coprésident Pierre Morange, rapporteur. L'un des freins au relèvement du niveau de sécurité est donc, comme toujours, financier. On a pu parler d'un coût de 15 à 20 millions d'euros ; confirmez-vous cette somme ?

M. Philippe Burnel. Je ne saurais ni la confirmer ni l'infirmier. Je n'ai pas les moyens d'estimer ce coût.

M. le coprésident Pierre Morange, rapporteur. Au regard des dépenses de soins, qui représentent plus de 200 milliards d'euros, ce serait en tout cas un investissement plutôt modique pour relever le niveau de sécurité et mettre en œuvre la proposition de la Cour sur le classement de la CNAMTS en tant qu'opérateur d'importance vitale.

M. Philippe Burnel. C'est un troisième aspect évoqué par la Cour, à propos duquel je n'ai plus en mémoire le détail de nos discussions avec elle. J'en ai retenu que la mesure ne changerait pas grand-chose à la réalité de l'action conduite. Mais je n'ai pas ce dossier suffisamment en tête pour vous répondre.

M. le coprésident Pierre Morange, rapporteur. Le réseau comporte d'innombrables entrées qui constituent autant d'opportunités pour les pirates informatiques. Cet aspect fait-il l'objet d'une réflexion en vue de relever le niveau de sécurité et de normer la sécurisation informatique pour combler les éventuelles lacunes du mécanisme ?

M. Philippe Burnel. Oui, et cette réflexion est d'autant plus nécessaire que le dispositif est plus ouvert. Jusqu'à présent, seul un petit nombre de personnes pouvaient y accéder, selon des conditions de sécurité et des systèmes d'authentification très rigoureux : il s'agissait presque d'un réseau fermé. En outre, son utilisation faisait l'objet d'une forme de travail à façon, la CNAMTS préparant les jeux de données à la demande pour les mettre à la disposition des acteurs, dans une quasi-relation *BtoB* de partenariats privilégiés permettant de savoir à qui on ouvrait les portes. Dans ces conditions, le système était de fait très sécurisé.

La CNAMTS va devoir rompre avec ce mode de fonctionnement artisanal pour ne pas créer des embouteillages qui rendraient l'ouverture des données purement formelle, et non réelle. Cela suppose d'organiser une infrastructure différente. L'ouverture devra respecter le référentiel de sécurité du SNDS tel que le prévoit la loi, rendu opposable par arrêté après avis de la CNIL. Ce référentiel est aujourd'hui quasiment achevé ; Franck Von Lennep pourra vous dire plus précisément que moi où il en est exactement.

C'est d'abord pour le PMSI qu'il est urgent de renforcer la sécurité. En effet, ses données ont été accessibles à certains acteurs sous la forme assez souple, que d'aucuns jugeaient même laxiste, de transmission sous forme d'un CD-Rom. Cette possibilité ayant été supprimée, plusieurs industriels se plaignaient à juste titre de ne plus pouvoir faire leur métier, ce qui les mettait en péril. Il a donc fallu redéfinir rapidement avec eux les conditions d'utilisation du PMSI. C'est ainsi que l'on a entrepris de construire des bulles de sécurité, environnements sécurisés dans lesquels est organisé le transfert de l'information en provenance de l'Agence technique de l'information sur l'hospitalisation (ATIH). L'acteur autorisé à recevoir cette information est physiquement désigné et les données sur lesquelles il travaille sont soumises à un traçage informatique, pour parer aux risques résiduels humains évoqués. Un industriel installé à Lyon a déjà mis en œuvre ce système pour pouvoir continuer de bénéficier des données du PMSI, anticipant ainsi l'arrêté relatif au référentiel de sécurité du SNDS.

M. le coprésident Pierre Morange, rapporteur. Pourriez-vous nous dire quelques mots de la gouvernance spécifique du SNDS, de l'articulation entre l'Institut national des données de santé (INDS) et la DREES et qu'en est-il du rattachement du comité stratégique du SNDS ?

M. Philippe Burnel. Le SNDS est conçu en trois lieux. D'abord, un lieu de pilotage stratégique, qui réunit les décideurs publics, sous la présidence de l'État, pour définir les évolutions du SNDS, identifier les bases de données à y intégrer, se prononcer sur l'opportunité d'un sous-échantillon d'accès plus facile, bref définir les règles d'alimentation et les grands principes régissant la vie du système.

Ensuite, le comité des producteurs, qui rassemble les producteurs de bases de données – la CNAMTS, l'ATIH, demain les structures du secteur médico-social telle la Caisse nationale de solidarité pour l'autonomie (CNSA). C'est un lieu de nature plus technique, où les acteurs s'accordent sur les formats de fichiers, les modalités de transmission, les modèles d'agrégation, la définition de délais, d'indicateurs, etc. Cet aspect est assez classique.

Le troisième lieu est plus nouveau, même s'il existait déjà sous la forme de l'Institut des données de santé (IDS) : c'est l'INDS. Il est doté de deux missions que l'on a initialement envisagé de séparer, avant que la concertation au sein de la commission *open data* ne montre que la quasi-totalité de nos interlocuteurs souhaitaient les voir réunies.

La première mission consiste à accueillir le secrétariat du comité d'experts chargé d'analyser les demandes d'accès des acteurs privés ou des administrations qui ne disposent pas d'un accès permanent, avant de les transmettre à la CNIL à laquelle il appartiendra *in fine* d'autoriser ou non le traitement. Le rôle du comité d'experts, qui était lui aussi déjà exercé, est assez simple dans son principe : il s'agit de déterminer si l'étude est sérieuse du point de vue méthodologique, s'il est justifié d'utiliser les données comme annoncé pour atteindre les finalités annoncées, si le nombre de données demandées n'est pas trop élevé, etc.

L'idée était ensuite d'englober ce guichet d'autorisation dans une structure de concertation ou de démocratie sociale au sein de laquelle toutes les parties prenantes demeurerait représentées et qui prolongerait, en quelque sorte, la commission *open data*. Il était important, en effet, que perdure un lieu où tous les acteurs pourraient régulièrement dialoguer, afin d'exprimer leurs besoins – ceux des industriels, des patients, des professionnels, indépendamment de la conception propre à l'administration de l'évolution du SNDS – et de préserver et garantir l'intérêt public.

Ce dernier point a été abondamment discuté. Les associations de patients, en particulier, craignaient que la séparation entre les experts, d'une part, et la démocratie sanitaire, de l'autre, n'encourage les experts à privilégier les recherches conduites par des scientifiques, de grande qualité méthodologique, au motif que le progrès des connaissances est en soi une finalité d'intérêt public. Les objectifs poursuivis par les patients, légitimes mais formalisés de manière moins experte, risquaient ainsi d'être pénalisés par la dichotomie entre les deux structures. D'où l'idée de réunir les uns et les autres pour mettre les experts en relation avec des interlocuteurs qui les laisseront juger de l'aspect méthodologique, mais pourront faire valoir auprès d'eux la contribution à l'intérêt public d'un dossier, même mal ficelé, et leur suggérer de donner des conseils au demandeur pour l'améliorer.

Ma conviction personnelle est que cette notion d'intérêt public, dont la définition est par nature ouverte, va évoluer et faire peu à peu l'objet d'une jurisprudence au sein de l'INDS. Nous en aurons probablement une conception plus large dans trois à cinq ans.

M. le coprésident Pierre Morange, rapporteur. Je ne suis pas certain que vous m'ayez répondu sur le contrôle *a posteriori* ou *a priori*, en lien avec le problème de la fluidité de l'accès aux bases de données.

Quant à la réunion des experts et de la démocratie sanitaire, elle renvoie à la question, à laquelle nous sommes régulièrement confrontés, de la possibilité d'un référentiel de normes de qualité pour les différents producteurs de soins. Car l'on observe des manipulations de la part de certains médias ou des manœuvres de *lobbying* émanant d'associations qui s'appuient sur des paramètres dont l'objectivité est contestable. Le dialogue que vous évoquez ne pourrait-il permettre d'établir enfin des normes de référence ?

M. Philippe Burnel. La question du contrôle *a priori* ou *a posteriori* a été l'un des sujets de discussion de la commission *open data*. Une partie de nos interlocuteurs nous y a mis en garde contre la multiplication des barrières, soutenant qu'il suffisait de contrôler *a posteriori* et de punir sévèrement les contrevenants, selon une vision du modèle que l'on pourrait qualifier d'anglo-saxonne. Telle n'était pas la position des représentants de l'État, au motif que, politiquement – au sens le plus noble du terme –, il est très délicat pour un État, vu notre culture nationale, d'assumer le risque qu'un dommage advienne, dans un domaine qui nécessite la confiance. Ce qui ne veut pas dire qu'il n'existe aucun contrôle *a posteriori*.

En ce qui concerne le référentiel et la qualité méthodologique, la loi prévoit que l'utilisation des données publiques a pour contrepartie la publication des études qu'elles auront servi à conduire, y compris dans l'hypothèse où leurs résultats ne seraient pas aussi favorables que l'utilisateur l'espérait. C'est une garantie de qualité.

Par ailleurs, le comité d'experts sera probablement amené à se doter d'une forme de jurisprudence sur différents sujets. On sait que, dans le domaine de la recherche scientifique, des essais sur les médicaments ou les dispositifs et, d'une manière générale, dans les procédures en santé, coexistent un *gold standard* et des méthodes dont le niveau de preuve est inférieur. À titre personnel, j'estime que ce sont les scientifiques qui doivent déterminer les bonnes méthodes, avant l'administration.

M. le coprésident Pierre Morange, rapporteur. Mais qu'en est-il de l'évaluation à laquelle procèdent différents acteurs – médias, associations – selon des critères dont la méthodologie scientifique est parfois pour le moins aléatoire ?

Mme la coprésidente Gisèle Biémouret. Le référentiel peut-il apporter plus de clarté ?

M. Philippe Burnel. J'ai connu ce débat lorsque j'étais chargé du PMSI. En dehors des chercheurs, les premiers à avoir demandé un large accès aux données du PMSI étaient les journalistes, ce qui nous a un peu surpris. C'est ainsi que l'équipe de *Science & Avenir*, qui a rejoint ensuite d'autres publications, a pu publier son premier palmarès. J'ai ensuite suivi l'aventure. J'ai observé qu'avec le temps les indicateurs, d'abord très frustes, se sont perfectionnés. Je me souviens ainsi d'un indicateur d'attractivité des hôpitaux selon la proportion de patients venus d'un autre département : il offrait une prime extraordinaire aux hôpitaux frontaliers, fussent-ils de qualité médiocre. De manière générale, la démarche a suscité de nombreux débats, certains jugeant préférable que les patients ne disposent d'aucune donnée plutôt que de données erronées.

À cette question, chacun donnera sa réponse ; peut-être le politique pourra-t-il s'en emparer, ce qui produira une réponse collective. Ma conviction est qu'il faut progresser dans ce domaine, que l'ouverture permet aussi de disposer d'une pluralité d'indicateurs et de diagnostics et que c'est en ce sens qu'il faut aller.

Nous en avons beaucoup discuté avec les médecins. Aujourd'hui, les indicateurs de qualité sont très largement diffusés chaque année dans les palmarès des établissements de santé ; ils ne concernent pas encore la médecine de ville, mais des systèmes d'affectation de scores en ville se dessinent. Nous avons, d'ailleurs, été interpellés par un syndicat médical et par l'Ordre national des médecins à propos des sites en ligne qui évaluent les professionnels. Nous avons répondu qu'une interdiction serait inenvisageable, et impossible à l'ère d'internet. La meilleure des réponses consiste à travailler ensemble à un système d'évaluation valable ; c'est ce que nous avons proposé.

M. le coprésident Pierre Morange, rapporteur. C'est ce à quoi je songeais.

Venons-en au modèle économique, qui n'est pas sans lien avec la question du contrôle *a priori* ou *a posteriori* : l'État peut légitimement préférer à la conception anglo-saxonne, pour des raisons qui touchent à l'éthique politique, un encadrement de l'accès aux banques de données, mais celui-ci suppose des moyens humains, techniques et financiers. Or on sait que le dispositif est pour le moins sous-dimensionné au regard du vivier de données à exploiter et des intérêts, voire des appétits – plus ou moins légitimes –, que celui-ci peut susciter. À l'époque où j'ai interrogé la CNIL à ce sujet, seules six personnes y étaient affectées au département chargé des questions de santé ; peut-être les effectifs ont-ils été renforcés depuis, mais cela donne une idée de l'inadéquation entre l'offre et la demande.

Les données ont été collectées par la puissance publique ; quel est le coût de leur collecte, de leur stockage, de leur sécurisation ? Le connaître permettrait de fonder le financement du dispositif par le biais d'une redevance.

Les comparaisons, notamment européennes, nous renseignent-elles sur l'existence de modèles économiques efficaces et efficaces ?

Dans ce contexte, quelles sont les données qui vous paraîtraient économiquement « rentables », toutes les précautions nécessaires étant naturellement prises pour garantir leur confidentialité et leur anonymisation ?

M. Philippe Burnel. Vaste et importante question !

D'abord, le fait d'ouvrir l'accès aux données, donc de permettre le renforcement de l'expertise privée, requiert probablement le renforcement de l'expertise publique. En d'autres termes, si l'on souhaite un dialogue – par exemple sur l'efficacité de telle stratégie thérapeutique, qui influe sur son taux de remboursement –, il est bon que les industriels puissent mieux documenter le dossier, mais il est indispensable que la partie non pas adverse, mais complémentaire, se dote elle aussi d'une expertise. Voilà pour le volet qui concerne l'exploitation des statistiques.

En ce qui concerne les moyens à mettre en œuvre, je ne saurais vous indiquer un coût : il convient que vous interrogiez directement l'assurance maladie. Dans la mesure où le SNIIRAM existait avant même d'être ouvert, il faudrait pouvoir faire la part de son coût imputable aux besoins de l'assurance maladie et des coûts marginaux induits par son ouverture. La Cour des comptes n'est pas parvenue à le faire.

M. le coprésident Pierre Morange, rapporteur. Vous ne le pouvez pas non plus ?

M. Philippe Burnel. Non : au poste où je suis, cela ne relève pas de ma mission – ce qui ne veut pas dire que l’administration en général ne peut pas vous répondre.

Les coûts marginaux dont je parle résultent de plusieurs éléments. D’abord, figurent toutes les démarches nécessaires à l’ouverture des données – création de magasins de données *datamarts*, éventuellement développement de capacités machines isolées, créations d’accès réseau particuliers –, bref tout ce qui relève de l’infrastructure. Ensuite, il faut y inclure l’ensemble des individus qui y contribuent. Enfin, la surveillance rendue nécessaire par l’ouverture, c’est-à-dire le renforcement du contrôle des accès par l’assurance maladie doit être comptabilisée : il est inutile de tracer des accès s’il n’y a personne pour les lire ; et si l’on autorise un acteur privé à créer une bulle de sécurité, il faut pouvoir l’auditer, par exemple pour vérifier que la bulle fonctionne dans les conditions annoncées et qu’il ne l’a pas réouverte au bout de six mois d’existence. Tout cela entraîne des surcoûts.

La loi ne dit pas que l’accès aux données sera payant mais ne dit pas non plus qu’il sera gratuit : la question n’est pas tranchée. Dans l’état actuel de la réflexion – mais je vous renvoie une fois encore à mon collègue Franck Von Lennep sur ce point –, l’accès serait gratuit pour les acteurs publics, pour éviter des complications inutiles puisque les faire payer reviendrait à ce qu’ils se paient eux-mêmes ; en revanche, il ne me paraîtrait pas illégitime que les acteurs privés contribuent aux coûts marginaux de l’ouverture.

Des gains de productivité sont par ailleurs envisageables. Le nombre réduit de personnels dédiés au sein de la CNIL, qui correspond bien à l’ordre de grandeur que vous indiquez, pose un véritable problème ; on le voit en ce moment, où ils ont à étudier beaucoup de textes issus de la loi de modernisation de notre système de santé tout en devant instruire des dossiers de demande d’autorisation individuelle. Parmi les solutions que la CNIL apporte à ce problème figure le recours à des méthodologies standard définies une fois pour toutes pour un type de données, assorti d’engagements de conformité, qui évite un examen individuel de chaque dossier.

On peut aussi – mais la démarche est très complexe du point de vue méthodologique, ce que Dominique Blum pourra vous expliquer plus savamment que moi – mettre à disposition des sous-échantillons de données, ce qui amoindrit considérablement le risque de ré-identification. Les conditions d’accès sont alors plus ouvertes, donc moins coûteuses en instruction.

Au total, il existe différentes méthodes pour gagner en productivité et réduire ainsi les surcoûts ; mais il est clair que l’ouverture représente un coût.

M. le coprésident Pierre Morange, rapporteur. Merci pour vos réponses précises et exhaustives qui vont enrichir notre réflexion. Peut-être aurons-nous à vous solliciter de nouveau, puisque le calendrier électoral nous conduira à procéder en deux temps : nous rendrons d’abord un pré-rapport, qui sera complété par un rapport définitif, livré à la fin de l’année-charnière 2017, voire début 2018, selon, notamment, l’état d’avancement des décrets d’application en cours de rédaction, qui nous permettront d’être plus précis dans nos préconisations. Sans préjuger des résultats de l’expression démocratique, cela témoigne de la permanence de l’activité parlementaire, en particulier celle de la MECSS.



Audition du Dr Dominique Blum, praticien hospitalier

La Mission d'évaluation et de contrôle des lois de financement de la sécurité sociale (MECSS) procède ensuite à l'audition, ouverte à la presse, du Dr Dominique Blum, praticien hospitalier, sur les données médicales personnelles inter-régimes détenues par l'assurance maladie, versées au Système national d'information inter-régimes de l'Assurance maladie (SNIIRAM) puis au Système national des données de santé (SNDS).

M. le coprésident Pierre Morange, rapporteur. Monsieur Dominique Blum, je vous souhaite la bienvenue. Praticien hospitalier, vos recherches sur les données du programme de médicalisation des systèmes d'information (PMSI) donneront à vos analyses une portée particulière, qu'elles portent sur la sécurisation, l'accessibilité, l'anonymisation ou le ré-identification et le chaînage, sujets auxquels vous attachez une attention soutenue.

M. Dominique Blum, praticien hospitalier. J'ai répondu à presque tous les éléments du questionnaire qui m'a été adressé, en m'inspirant, pour les sujets pour lesquels je ne suis pas le plus compétent, du legs d'André Loth. Il a précédé Philippe Burnel à la « mission PMSI » qui a mis le PMSI sur les rails dans les années 1990, et il fut le premier rédacteur de l'article 193 de la loi de modernisation de notre système de santé du 26 janvier 2016. Je fais la part, dans le document, de mes réponses personnelles et de celles qu'André Loth m'a inspirées et auxquelles j'adhère.

J'appartenais à la mission PMSI depuis le milieu des années 1990 et je me suis intéressé aux problèmes de l'anonymat et de la ré-identification que soulevait ce programme. Le résumé de sortie anonymisé (RSA) à l'hôpital est une de mes inventions. Elle est due à ce que la Commission nationale de l'informatique et des libertés (CNIL) nous avait informé que le résumé de sortie standardisé, que tous les hôpitaux et toutes les cliniques produisaient, avait un caractère indirectement nominatif. J'ai alors proposé un dispositif d'appauvrissement du contenu qui a consisté à remplacer la date de naissance du patient par son âge et les dates d'entrée et de sortie d'hospitalisation par le mois de sortie et la durée du séjour, et quelques autres modifications de cette nature.

Nous étions persuadés, à l'époque, que les bases étaient ainsi véritablement anonymisées. Mais, en 1997, pensant qu'il fallait être précautionneux, j'ai repris la base nationale et je me suis rendu compte qu'en combinant les informations, même appauvries, on aboutissait à des combinaisons uniques dans la base de données nationales. La note interne que j'ai alors publiée a amené le ministère à consulter la CNIL, laquelle a demandé que l'on procède à des adaptations pour les informations placées dans les fichiers distribués à l'extérieur – à l'époque, à la demande des journalistes en particulier, les bases de données étaient diffusées sous forme de cédérom.

M. le coprésident Pierre Morange, rapporteur. Quand était-ce ?

M. Dominique Blum. En 1996.

À partir de 2002, je me suis éloigné du PMSI jusqu'en 2009, année de mon retour à la direction de la recherche, des études, de l'évaluation et des statistiques (DREES). J'ai alors constaté que le résumé de sortie anonymisé avait été extraordinairement enrichi et contenait de très nombreuses informations supplémentaires mais, par ailleurs, qu'aucune disposition n'avait été prise pour éviter les risques de ré-identification. Aussi ai-je conduit, en 2010, de nouveaux travaux, cette fois sur les données collectées en 2008, et j'ai démontré que si l'on

connaît quelqu'un – j'entends par cela que si l'on connaît son âge, son sexe, le code postal de son lieu de résidence et la date de son entrée dans un établissement donné, ce qui est le cas d'un employeur, par exemple – on a 89 % de chances de le retrouver dans la base nationale, pourtant considérée comme anonyme ; la proportion passe à 100 % si la personne concernée a été hospitalisée au moins deux fois dans l'année. En effet, les combinaisons de données de la base nationale sont uniques si l'on dispose des informations que j'ai citées et quand on sait si le patient est sorti de l'établissement vivant ou mort.

J'avais mis en évidence, en 1998, que le mode de sortie « décès » est un élément extrêmement discriminant ; aussi la CNIL avait-elle demandé que cette information soit supprimée des fichiers. Malgré cela, les éléments contenus dans le résumé de sortie anonymisé sont restés tout aussi ciblant. D'ailleurs, ayant renouvelé l'étude en 2015 sur les données collectées en 2013, je suis parvenu à des conclusions identiques : si on les connaît, neuf patients sur dix peuvent être ré-identifiés à coup sûr, et tous peuvent l'être s'ils ont été hospitalisés au moins deux fois dans l'année. De plus, le programme est un chaînage longitudinal et transversal, géographique et chronologique ; il résulte de ces caractéristiques que, quelle que soit la date de son hospitalisation depuis 2006 et quels que soient le secteur et l'établissement dans lequel le patient a été hospitalisé, on peut retrouver tous ces éléments si l'on en a trouvé un seul.

J'ai appelé l'attention des pouvoirs publics à ce sujet en 2011, car je tiens à ce que la base de données puisse continuer d'être utilisée dans de bonnes conditions selon les finalités pour lesquelles elle a été créée, et aussi que les chercheurs puissent continuer à travailler sur ces données. Or chacun comprend que si une fuite se produisait qui permettrait la divulgation d'éléments de santé d'une célébrité ou, pire encore peut-être, d'un citoyen quelconque qui aurait été identifié à son insu par son employeur ou par son assureur, et que cette intrusion était révélée, le grand public remettrait en cause le principe même de la base de données.

M. le coprésident Pierre Morange, rapporteur. On peut aussi s'inquiéter de la possible marchandisation des données de santé au bénéfice des assureurs, étant donné les avantages qu'il y aurait pour eux à pouvoir croiser ces métadonnées et les informations issues du séquençage génomique utilisées aux fins de médecine préventive. La possibilité de fuites renforcerait le risque de contrats d'assurance santé personnalisés sur le fondement de ces éléments, ce qui réduirait à néant le principe de mutualisation, pierre angulaire de notre système de protection sociale.

M. Dominique Blum. En effet. On entend souvent dire que tout cela n'est pas grave au regard du nombre d'informations personnelles que chacun stocke dans ses appareils électroniques connectés. Certes. Mais outre que la qualité du recueil des informations par les appareils connectés n'est pas encore celle d'un simple tensiomètre, la différence fondamentale est que la mise à disposition des informations personnelles dans les objets connectés est volontaire alors que, pour le PMSI ou le Système national d'information inter-régimes de l'Assurance maladie (SNIIRAM), elle est obligatoire, systématique et exhaustive : personne ne passe entre les gouttes. C'est pourquoi la question est particulièrement sensible.

Il m'a souvent été opposé au fil des ans par des confrères médecins de l'information médicale qu'il y avait peut-être un risque mais que, jusqu'alors, la preuve n'était pas apportée qu'il se soit concrétisé.

M. le coprésident Pierre Morange, rapporteur. Raisonnement indigent !

M. Dominique Blum. Il peut sembler l'être, d'autant que si des fuites se produisaient, elles résulteraient d'intrusions illégales dont on imagine que ceux qui s'y livreraient ne se vanteraient pas. Cela dit, vous avez sans doute entendu parler de ces pirates qui mettent au point des logiciels informatiques malveillants dits *ransomware*.

M. le coprésident Pierre Morange. Oui, ils prennent des données « en otage » et demandent des rançons de plusieurs dizaines de milliers de dollars.

M. Dominique Blum. Je souhaitais donc que l'on anticipe ce risque en prenant des dispositions permettant de pérenniser dans de bonnes conditions l'existence de la base de données, et je considère que les dispositions de la loi de janvier 2016 relatives au système national des données de santé (SNDS) répondent parfaitement aux objectifs d'accessibilité aux personnes autorisées. Elles emportent certes des contraintes, mais si les dispositions prévues sont vécues comme telles par ma corporation, celle des médecins de l'information médicale, c'est en raison du laxisme qui prévalait auparavant, lorsque les bases de données du PMSI étaient distribuées sous la forme de cédéroms qui étaient reproduits en dépit des engagements formels de n'en rien faire, pris par ceux qui auraient dû en être les seuls dépositaires et qui, dans les colloques scientifiques, proposaient aimablement à leurs collègues de leur en donner copie. Il était nécessaire de revoir les procédures. J'estime que le nombre de cédéroms où figurait la base de données du PMSI était à peu près le double de ceux qui étaient officiellement distribués par l'Agence technique de l'information sur l'hospitalisation (ATIH). Un courriel de l'ATIH m'en a apporté la preuve : il était envoyé à tous ceux qui avaient reçu le cédérom l'année précédente, et la liste des destinataires était visible ; or dans cette liste ne figuraient pas certains détenteurs de la base de données que je connaissais.

Je pense que les précautions nécessaires ont été prises suffisamment tôt. Cela répond d'ailleurs à la question de savoir pourquoi les bases de données ont été peu utilisées par ceux qui y avaient accès : que le PMSI soit assez mal documenté et que le SNIIRAM soit très difficile d'accès les a protégées.

M. le coprésident Pierre Morange, rapporteur. Ce n'est donc ni par inertie ni par indifférence ?

M. Dominique Blum. Non. Je parle des usages autres que les usages par les services de l'État ou par les médecins. À peu près tous les médecins de l'information médicale avaient accès à la base de données. C'est eux qui produisent l'information, mais pour analyser les données et faire les comparaisons pertinentes, ils en avaient besoin. La plupart d'entre eux l'obtenaient *via* des agences régionales de santé, ce qui était illégal, mais pour un usage qui, à ma connaissance, n'a jamais donné lieu à dérapage.

M. le coprésident Pierre Morange, rapporteur. Vous voyez dans le chaînage un risque évident de ré-identification ; pour sa part, la Cour des comptes juge « obsolète » l'algorithme de cryptage utilisé pour occulter les identifiants nominatifs dans le SNIIRAM. Comment recommandez-vous d'assurer la sécurité du dispositif ? En une période où l'évolution technologique est extraordinairement rapide et où les pirates informatiques s'en prennent même à des systèmes hyper-sécurisés, la multiplicité des portes d'entrée n'accroît-elle pas les risques de divulgation de données confidentielles ou de détournement de fonds publics ?

M. Dominique Blum. Je n'ai pas envisagé l'hypothèse d'une situation aussi catastrophique.

M. le coprésident Pierre Morange, rapporteur. Mieux vaut la prévenir !

M. Dominique Blum. J'ai considéré qu'il fallait confiner les données dans des espaces confiés à des spécialistes de la sécurité ; c'est ce que j'ai recommandé. Les pirates informatiques qui se cachent derrière le nom de Rex Mundi ont certes piraté Labio, mais ce laboratoire ne respectait pas les règles de sécurité ; je crois qu'il a d'ailleurs été poursuivi pour cette raison. Confier la base de données à des spécialistes de la sécurité et créer, comme le dit la loi, un dispositif prévoyant des habilitations pour l'accès à ces informations et un système de traçage et de surveillance de tous les accès, c'est viser la sécurité maximale.

M. le coprésident Pierre Morange, rapporteur. Selon vous, la professionnalisation du dispositif et la maîtrise et le filtrage des accès sont déjà possibles ?

M. Dominique Blum. Je le pense. Le Centre d'accès sécurisé aux données (CASD), hébergé par l'Institut national de la statistique et des études économiques (Insee) à Paris, organise l'accès à des données individuelles détaillées dans d'autres domaines que la santé. Les données fiscales, apparemment considérées comme encore plus sensibles que les données médicales par ceux qui les mettent à disposition de leurs utilisateurs respectifs – ce que je regrette – sont par ce biais rendues accessibles aux chercheurs de France et de pays étrangers, dans un cadre de sécurité qui a fait la preuve de son efficacité. Le référentiel de sécurité visé dans le projet de décret pour l'accès aux données du SNDS est actuellement analysé par la CNIL. Ma recommandation est de ne pas multiplier les points d'accès et de confier la sécurité du dispositif à des spécialistes. De ce point de vue, je reste très critique sur la responsabilité de l'ATIH qui, lors de la création du PMSI, n'avait pas mesuré la nécessité de sécuriser les informations collectées, car elle n'en avait ni les compétences ni la culture. S'en serait-elle préoccupée que certaines informations transmises à la CNIL n'auraient pas été un peu biaisées et que des dispositions auraient été prises depuis longtemps pour assurer la sécurité du système.

M. le coprésident Pierre Morange, rapporteur. Selon vous, quelle pourrait être la structure pertinente pour assurer la sécurité informatique ? Visez-vous un organisme en particulier ou un service dont il faudrait étoffer les ressources humaines ? Nous restons sur notre faim, car la première des obsessions est la qualité du coffre-fort informatique.

M. Dominique Blum. Il existe dans les ministères des services spécialisés dans la sécurité informatique ainsi que des sociétés privées. En tout cas, il faut confier cette tâche à des opérateurs dont c'est le cœur de métier. Des dispositifs bricolés nous conduiraient aux mêmes résultats qu'auparavant avec, en outre, un risque d'éparpillement. Si les données sont disséminées entre le PMSI, le SNIIRAM, le Centre d'épidémiologie sur les causes médicales de décès (CépiDc), les pouvoirs publics comme les chercheurs éprouveront les plus grandes difficultés à effectuer des rapprochements, qui s'avèrent nécessaires pour certains travaux.

Prenons l'exemple des médicaments hospitaliers : les données du PMSI ne contiennent aucune information à ce sujet mais elles sont versées au SNIIRAM, dont les données comprennent des informations sur les médicaments délivrés en ambulatoire. Mettre en regard les données du PMSI et du SNIIRAM se révèle utile pour rapprocher les éléments relatifs à la clinique hospitalière de ceux relatifs aux médicaments en ambulatoire.

M. le coprésident Pierre Morange, rapporteur. Qu'en est-il du risque lié au chaînage, dont vous remettez en cause le principe, si j'ai bien compris ? Ne peut-on craindre une perte de substance dans l'exploitation de ces mêmes données ?

M. Dominique Blum. Je ne remets pas en cause le principe du chaînage mais le fait qu'il existe sans dire son nom. Il avait été indiqué à la CNIL que le chaînage permettait d'apparier les informations cliniques aux informations relatives à la facturation. Comme il n'y a pas de factures dans les hôpitaux, cela m'avait paru surprenant. J'ai donc approfondi la question et il est apparu que le chaînage, loin de ne concerner qu'un seul séjour, était de nature longitudinale, information qui n'avait pas été communiquée à la CNIL.

En outre, de 2006 à juillet 2013, il contenait un champ non documenté, la datation des séjours. Si l'on comprend bien les motivations, légitimes, de ceux qui l'ont intégré – il permet d'éviter les superpositions et les saucissonnages de séjours, toutes stratégies permettant aux établissements hospitaliers de frauder et d'obtenir des doubles financements –, on comprend moins bien les modalités de sa mise en œuvre. On peut se protéger contre des éléments que l'on connaît mais pas contre des éléments que l'on ignore.

À cause de ce manque de transparence, pendant des années, les cédéroms que j'évoquais contenaient non seulement les données de la base du PMSI mais également les données de chaînage. Les personnes qui demandaient la communication desdits cédéroms, persuadées qu'ils leur permettaient d'apparier facturation et données cliniques, cochaient systématiquement dans les formulaires adressés à la CNIL et à l'ATIH la case « j'ai besoin du chaînage ». Or ce fichier de chaînage ne doit absolument pas être diffusé dans le cadre de l'accès aux *open data*.

Je fais partie de ceux qui estiment qu'il est tout à fait possible, en procédant à des appauvrissements techniquement très pointus des bases de données du PMSI, d'en extraire des données qui conservent la même granularité individuelle mais qui sont totalement anonymisées afin d'empêcher toute ré-identification.

M. le coprésident Pierre Morange, rapporteur. Ces techniques nécessitent de mobiliser des moyens humains et financiers, mais elles auraient l'avantage de permettre des économies à moyen et long terme, car les services concernés pourraient alors concentrer leurs efforts de surveillance sur des sujets plus sensibles.

M. Dominique Blum. C'est un sujet très controversé parmi les experts. Il y a des cas pour lesquels il est compliqué de produire de telles bases de données, surtout des bases qui présentent encore un intérêt. Parmi les techniques qui s'affrontent, certaines aboutissent à une forte déperdition de l'information utile.

Prenons l'exemple du palmarès des hôpitaux, que Fabrice Malye et son équipe ont fait paraître à l'origine dans *Science et Avenir*, puis dans *L'Express* et aujourd'hui dans *Le Point*. La méthodologie qu'ils ont adoptée permettrait tout à fait de recourir à une base de données qui, tout en étant exempte de risques de ré-identification, aurait la même granularité et serait aussi précise s'agissant des contenus sensibles – diagnostics, actes, médicaments. Il est tout à fait excessif de lui remettre un fichier pour lequel il prend lui-même le risque d'une fuite.

M. le coprésident Pierre Morange, rapporteur. Y a-t-il un consensus sur le degré de précision ?

M. Dominique Blum. La question n'est pas tant le degré de précision que la possibilité de produire des données ayant les caractéristiques que je viens de décrire.

Tout le monde s'accorde sur le fait qu'il est possible de produire pour l'*open data* des données agrégées, qui fourniront, par exemple, la distribution des diagnostics au sein de groupes. La DREES, l'ATIH et d'autres organismes comme la Fédération nationale des observatoires de santé (FNORS) produisent de telles données. Toutefois, lorsqu'on les propose aux journalistes, ils poussent de hauts cris, car ils estiment que l'information ne leur parvient que déjà filtrée et qu'on leur cache des choses. Le véritable enjeu, ce sont les bases de données qui conservent la granularité individuelle.

M. le coprésident Pierre Morange, rapporteur. Les méthodes que vous préconisez comme permettant d'évacuer le risque des ré-identifications font-elles consensus auprès de l'État et de l'assureur obligatoire ?

M. Dominique Blum. Pas encore, car c'est un sujet qui reste très discuté au niveau international. Les techniques sont complexes et la France n'est pas très en pointe en ce domaine, il faut bien le dire. C'est un regret que j'ai personnellement. J'essaie de monter un groupe de recherche et je cherche des appuis pour pouvoir accéder aux données dans des conditions normales, c'est-à-dire sans avoir à dépenser de l'argent ; car aujourd'hui, il faut payer.

M. le coprésident Pierre Morange, rapporteur. Quelle est votre expérience des bases de données européennes ou américaines ? Quelle exploitation en est faite ? Quels en sont les bénéfices pour la santé publique ? Quels modèles économiques soutiennent la sécurisation des dispositifs ?

M. Dominique Blum. Pour vous répondre, je m'inspirerai des réflexions d'André Loth. De l'analyse des expériences étrangères, il a tiré la conclusion que la France était le seul pays à disposer, avec le SNIIRAM et le PMSI, de bases de données exhaustives couvrant l'ensemble du système hospitalier et l'ensemble du système ambulatoire. Ailleurs, les bases de données sont parcellaires. Aux États-Unis, par exemple, elles couvrent seulement les bénéficiaires du dispositif Medicare ou bien les anciens combattants.

Je n'ai pas connaissance de travaux portant sur le modèle économique. Les études qui ont particulièrement retenu mon attention sont celles qui concernent la capacité à produire des bases en *open data*, je pense en particulier aux analyses menées au Massachusetts Institute of Technology (MIT).

Les débats sur les bases en *open data* ne sont clos nulle part. Il faut rassembler de multiples avancées et les adapter à nos besoins. La question est très souvent abordée de manière assez théorique, comme si la donnée de santé se résumait à un seul diagnostic. Or la particularité des dossiers médicaux, c'est de comporter non pas un diagnostic mais plusieurs diagnostics. Les questions statistiques que cela pose sont beaucoup plus pointues que celles qui sont exposées dans les présentations traditionnelles.

M. le coprésident Pierre Morange, rapporteur. Quel est votre avis sur la réglementation européenne sur la protection des données ?

M. Dominique Blum. Pour tout vous dire, je suis un peu surpris par certains points, notamment par l'emploi du terme « raisonnablement » qui revient dans beaucoup de textes officiels. Il est ainsi indiqué que le niveau de protection doit être adapté en fonction des

moyens qu'il est raisonnablement possible de mettre en œuvre pour essayer d'attaquer la base de données. Je ne pense pas que cet élément puisse réduire les risques pesant sur les données du PMSI et du SNIIRAM ou influencer sur la manière de les évaluer.

En revanche, il y a une approche intéressante concernant la responsabilisation des personnes qui conduisent des projets de recherche sur les données. Chaque porteur de projet devra lui-même procéder à une évaluation des risques et déterminer les mesures à prendre au sein de son organisme d'appartenance pour que le traitement des données se déroule dans des conditions de sécurité satisfaisantes. Il est ensuite prévu qu'un audit contrôle leur bonne mise en œuvre. C'est là une approche assez différente de celle qui prévaut actuellement, mais j'ai cru comprendre que le référentiel de sécurité en cours d'analyse à la CNIL tendrait vers un tel dispositif. Cela n'a rien d'étonnant puisque le règlement européen entrera en vigueur en 2018.

M. le coprésident Pierre Morange, rapporteur. Est-il possible de procéder à des catégorisations sociales d'assurés ou de patients dans les données agrégées ?

M. Dominique Blum. À ma connaissance, elles ne comportent pas d'informations sur les catégories socio-professionnelles (CSP) des individus, sauf certains travaux de l'INSERM portant sur des cohortes.

M. le coprésident Pierre Morange, rapporteur. On pourrait imaginer que des assureurs fassent une exploitation commerciale de telles données. C'est de la science-fiction sanitaire, je vous l'accorde, mais la tendance naturelle des entreprises à faire du profit pourrait laisser envisager cette éventualité.

M. Dominique Blum. La CSP n'est pas intégrée dans les bases de données, comme je vous le disais. Au niveau organisationnel, nous nous sommes prémunis d'un tel risque en faisant en sorte que les organismes d'assurance privés aient recours à des tiers de confiance qui interviennent pour eux.

M. le coprésident Pierre Morange, rapporteur. Oui, mais qu'en est-il des données concernant les bénéficiaires de la protection universelle maladie, de la couverture maladie universelle complémentaire (CMU-C) ou encore de l'aide au paiement d'une complémentaire santé (ACS), toutes prestations conditionnées à un niveau de revenus ? L'enjeu n'est pas mince : rappelons qu'elles sont versées à quelque 6,5 millions de personnes.

M. Dominique Blum. La loi du 26 janvier 2016, qui s'appliquera, je l'espère, bientôt, prévoit que les études menées sur les bases de données devront au préalable faire l'objet d'une validation par un organe *ad hoc* de l'Institut national des données de santé. L'utilisation aux fins que vous évoquez devrait donc être prohibée.

M. le coprésident Pierre Morange, rapporteur. Peut-on envisager de remplacer le contrôle budgétaire des dépenses de santé par masses comptables par un contrôle statistique des actes ?

M. Dominique Blum. Je dois dire que ce n'est pas du tout mon domaine.

M. le coprésident Pierre Morange, rapporteur. J'ai bien conscience que cette question se situe à la périphérie de votre domaine de recherche, mais nous aimerions tout de même savoir si, selon vous, un changement dans l'exploitation des données peut aboutir à envisager une nouvelle manière de gérer les moyens consacrés au système de soins. Au lieu d'une approche macroéconomique, il s'agirait de rechercher, par une approche plus fine et

plus individualisée, un meilleur rapport coût-efficacité au service de chacun de nos concitoyens.

M. Dominique Blum. Les deux approches sont complémentaires. L'approche macro peut nourrir l'approche micro et vice-versa. Toujours est-il que les données PMSI dont nous disposons sont d'une précision telle qu'elles permettraient de travailler finement sur les actes de soins.

M. le coprésident Pierre Morange, rapporteur. Pour en revenir à un univers qui vous est plus familier, pourriez-vous nous décrire les relations entre les communautés de chercheurs et les instances administratives autorisant les recherches sur les bases de données ? Avez-vous constaté que des barrières s'étaient élevées entre eux ?

M. Dominique Blum. J'ai plutôt l'impression que les chercheurs qui le souhaitent parviennent assez facilement à accéder à ces données. Ceux de l'Institut de recherche et de documentation en économie de la santé (IRDES) produisent depuis longtemps des travaux à partir des bases du PMSI ou du SNIIRAM. Les chercheurs de l'INSERM, qui exploitent les données du SNIIRAM pour leurs études de cohortes, sont même en avance sur beaucoup d'autres et ont une maîtrise de cette base de données, qui nécessite un investissement humain très fort.

En revanche, le système vers lequel on se dirige a rendu plus complexe la situation pour les médecins spécialistes de l'information médicale. Plus de 200 d'entre eux ont signé cet été une pétition pour réclamer un accès simplifié aux données, dont eux-mêmes sont pourtant à l'origine. Les médecins attachés aux départements d'information médicale (DIM) ne disposent pas d'un cadre législatif leur permettant d'accéder aux données à l'échelle du territoire alors que la nouvelle loi créant les groupements hospitaliers de territoire (GHT) les met dans une position de pivot. Ils ne parviennent pas à obtenir dans des délais satisfaisants les autorisations pour accéder aux données, quand ils ne se heurtent pas à des refus.

Il y a une incompréhension totale de la part de mes confrères et c'est à moi qu'ils adressent des reproches, car ils estiment que c'est depuis que j'ai appelé l'attention sur ce sujet qu'ils ne peuvent plus accéder aux données ! Je leur réponds qu'avant, ils y avaient seulement accès, illégalement, par le biais des agences régionales de santé.

M. le coprésident Pierre Morange, rapporteur. Quelles mesures concrètes envisagez-vous pour mettre fin à cette situation ?

M. Dominique Blum. Il faudrait que la CNIL traite plus rapidement leurs demandes et surtout que les décrets d'application soient publiés au plus vite. Comme l'existence de risques est désormais officielle, chaque organisme – et c'est compréhensible – a voulu prendre le maximum de précautions et anticiper le renforcement de la sécurisation des données.

M. le coprésident Pierre Morange, rapporteur. Cette tétanisation des esprits se traduit par des délais de quelle longueur ?

M. Dominique Blum. Ils sont supérieurs à douze mois.

M. le coprésident Pierre Morange, rapporteur. Merci pour la précision de vos réponses et la qualité de vos travaux.

La séance est levée à douze heures cinq.