

Septembre 2013

RAPPORT SUR LA GOUVERNANCE ET L'UTILISATION DES DONNÉES DE SANTÉ

***Pierre-Louis BRAS, inspecteur général des affaires sociales
avec le concours d'André LOTH, administrateur général, directeur de
projet à la DREES***



MINISTÈRE DES AFFAIRES SOCIALES ET DE LA SANTÉ

La Ministre

Cab/MT/BM/MGD/D.13-3306

Paris, le 16 AVR 2013

Monsieur l'Inspecteur général,

L'utilisation des données de santé est un élément fondamental du pilotage de notre système de santé. Les débats récents ont révélé la nécessité de définir un cadre adapté à l'ensemble des enjeux liés à ces données, qu'il s'agisse de la sécurité sanitaire, de l'épidémiologie ou de la compréhension du fonctionnement de notre système de santé.

S'il importe de développer l'accès du grand public aux données de santé qui lui sont nécessaires pour s'informer sur le système de santé, l'offre de soins, les parcours de soins et les produits de santé, s'il convient d'étudier les conditions dans lesquelles les industriels pourraient utiliser ces données pour réaliser des études post AMM, il est tout autant essentiel pour les pouvoirs publics de garantir le secret médical et la protection des données personnelles dans l'utilisation des bases de données médico-administratives.

De plus, la dispersion de la gouvernance des données de santé, de leur accès et de leur utilisation, issue de décisions successives, rend difficile l'appréhension globale des principes d'une politique des données de santé.

Je vous demande de me proposer, dans le cadre de la mission que je vous confie, les voies et moyens de la mise en place d'un dispositif assurant à la fois une gouvernance d'ensemble des données de santé et garantissant leur mise à disposition dans des conditions adaptées aux finalités poursuivies par les différents acteurs.

Monsieur Pierre-Louis BRAS
Inspecteur Général des Affaires Sociales
39-43, quai André Citroën
75739 PARIS CEDEX 15

.../...

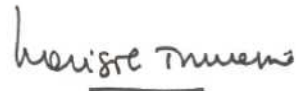
Vous vous attacherez en particulier à :

- proposer aux pouvoirs publics une organisation efficiente du pilotage des données de santé à des fins de santé publique et de sécurité sanitaire, garantissant une bonne articulation avec les institutions responsables des bases de données médico-administratives ;
- étudier les conditions fiables et sécurisées de la mise en place d'un dispositif d'accès et d'utilisation des bases de données médico-administratives adapté à la nature et à la réalité des besoins des différents acteurs.

Vous bénéficierez, pour mener à bien vos travaux, de l'appui de la direction de la sécurité sociale, de la direction générale de la santé et de la direction de la recherche, des études, de l'évaluation et des statistiques. Il vous reviendra de consulter l'ensemble des acteurs et particulièrement les représentants des patients, de la société civile, des professionnels et des établissements de santé, des industriels des produits de santé, des sociétés de service et de conseil en santé ainsi que les institutions responsables des bases de données de santé.

Vous vous rapprocherez également des Professeurs Bégaud et Costagliola, auxquels j'ai confié une mission spécifique sur la pharmaco-surveillance, ainsi que du secrétaire général de la modernisation de l'action publique, afin d'articuler votre réflexion avec la politique d'ouverture des données publiques animée par la mission Etalab.

Vos conclusions me seront remises sous trois mois.



Marisol TOURAINE

Sommaire

Synthèse	7
Introduction.....	9
Chapitre 1 Le constat.....	11
1 La base centrale pour les données de santé individuelles : le SNIIRAM	11
1.1 Des objectifs initiaux limités.....	12
1.2 Le contenu et l'organisation du SI	14
1.3 La richesse et les limitations du SI.....	21
2 Les accès au SNIIRAM	26
2.1 Les risques de réidentification	26
2.2 Les accès au SI : état des lieux et difficultés.....	31
2.3 Les difficultés particulières des appariements de données et le problème du NIR.....	38
2.4 Une complexité et des délais excessifs	40
Chapitre 2 Les propositions.....	43
1 Les enjeux de l'accès aux données de santé	43
1.1 Les bénéfices potentiels	43
1.2 Le mésusage éventuel des données	45
2 Les principes	47
2.1 Les finalités du SI	47
2.2 La propriété de la base	48
2.3 Les conditions d'accès au SI	49
2.4 La protection des données relatives aux professionnels de santé.....	50
3 Les actions à conduire	56
3.1 Permettre un accès public à des lots de données qui ne présentent pas de risques d'atteinte à la vie privée.....	56
3.2 Organiser des accès contrôlés pour exploiter pleinement le potentiel actuel et futur, en termes sanitaire et économique	60
3.3 Rendre le SI plus accessible	69
4 Le dispositif économique et institutionnel.....	72
4.1 Quel modèle économique ?	72
4.2 Quelle organisation pour gérer la base ?	73
4.3 Quelle gouvernance ?	76
ANNEXES.....	78

Annexe 1 – Liste des personnalités rencontrées.....	79
Annexe 2 : Glossaire	81
Annexe 3 : Des liens utiles.....	83
Annexe 4 : Principales sources de données en santé.....	84
Annexe 5 : Rapport OCDE – Strengthening health information infrastructure for Health Care quality governance, <i>Good Practices, New Opportunities and Data Privacy Protection Challenges</i>	86
Des améliorations dans les systèmes de santé et des évolutions dans la gouvernance engagées dans la majorité des pays depuis 5 ans	86
La France bien positionnée pour davantage capitaliser sur son patrimoine de données en santé	87
Annexe 6 : Panorama des principaux dispositifs existants à l'étranger	90
Annexe 6.1 : « Health & Social Care Information Centre » du NHS	90
Annexe 6.2 : Système QOF (Quality Outcome Framework) du NHS	92
Annexe 7 : Gestion des données à caractère personnel - Comparaison des systèmes d'information sanitaire européens et canadien (Congrès ADEL-EMOIS 2012)	98
Annexe 8 : Loi Informatique et Libertés, extraits.....	110
Annexe 9 : Extraits du Code de la sécurité sociale	118
Annexe 10 : Extraits du Code de la santé publique	120
Annexe 11 : « Arrêté SNIIRAM » du 19 juillet 2013	122

Synthèse

La première partie du rapport est un constat

Elle rappelle en premier lieu (1.1) les finalités initiales des bases SNIIRAM et PMSI, décrit leur réunion en une grande base -appelée *Système d'Information (SI)* dans la suite du rapport-, détaille son contenu et son organisation (1.2), sa richesse exceptionnelle et ses limitations (1.3).

Elle expose en second lieu les difficultés actuelles, juridiques et pratiques, d'accès aux données :

- comment, dans une base supposée anonyme, il peut exister un risque de ré-identification indirecte des personnes (2.1) ;
- ce qu'est le régime actuel des droits d'accès aux SI, ce que sont ses origines, sa complexité et ses défauts (2.2) ;
- pourquoi les contraintes actuelles sur l'emploi du NIR (numéro de sécurité sociale) protègent moins les données personnelles qu'elles n'entravent les recherches (2.3) ;
- pourquoi enfin la difficulté d'accéder aux données n'est pas seulement juridique mais liée aussi à leur complexité et à l'insuffisance des services de mise à disposition et de mise en forme des données (2.4).

La seconde partie du rapport présente les enjeux et les propositions

Les enjeux (1) sont d'une part le risque de ré-identification et le risque de mésusage et d'autre part les grands bénéfices, démocratiques, sanitaires et économiques dont est porteur le SI.

Au terme d'une revue critique des arguments, les principes proposés (2) sont les suivants :

- *Le SI a des finalités très larges et constitue donc un bien public qui ne peut être approprié par aucun des acteurs du système. Il doit être administré dans l'intérêt commun par une autorité légitime en concertation avec l'ensemble des parties prenantes ;*
- *Dès lors que les données présentent un risque de réidentification des patients, l'accès doit en être restreint ;*
- *Les lots de données qui ne présentent pas un tel risque de réidentification doivent être communicables à tous ou rendus publics ;*
- *La diffusion des tarifs moyens des professionnels de santé ne devra plus être réservée à l'assurance maladie mais le débat sur la transparence des données nominatives relatives à l'activité des professionnels mérite d'être ouvert dans un cadre concerté avec les intéressés.*

Pour mettre en œuvre ces principes, il est proposé (3) :

- De distinguer autant que possible les lots de données clairement anonymes des lots de données indirectement nominatifs. A partir d'une expertise publique sur les risques de réidentification, les lots de données qui peuvent sans risques être communiqués ou rendus publics seront définis sous le contrôle de la CNIL ;
- D'ouvrir l'accès aux lots de données anonymes en distinguant la publication (gratuite) et des extractions ou des tableaux de bord à façon (payants) ;
- De limiter l'accès aux données indirectement nominatives du SI, avec pour critères la finalité d'intérêt public, la qualité du protocole, le besoin d'accéder aux données, la sécurité des procédures et la qualité du demandeur :

- Les droits d'accès permanents seront accordés par le ministre de la santé sous le contrôle de la CNIL aux organismes publics, à partir d'une appréciation de l'équilibre entre les risques (réidentification) et les bénéfices (vigilance sanitaire, connaissance médico-économique) ;
- Les organismes de vigilance sanitaire (ANSM, INVS, HAS) doivent se donner les moyens d'exploiter pleinement le potentiel du SI ;
- Les droits d'accès ponctuels seront accordés, par un dispositif unique qui permettra de vérifier que le « bénéfice » collectif potentiel attendu de la recherche envisagée justifie la mise à disposition des données, avec des moyens accrus, pour un contrôle a priori des risques de mésusage ;
- Les appariements autorisés dans le cadre cette procédure ne nécessiteront pas de décret en Conseil d'État ;
- Les données seront hébergées dans des conditions sûres et les procédures d'accès seront sécurisées.
- Si ces orientations sont retenues, à l'issue de la concertation avec les parties prenantes, leur mise en œuvre impliquera des modifications législatives concernant notamment :
 - les dispositions relatives au SNIIRAM et au PMSI dans les codes de la sécurité sociale et de la santé ;
 - les dispositions de la loi Informatique et Libertés relatives à la recherche en santé, à l'analyse et à l'évaluation des activités de soins et de prévention ainsi qu'à l'emploi du NIR ;
- L'exploitation du potentiel exceptionnel de la base et le développement des usages, spécialement pour les chercheurs, passent par un plan d'urbanisation du SI et la mise en place d'une grande plate-forme de services aux utilisateurs ;
- Les besoins et les priorités pour l'élargissement du périmètre du SI seront déterminés en concertation avec les parties prenantes.

Enfin sont proposées (4) les grandes lignes d'un modèle économique et institutionnel :

- Pour couvrir les coûts de fonctionnement, les accès à la base sécurisée, les extractions et les travaux à façon donneront lieu à paiement par l'organisme commanditaire de l'étude. Cela s'appliquera aussi aux laboratoires pharmaceutiques ;
- Trois solutions institutionnelles alternatives sont proposées à la discussion pour l'organisation de la base et du guichet associé :
 - Une structure autonome,
 - Une structure adossée à la CNAMTS mais disposant de moyens dédiés,
 - Une structure adossée à la Direction statistique du ministère des affaires sociales, la DREES ;
- Une gouvernance qui relèvera en dernière instance des ministres chargés de la santé et de la sécurité sociale, s'appuyant sur un Haut Conseil des données de santé réunissant les parties prenantes et doté d'un Conseil scientifique.

Introduction

Par lettre du 16 avril 2013, la ministre des affaires sociales et de la santé a saisi l'auteur de ce rapport d'une mission sur la gouvernance et l'utilisation des données de santé. Le sujet est complexe. Il concerne surtout la plus importante des bases publiques de données de santé dans notre pays, voire dans le monde, le *Système national d'information inter-régime de l'assurance maladie* (SNIIRAM).

Issues des feuilles de soins et des résumés de sortie hospitaliers, dont on a retiré tous les éléments directement identifiants, les données du SNIIRAM décrivent l'offre et la consommation de soins ainsi que l'état de santé des quelque 65 millions d'habitants de notre pays dans la période récente.

Utiliser davantage ces données pour produire de la connaissance et informer le public d'une part, protéger leur confidentialité d'autre part : ces exigences sont par nature contradictoires ou difficiles à concilier. La contradiction se révèle d'autant plus aiguë que les données sont nombreuses, que les enjeux sanitaires, démocratiques, économiques sont importants mais que la gouvernance des données est éclatée ou que ses principes manquent de clarté.

Ce sont surtout les difficultés d'accès et la sous-utilisation des données qui ont été mises en cause récemment :

- Déjà les chercheurs en épidémiologie, sous l'égide du Pr. Goldberg, avaient dressé un tableau précis et critique des obstacles auxquels se heurtent les chercheurs, dans deux rapports du Haut Conseil de la santé publique¹ de 2009 et de 2012 ;
- L'Institut des données de santé a aussi plaidé pour davantage d'ouverture des données à la société civile ;
- La DREES a réuni à l'automne 2012 un groupe d'experts² pour identifier les obstacles juridiques et techniques à l'usage des données par les chercheurs et les administrations et y proposer des solutions ;
- Plus récemment, des tribunes dans la presse³ et une pétition de janvier 2013⁴ ont suscité un vif débat. « Libérez les données de santé » demandait la pétition qui, accusant l'assurance maladie et les pouvoirs publics de séquestrer les données du SNIIRAM, exigeait que soit appliqué le principe de libre réutilisation des données publiques⁵ (*open data*).

Cependant les acteurs s'accordent sur une limite : l'utilisation des bases constituées à partir de données de santé individuelles, pour des finalités collectives d'études ou de recherches, ne doit pas remettre en cause le droit des personnes au respect de leur vie privée, principe de valeur constitutionnelle⁶. Sans confiance des citoyens dans la protection des données confidentielles, il ne peut exister de système d'information en santé viable. Il doit donc être admis que si des données de santé sont directement ou indirectement nominatives, l'accès doit en être restreint et contrôlé.

¹ Voir en annexe 3 une « Liste des liens utiles ».

² Conduit par André LOTH (DREES), ce groupe était constitué de Marcel GOLDBERG (INSERM), Hélène CAILLOL (CNAMTS), Alain FONTAINE (DGS), Renaud LEGAL (DREES), Emmanuel CHION (DSS), Jérôme DUPONT (DGOS) et Frédérique POTHIER (SG-DSSIS).

³ Notamment la tribune publiée le 15 janvier 2013 dans *Le Monde* sous les signatures de Jean de Kervasdoué et Didier Sicard, ce dernier étant aussi président du Comité des experts de l'Institut des données de santé.

⁴ <http://www.opendatasante.com/>

⁵ Loi du 17 juillet 1978 sur l'accès aux documents administratifs (voir notamment les articles 10 et suivants ajoutés par l'ordonnance du 29 avril 2009).

⁶ 99-416 DC du 23 juillet 1999, loi portant création d'une couverture maladie universelle.

Précisons toutefois de quelles données de santé il s'agit : il n'est pas question ici ni de l'accès des personnes à leur dossier médical, ni des règles de partage des informations médicales entre professionnels de santé pour soigner un même malade, ni, non plus, des données nominatives qui servent à accorder ou refuser un remboursement à un assuré.

Les données dont il est question dans ce rapport sont des données destinées à la recherche, aux statistiques, à l'aide à la décision et à l'information du public. Elles permettent de mieux connaître le système de santé, pour mieux l'utiliser, pour en débattre démocratiquement et pour l'améliorer, et notamment :

- de mieux connaître les risques sanitaires, les effets des médicaments, l'état de santé de la population et ses déterminants, l'offre de soins, la qualité des soins et leurs coûts, l'évolution des dépenses...,
- de concevoir des réformes ou d'en évaluer les effets,
- et de développer la recherche en santé dans tous ces domaines.

Les réflexions et les polémiques ont mis en évidence le besoin d'un débat et d'une doctrine. L'exigence d'ouverture traduit aussi les progrès considérables déjà réalisés, par les équipes de la CNAMTS et de l'ATIH notamment, pour constituer et utiliser ces bases dont le potentiel avait été initialement sous-estimé.

La sécurité sanitaire, la démocratie sanitaire, la recherche en santé et le pilotage du système de santé bénéficieront du Système d'information national de santé ; plus largement, cette grande infrastructure contribuera au progrès économique du pays. Le présent rapport propose des principes pour sa gouvernance, son amélioration et son bon usage.

Chapitre 1 Le constat

1 La base centrale pour les données de santé individuelles : le SNIIRAM

Parmi de nombreuses sources de données sur la santé⁷, la France dispose de plusieurs bases administratives nationales à vocation exhaustive, comportant des données individuelles de santé :

- La principale base médico-administrative est le Système national d'information inter-régimes de l'assurance maladie (SNIIRAM). Géré par la Caisse nationale de l'assurance maladie (CNAMTS), il tire initialement ses données des « feuilles de soins⁸ », soient environ 1,2 milliard d'enregistrements par an, traitées puis retransmises par les caisses primaires du Régime Général et les autres régimes -ou organismes- d'assurance maladie obligatoire (par exemple régime agricole, sections locales mutualistes⁹...). Sa montée en charge a commencé en 2003 ;
- Les informations médico-administratives sur les séjours hospitaliers, tirées du Programme de médicalisation des systèmes d'information (PMSI)¹⁰, sont conservées dans une autre base nationale, gérée par l'Agence technique de l'information sur l'hospitalisation (ATIH), *qui en transmet toutefois une copie au SNIIRAM* depuis 2007 ;
- Les informations sur les décès sont aussi partiellement incluses dans le SNIIRAM :
 - o Le Répertoire national d'identification des personnes physiques (RNIPP) géré par l'INSEE et son « miroir » à la CNAV, le Système national de gestion des identités (SNGI), comportent en plus des éléments d'identité, le statut vital des personnes, qu'ils fournissent aux organismes de sécurité sociale ; c'est ainsi que la date de décès *se retrouve aussi dans le SNIIRAM* ;
 - o Le Centre d'épidémiologie sur les causes médicales de décès (CépiDc) est un laboratoire de l'INSERM, destinataire de la partie médicale des certificats de décès ; ces données ont fait l'objet d'un appariement – encore expérimental - avec le

⁷ Une liste détaillée des sources de données sur la santé, tirée du rapport de 2009 du Haut Conseil de la santé publique sur les systèmes d'information pour la santé publique (Documentation française et site Internet du HCSP) figure en annexe 4.

⁸ Utilisées pour constater les soins et en demander le remboursement (à l'assuré) ou le paiement (au professionnel ou à l'établissement), les « feuilles de soins » sont définies par le Code de la sécurité sociale (articles L.161-33 et R.161-39 et suivants CSS). L'information provient des feuilles de soins électroniques (système SESAM-Vitale) ou est saisie par les caisses à partir des formulaires papier. Après contrôle, les feuilles de soins deviennent des « décomptes » mentionnant le montant pris en charge par l'assurance maladie.

⁹ Les SLM sont des mutuelles de fonctionnaires ou d'étudiants gérant aussi la part obligatoire de l'assurance maladie pour le compte du Régime général.

¹⁰ Le *Programme de médicalisation des systèmes d'information hospitaliers* (PMSI) est régi par les articles L. 6113-7 et 6113-8 du Code de la santé publique. Sa généralisation a débuté dans les années 1990 ; il repose sur un recueil de *résumés de sortie standardisés*, à l'issue des séjours hospitaliers en Médecine, Chirurgie et Obstétrique (MCO) ; le recueil de données a été ensuite élargi à l'hospitalisation à domicile (HAD) aux Soins de Suite et de Réadaptation (SSR) et à la psychiatrie. Ces documents, sans le nom du malade et avec un numéro de sécurité sociale chiffré, sont transmis à l'Agence Technique de l'Information sur l'Hospitalisation (ATIH) et ils servent de base au financement des hôpitaux (en MCO) : c'est le principe de la *Tarification à l'activité* ou T2A instaurée depuis 2004. L'ATIH transmet une copie de ces données au SNIIRAM depuis 2007.

SNIIRAM, appariement dont la mission estime (voir Chapitre II) qu'il devrait devenir permanent.

Ainsi il existe un SNIIRAM au sens étroit, la base d'origine : ce sont essentiellement des données de facturation sur les « soins de ville » (et en cliniques privées).

Et il existe un SNIIRAM au sens large, la base actuelle, qui inclut aussi les informations d'activité médicale des hôpitaux et des cliniques, issues du PMSI. Ce SNIIRAM élargi (SI dans la suite du texte) désigne donc aujourd'hui le mariage de deux bases, l'une hospitalière, l'autre ambulatoire. Décrivant tous les soins reçus par les 65 millions de personnes résidant dans le pays, cette base occupe désormais en France la place centrale, et c'est surtout d'elle qu'il a été question récemment dans le débat public sur les difficultés d'accès aux données de santé.

Malgré des objectifs initiaux limités (1.1), le contenu et l'organisation du SI en une série de bases spécialisées (1.2) en font une des premières bases de données au monde même si elle comporte des limites (1.3).

1.1 Des objectifs initiaux limités

Le texte fondateur du SNIIRAM est l'article L. 161-28-1, inscrit dans le Code de la sécurité sociale par la loi du 23 décembre 1998 :

« Il est créé un système national d'information interrégimes de l'assurance maladie qui contribue :

1° A la connaissance des dépenses de l'ensemble des régimes d'assurance maladie par circonscription géographique, par nature de dépenses, par catégorie de professionnels responsables de ces dépenses et par professionnel ou établissement ;

2° A la transmission en retour aux prestataires de soins d'informations pertinentes relatives à leur activité et leurs recettes, et s'il y a lieu à leurs prescriptions ;

3° A la définition, à la mise en œuvre et à l'évaluation de politiques de santé publique [ce troisième objectif a été ajouté en 2004]. »

Cet article fait suite à un article L. 161-28 qui affirme que « les caisses d'assurance maladie ont pour mission de participer à la maîtrise de l'évolution des dépenses ». Les finalités initiales du SNIIRAM ont été envisagées principalement sous l'angle de la maîtrise des dépenses. La référence aux « politiques de santé publique », ajoutée dans un second temps par la loi du 9 août 2004, a élargi opportunément la perspective¹¹.

La « maîtrise médicalisée des dépenses » suppose une information suffisamment détaillée¹² et complète sur les pratiques de soins et de prescriptions. La qualité et l'exhaustivité des retours

¹¹ Les finalités du SNIIRAM sont déjà reformulées de manière plus large dans les arrêtés successifs relatifs à la mise en œuvre du SNIIRAM qui le régissent (le dernier est celui du 19 juillet 2013, figurant en annexe 11) et dans le contrat d'objectif entre l'État et l'UNCAM.

¹² C'est la raison initiale du codage des actes et des prestations prévu à l'article L. 161-29 CSS (le codage des pathologies, également prévu, n'a pas été mis en œuvre hors de l'hôpital). Le codage, notamment des

d'information vers les professionnels sont essentiels. Un transfert et un regroupement des données au niveau d'un SI national inter-régimes étaient nécessaires. En effet, les feuilles de soins sont d'abord transmises aux différentes caisses d'assurance maladie, pour permettre la « liquidation » des feuilles de soins. Au niveau d'une caisse ou d'un régime, la prise en compte de l'activité des professionnels n'est que partielle. Le SNIIRAM a été créé pour avoir une vision complète.

Conçu à l'origine comme un outil de l'assurance maladie pour la maîtrise des dépenses et le dialogue avec les professionnels de santé de ville, le SNIIRAM s'est enrichi dans son contenu et ses usages, surtout à partir du moment où il a reçu les données hospitalières du PMSI. Le PMSI, généralisé dans les années 1990, avait été conçu de son côté pour mesurer l'activité des hôpitaux et des cliniques (cf. notamment la loi hospitalière de juillet 1991), avant son utilisation à partir de 2004 pour servir de base à la tarification à l'activité (T2A).

Il peut sembler surprenant que les données de facturation des soins se trouvent initialement dans deux bases différentes : PMSI pour les hôpitaux et SNIIRAM pour les soins de ville (en fait l'assurance maladie recueille les flux de facturation des professionnels libéraux mais aussi des cliniques¹³). Ce dualisme est un produit de l'histoire : les relations avec les professionnels de ville sont gérées par l'assurance maladie alors que l'Etat gère la relation avec les hôpitaux. C'est pourquoi la base PMSI a été constituée auprès d'un organisme distinct de l'assurance maladie, l'Agence technique de l'information sur l'hospitalisation (ATIH), créé par l'Etat pour gérer les référentiels techniques du PMSI.

Quoi qu'il en soit, ces données ont en commun de décrire des prestations¹⁴ dispensées aux bénéficiaires de l'assurance maladie et financées par elle directement ou indirectement. C'est pourquoi il a été prévu dès 2002¹⁵ que ces données seraient regroupées dans une même base, le SNIIRAM, transmises par les organismes d'assurance maladie d'une part et par l'ATIH d'autre part. Cependant, s'agissant de systèmes conçus indépendamment, il a fallu quelques efforts pour que les données relatives à une même personne puissent être « chaînées » entre elles alors que le nom et le numéro de sécurité sociale sont occultés (sur l'anonymat de la base voir infra § 2.1).

Cette jonction entre les deux recueils de données, hospitalier et ambulatoire, permettant le suivi longitudinal des soins reçus par une même personne, est opérationnelle depuis 2009 : c'est le progrès décisif qui a contribué à donner sa nouvelle dimension au SNIIRAM, pour la santé publique et la connaissance des pratiques de soins.

L'élargissement des usages du SNIIRAM comme outil de connaissance et d'information a été très tôt l'œuvre des équipes de la CNAMTS et il était d'emblée prévu que la CNAMTS rendrait compte de la qualité de son nouveau système d'information à un Comité pour la transparence des statistiques de l'assurance maladie (COTSAM).

médicaments et des actes techniques médicaux, a été rendu possible par la dématérialisation des feuilles de soins s'appuyant sur les cartes *Vitale* (dont la généralisation a débuté aussi en 1998).

¹³ L'assurance maladie reçoit directement les bordereaux de facturation des cliniques mais les résumés de sortie comportant les codes des diagnostics sont anonymisés puis transmis au SNIIRAM séparément, via l'ATIH.

¹⁴ La tarification à l'activité (T2A) des activités hospitalières de médecine, chirurgie et obstétrique (MCO) est une tarification forfaitaire par types de séjours (les séjours hospitaliers sont classés en Groupes homogènes de séjours sur la base des données médicales que comporte chaque résumé de sortie).

¹⁵ Le déversement du PMSI dans le SNIIRAM se fonde sur les articles L. 161-28-1, L. 161-29, R. 161-29 et R. 161-30 du CSS. Il a été décidé par un arrêté ministériel du 11 avril 2002, visé par la CNIL.

Un autre organisme a succédé au COTSAM, regroupant non plus seulement des professionnels de santé et des experts mais aussi d'autres acteurs (État, régimes d'assurance maladie obligatoire et fédérations d'assurance maladie complémentaire rejoints ensuite par les associations d'usagers du système de santé, les agences sanitaires et les fédérations hospitalières, plus des membres associés). Créé par une loi d'août 2004, l'Institut des données de santé (IDS) a pour mission « d'assurer la cohérence et de veiller à la qualité des systèmes d'information utilisés pour la gestion du risque maladie et de veiller à la mise à disposition de ses membres, de la Haute Autorité de santé, des unions régionales des professionnels de santé ainsi que d'organismes désignés par décret en Conseil d'État, à des fins de gestion du risque maladie ou pour des préoccupations de santé publique, des données issues des systèmes d'information de ses membres, dans des conditions garantissant l'anonymat fixées par décret en Conseil d'État pris après avis de la Commission nationale de l'informatique et des libertés. »

Lieu d'expression et d'information des acteurs, l'IDS s'est aussi vu reconnaître un rôle dans l'attribution des droits d'accès au SNIIRAM, a milité pour l'élargissement des droits d'accès et a contribué à l'élargissement de ses usages.

1.2 Le contenu et l'organisation du SI

Le contenu du SI peut se décrire en énumérant les données qui le constituent et leur origine (121) ou en considérant la manière dont ces données sont organisées et mises à la disposition des utilisateurs (122).

121 Les données qui constituent le SI

Le contenu du SI est décrit dans un protocole inter-régimes –très détaillé- approuvé par un arrêté ministériel, prévu à l'article L. 161-28-1 du CSS.

Les données brutes se présentent comme un entrepôt de données qui résultent schématiquement de l'addition :

- de données sur les bénéficiaires (fichiers administratifs et fichiers du service médical) ;
- de données sur les prestataires (répertoire des professionnels et répertoire des structures) ;
- des décomptes qui récapitulent les demandes de prise en charge et la réponse apportée. Ce sont principalement les « feuilles de soins » et les remboursements, mais on doit citer aussi les relevés de soins externes hospitaliers¹⁶, les bordereaux de facturation des cliniques, les avis d'arrêts de travail et les indemnités journalières, les rentes pour accidents du travail et maladie professionnelles etc. ;
- et des résumés de sortie hospitaliers transmis depuis 2007 par l'Agence technique d'information sur l'hospitalisation (ATIH).

¹⁶ Depuis 2009

Données sur les bénéficiaires

- numéro d'anonymat (expression incluant le NIR, chiffrée deux fois de manière irréversible)
- sexe
- mois et année de naissance
- date de décès (le cas échéant)
- code postal du domicile
- hébergement dans un EHPAD (le cas échéant)
- code CIM 10* en cas d'affection(s) de longue durée (ALD) ou de maladie professionnelle
- médecin traitant
- bénéficiaire de la CMU complémentaire ou de l'aide à la complémentaire santé (le cas échéant)
- régime d'assurance maladie et organisme gestionnaire
- rente pour accident du travail et/ou invalidité le cas échéant...

Classification internationale des maladies (10^{ème} révision) de l'OMS

Données sur les professionnels de santé ayant prescrit ou réalisé la prestation de soins

- nom et prénoms (en clair ou remplacés par un code, selon l'habilitation du lecteur)
- profession (et spécialité s'il y a lieu)
- lieu d'exercice (géocodage)
- statut conventionnel (secteur 1 ou 2 par exemple)...

Données issues des feuilles de soins (et des bordereaux de facturation des cliniques privées, d'une partie des établissements médico-sociaux et des factures de soins externes des hôpitaux)

- nature de la ou des prestation(s) remboursable(s) avec leur code détaillé (codes CIP ou UCD pour les médicaments*, codes CCAM** pour les actes médicaux, codes des examens biologiques, des actes d'auxiliaires médicaux, des dispositifs médicaux...),
- date(s) de la ou des prestation(s),
- date de la feuille de soins,
- date de la prescription (s'il y a lieu),
- adressage par le médecin traitant (le cas échéant),
- pour chaque acte ou prestation : tarif pratiqué, base de remboursement, taux de remboursement***, somme perçue,

- compléments d'acte éventuels ;
- lieu de la prestation et indemnités kilométriques en cas de visite à domicile...

*La codification du *Club des industries pharmaceutiques* (CIP) figure dans le code à barres sur les boîtes des médicaments vendus en pharmacie. Les hôpitaux, qui achètent généralement leurs médicaments en vrac et les conditionnent eux-mêmes, utilisent la codification en *Unités communes de dispensation* (UCD).

**Classification commune des actes médicaux

***Le taux de remboursement par le régime obligatoire dépend de plusieurs facteurs : soins en rapport ou non avec une ALD, « risque » au titre duquel les soins ont été dispensés (maladie, maternité ou accident du travail et maladie professionnelle), régime de sécurité sociale...

Données issues des résumés de sortie hospitaliers* (tous les séjours en établissements de Médecine-Chirurgie et Obstétrique mais aussi désormais en Hospitalisation à Domicile, en Soins de Suite et Réadaptation et en Psychiatrie)

- numéro d'anonymat (permettant le chaînage entre les soins pour une même personne)
- sexe
- mois et année de naissance
- code postal du domicile
- établissement (numéro FINESS)
- dates d'entrée et de sortie
- modes d'entrée et de sortie (transfert, décès, retour à domicile ou autre destination...)
- diagnostic principal (code CIM 10)
- diagnostics associés (codes CIM 10)
- actes « classants** » (codes CCAM)
- degré de dépendance (pour les hospitalisations en SSR notamment)...
- nombre de jours en soins intensifs ou réanimation
- Groupe homogène de séjours (GHS) et tarif applicable
- médicaments (codes UCD) et dispositifs médicaux facturables en sus du GHS

*Les formats des résumés de sortie ont tendance à varier d'une année sur l'autre et ils diffèrent entre MCO, HAD, SSR et Psy.

**Dans les résumés de sortie servant de base à la rémunération des établissements, il n'est pas obligatoire de saisir tous les actes : en règle générale, seuls sont pris en compte pour le classement dans un groupe homogène de séjours (groupe tarifaire) les actes effectués au bloc opératoire (en MCO) ou les actes de rééducation (en SSR). Toutefois, les médecins libéraux étant rémunérés à l'acte, les actes qu'ils facturent figurent tous dans les bordereaux de facturation des cliniques privées.

Les données d'identité des personnes (nom, prénoms, date et lieu de naissance, NIR¹⁷) sont supprimées et remplacées par un « numéro d'anonymat » ou numéro de chaînage, qui est obtenu par un chiffrement irréversible du NIR (cf. infra § 2.1 sur cette procédure de chiffrement), de sorte que les soins successifs relatifs à une même personne puissent être rattachés à un numéro d'anonymat identique¹⁸ (c'est le principe du chaînage).

Le chaînage avec l'hôpital a constitué comme on l'a vu l'étape la plus importante (réalisé en 2009, avec un rattrapage pour 2007 et 2008, il a été « industrialisé » en 2010) mais le contenu du SI a connu d'autres progrès depuis que l'entrepôt de données a commencé à être alimenté en 2003 :

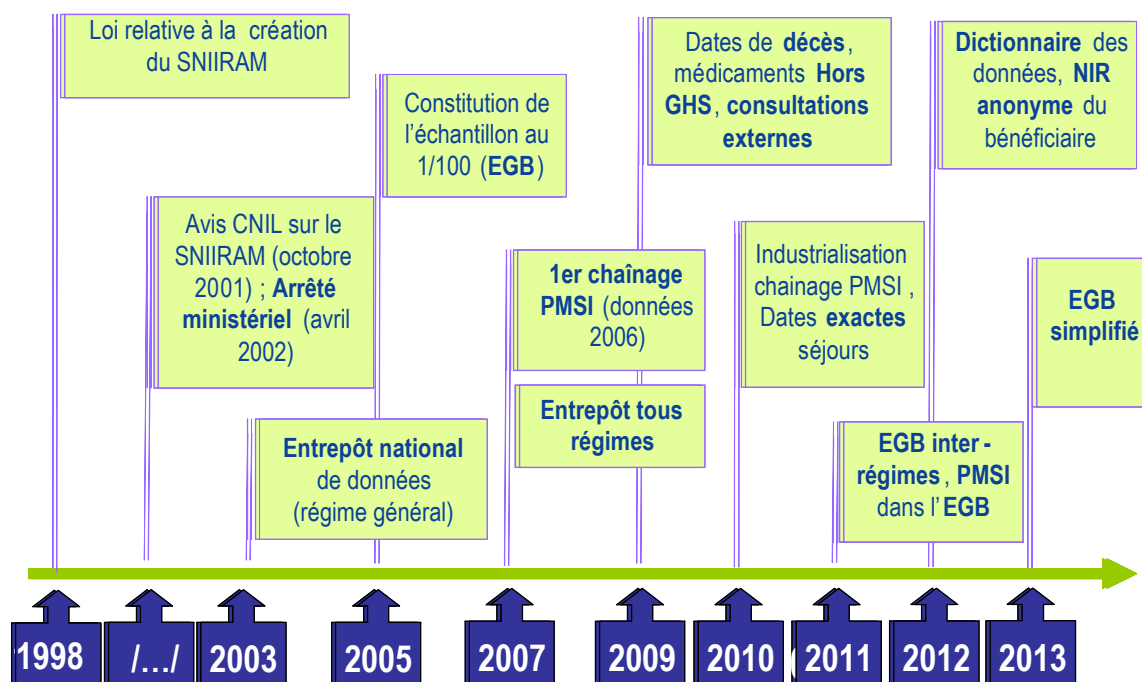
- la contribution de certains petits régimes et de certaines sections locales mutualistes ayant été plus tardive, on approche seulement aujourd'hui l'exhaustivité ;
- les dates de décès, les consultations externes des hôpitaux et les médicaments rétrocedés à l'hôpital (pour les patients externes) ont été ajoutés en 2009 ;
- le codage a été étendu à d'autres produits de santé (dispositifs médicaux et matériels) ;
- l'allongement de la profondeur historique autorisée a été obtenu de la CNIL en 2011 ;
- le NIR chiffré des bénéficiaires ayants droit a commencé à être ajouté en 2012 (en sus du NIR de l'assuré), ainsi que le géocodage des professionnels.

La durée de conservation des données du SNIIRAM initialement fixée à 2 ans plus l'année en cours, a été sensiblement allongée en 2011 après l'épisode du *Mediator*, quand l'intérêt de la base pour des études pharmaco-épidémiologiques a été mieux compris tant par les pouvoirs publics que par la CNIL : la durée de conservation des données en ligne est désormais de 3 ans plus l'année en cours (et de 10 ans plus l'année en cours pour la base hospitalière de l'ATIH). La CNAMTS et l'ATIH sont en outre autorisées à conserver les données antérieures, archivées sur 10 ans, qu'il est possible d'exploiter en cas de nécessité (la profondeur totale de 14 ans excède pour l'instant l'âge des plus anciennes données du SNIIRAM). Enfin, la profondeur autorisée est de 20 ans pour l'Echantillon généraliste des bénéficiaires (EGB, échantillon au 1/97).

Le schéma ci-après met en évidence l'ampleur du travail réalisé sur la décennie, par la CNAMTS, pour améliorer un SI qui reste une base jeune, à l'aune des très grands projets similaires.

¹⁷ Le numéro d'inscription au répertoire national d'identification des personnes physiques (NIR, RNIPP) est communément appelé numéro de sécurité sociale parce qu'il est réservé à cet usage en vertu d'une doctrine constante de la CNIL. Il est attribué à la naissance (pour les personnes nées en France) ou lors de l'installation en France. Les ayants droit d'un assuré social (enfants, conjoint non assuré...) utilisent aujourd'hui le numéro de l'assuré (et non le leur) sur les feuilles de soins.

¹⁸ En pratique, c'est plus complexe pour les ayants droit lorsque leurs données de soins sont rattachées au NIR de l'assuré : leur numéro d'anonymat est obtenu en chiffrant le NIR de l'assuré auquel on ajoute l'année de naissance et le sexe du bénéficiaire des soins. Le résultat est qu'un ayant droit va changer de numéro d'anonymat lorsqu'il deviendra lui-même assuré (et qu'on peut confondre les jumeaux de même sexe !). Cette difficulté technique sera réglée lorsque tous les ayants droit seront identifiés, dans la base, par leur propre NIR chiffré (évolution en cours).



Source CNAMTS

122 L'organisation du SI pour les utilisateurs : extractions, publications, magasins et outils

L'entrepôt de données ou *datawarehouse*, alimenté par les régimes d'assurance maladie et l'ATIH n'est pas fait pour être interrogé directement.

Des données peuvent en être extraites à la demande et communiquées¹⁹ sous certaines conditions (détaillées au § 2.3), en vue d'études spécifiques impliquant ou non l'appariement avec des données d'enquêtes ou avec les données d'une autre base administrative.

Le SI alimente aussi en chiffres le site *ameli.fr* de la CNAMTS, les tableaux de bord diffusés à ses membres par l'Institut des données de santé (IDS) ou par l'Institut statistique des professionnels de santé libéraux (isPL). Il alimente également les sites en ligne *SCORE-santé* (de la Fédération nationale des ORS) et *Eco-Santé* (de l'IRDES) qui sont, avec le site de la DREES, les sources d'informations les plus largement utilisées en économie de la santé par les enseignants, les étudiants et les administrations.

Mais le SI a été aussi organisé par la CNAMTS pour permettre -aux personnes habilitées- d'effectuer des requêtes sur le SNIIRAM en utilisant pour ce faire les « magasins », appelés *datamarts*. Ces derniers ont été conçus pour que les utilisateurs, en fonction des différents profils d'habilitation et à condition d'avoir suivi la formation, les interrogent via le portail du SNIIRAM, en s'aidant des outils que la CNAMTS a progressivement mis à leur disposition (forums et bientôt dictionnaire de données en ligne).

¹⁹ La base complète du PMSI peut être aussi acquise auprès de l'ATIH (voir infra § 2.2 et 2.3).

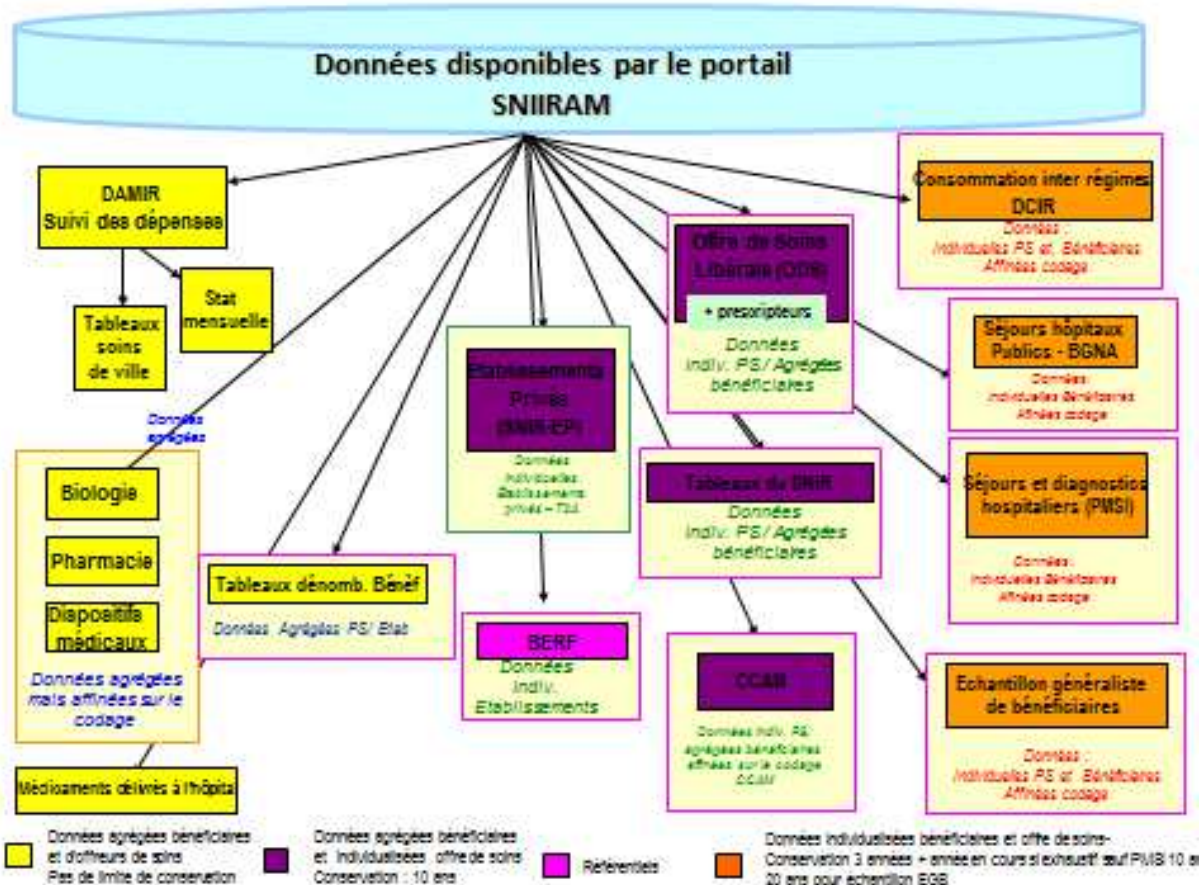
Les datamarts du SNIIRAM, au nombre de 16 aujourd'hui, sont regroupés par thèmes :

- Le **thème des bénéficiaires** inclut essentiellement :
 - o Le **datamart de consommation inter-régimes (DCIR)**, constitué des données individuelles par bénéficiaire (tous ceux ayant consommé des soins dans l'année, soit 93 % de la population française²⁰). C'est le niveau le plus détaillé pour les interrogations de la base et donc celui pour lequel les enjeux liés à l'accès sont les plus importants. Signalons dès à présent que parmi les données du DCIR, sont considérés comme « sensibles » parce que particulièrement susceptibles de permettre une réidentification, quatre catégories de données : dates des soins, code postal du domicile, mois et année de naissance, date de décès (éventuel). La CNIL peut en interdire le croisement (soit limiter l'accès à une seule d'entre elles, les autres étant « floutées » c'est-à-dire rendues moins précises : jour→mois, mois→année, code postal→département) ;
 - o L'**échantillon généraliste des bénéficiaires (EGB)**. C'est un échantillon au 1/100²¹ de la population des bénéficiaires, et c'est aussi le banc d'essai des améliorations apportées à la base pour en faciliter l'utilisation ;
 - o Les **données hospitalières détaillées du PMSI** transmises par l'ATIH (tous les résumés de sortie individuels), qui constituent un datamart supplémentaire dans le thème « bénéficiaires ». Elles ne sont pas intégrées dans le DCIR et sont juxtaposées au datamart préexistant « Budget Global » (dit BGNA) sur les frais de séjours à l'hôpital (données de qualité inégale²²). Nota : les données PMSI peuvent être également acquises directement auprès de l'ATIH, laquelle met aussi en ligne des statistiques agrégées sur son site SNATIH ;
- Le **thème des dépenses** inclut divers datamarts, notamment les DAMIR (dépenses de l'assurance maladies inter-régimes) où on trouve les volumes et les dépenses agrégées par code de prestation. D'autres datamarts regroupent les dépenses par grands ensembles (soins de ville, médicaments en ville, médicaments à l'hôpital, examens de biologie en ville etc.) ;
- Le **thème de l'offre de soins** inclut plusieurs datamarts constitués de données individuelles sur telle ou telle profession afin de connaître sa démographie, son activité, les tarifs qu'elle pratique et les dépenses qu'elle génère. Précis sur les professionnels de santé, ces datamarts sont constitués de données agrégées pour ce qui est des bénéficiaires des soins. Selon les profils d'habilitation des utilisateurs, l'identité de chaque professionnel de santé est soit codée soit en clair.

²⁰ Le SI couvre les soins délivrés à la population résidant en France y compris les bénéficiaires de l'aide médicale d'Etat (les soins aux étrangers de passage n'y figurent pas).

²¹ Au 1/97 plus précisément.

²² Il est prévu dans le cadre du projet FIDES (Facturation individuelle des établissements de santé) que les établissements publics et assimilés transmettent directement, et au fil de l'eau, des factures individuelles à l'assurance maladie, ce qui garantirait une cohérence entre factures nominatives et résumés de sortie anonymes (il n'a pas été envisagé de fusion pure et simple de ces deux documents), mais ce projet connaît des retards.



Source CNAMTS

Aussi curieux que cela puisse paraître dans un climat de polémique sur l'accès aux données, le but initial de ces datamarts était surtout de permettre l'accès à la base de personnes extérieures à la CNAMTS, puisque les services de l'assurance maladie ont longtemps continué de se servir des outils dont ils disposaient précédemment pour élaborer leurs statistiques mensuelles et leurs tableaux de bord (et parce que les échelons régionaux du service médical et les défunctes Unions régionales des caisses d'assurance maladie disposaient d'accès aux bases régionales du Régime Général destinées à la gestion du risque).

Outre les datamarts, on doit citer parmi les outils mis à la disposition des utilisateurs du SI :

- Des formations pour les utilisateurs (neuf formations différentes, couvrant l'architecture du SI, l'outil de requête et les différents datamarts) ;
- des bibliothèques de requêtes ou des tableaux de bord ou des « cubes de données » alimentés (notamment) par le SNIIRAM. Il s'agit de requêtes standards auxquelles la réponse, sous forme de tableaux, est mise à jour automatiquement en tenant compte des données les plus récentes. Ces tableaux de bord peuvent être mis à disposition sur le portail du SNIIRAM, ou sur le site *ameli.fr* de la CNAMTS (ou sur le site de l'ATIH pour ce qui concerne les données hospitalières) ;
- le forum des utilisateurs sur le portail du SNIIRAM où l'on peut poser des questions et proposer des réponses (en pratique ce sont surtout les responsables du SI qui répondent aux questions) ;

- un dictionnaire de données dont l'usage était jusqu'à présent malaisé (parce que les utilisateurs devaient l'installer sur leur poste de travail) mais qui devrait être remplacé par une version en ligne en septembre 2013.

Le portail lui-même intègre par ailleurs des outils de sécurité qui enregistrent et vérifient l'habilitation des utilisateurs. Il donne accès aux datamarts et aux autres outils.

1.3 La richesse et les limitations du SI

Il ressort de cette description du SI qu'il est une base de données d'une taille et d'une richesse exceptionnelle (141) mais dont le périmètre restreint limite les usages (142)

131 Une base de données d'une richesse et d'une qualité exceptionnelles

La France, pays centralisé de 65 millions d'habitants dispose aujourd'hui avec l'ensemble SNIIRAM-PMSI d'une des plus grandes, voire de la plus grande base médico-administrative au monde : elle regroupe par an 1,2 milliard de « feuilles de soins », 500 millions d'actes médicaux et 11 millions de séjours hospitaliers²³ (en médecine, chirurgie et obstétrique), avec potentiellement une profondeur historique de 14 ans (20 ans pour l'EGB et le PMSI).

Il existe hors de France des bases de données comportant des informations plus étendues que celles du SI français²⁴, notamment des résultats d'examens et des informations transmises par les médecins sur les diagnostics et les facteurs de risques (alcool, tabac, masse corporelle...), ou des bases plus faciles à apparier avec des sources de données complémentaires, mais aucune base à notre connaissance ne réunit autant d'informations médicales chaînées sur une population aussi vaste²⁵. La base anglaise GPRD (*General Practice Research Database*) est pour l'instant, avec moins de 4 millions de dossiers actifs²⁶, très loin des effectifs du SI, de même que les bases américaines de *Medicare*, de certains *HMOs* ou des *Veterans* qui comportent en outre des biais de sélection. Jusqu'à présent, seuls des petits pays ont réussi à chaîner systématiquement leurs données de soins primaires avec leurs données hospitalières.

On peut donc citer parmi les points forts de la base française :

- l'exhaustivité de la population (effectif et absence de biais), qui confère au SI sa grande puissance statistique,
- les données détaillées sur les consommations de soins en ville et à l'hôpital, notamment tous les traitements médicamenteux prescrits au long cours et toutes les hospitalisations,
- le chaînage (suivi longitudinal),
- la chronologie précise,

²³ Séjours hospitaliers hors "séances" (chimiothérapies, radiothérapies, dialyses...).

²⁴ Il existe aussi en France des panels de médecins constitués par des sociétés d'étude privées qui recueillent ces types de données.

²⁵ *Strengthening Health Information Infrastructures for Health Care Quality Governance* - OCDE report preliminary version April 2013

²⁶ Le nombre de dossiers actifs comportant un chaînage entre ville et hôpital serait sensiblement inférieur selon l'information recueillie par la mission auprès du NHS.

- la qualité d'ensemble plutôt bonne des données. Il faut noter à ce sujet que l'exactitude du NIR des cartes *Vitale*²⁷ et son mode de recueil automatique contribuent fortement à la qualité de l'ensemble. De même, la saisie des codes des médicaments est pour l'essentiel automatisée. Les codes des diagnostics sont en revanche un peu moins fiables.

Le nombre croissant de demandes d'accès provenant des chercheurs, et le nombre de publications basées sur des données du SNIIRAM, illustrent la richesse de la base, naguère sous-estimée et encore sous-utilisée. Les chercheurs sont ainsi devenus depuis 2012 les principaux utilisateurs du SI en nombre de requêtes : la mise à disposition d'informations riches, structurées et déjà recueillies, intéressantes par elles-mêmes ou pouvant compléter utilement des informations collectées par enquêtes, fournissent un matériau d'une richesse désormais évidente.

Requêtes effectuées en 2012 sur l'échantillon de bénéficiaires (Source CNAMTS)

	Nombre total de requêtes	%
Régime Général	6 315	28%
Sections locales mutualistes	184	1%
ARS	214	1%
Ministère Santé	189	1%
Organismes de recherche	12 649	57%
Agences sanitaires et assimilés (2)	1 946	9%
IDS	772	3%
Total	22 296	

(1) Organismes de recherche = INSERM, IRDES, CNRS, CETAF, organismes habilités par l'IDS

(2) Sphère Santé = INVS, ANSM, INCA, HAS, HCAAM, ATIH, OFDT, Fonds CMU, Agence de la biomédecine

132 Les limitations du SI : un périmètre restreint au regard des usages possibles

Compte tenu du temps nécessaire pour faire aboutir des projets de cette taille, le SI doit être considéré, on l'a vu, comme une base jeune. Les limites actuelles doivent donc être appréciées relativement aux progrès déjà accomplis, mais surtout au regard des études que l'on entend réaliser. Les besoins en données dépendent des usages. Pour ne prendre que deux exemples :

- Pour apprécier l'impact des médicaments en vie réelle, il est utile dans bien des cas de connaître les facteurs de risque associés aux comportements individuels (consommations de tabac, alcool etc.), voire à des facteurs génétiques ;

²⁷ Pour les besoins de la télétransmission des feuilles de soins, tous les organismes d'assurance maladie ont procédé à partir de 1998 à une certification systématique de leurs fichiers –pour tous les bénéficiaires- en les confrontant aux données du *Répertoire national d'identification des bénéficiaires de l'assurance maladie* (RNIAM, géré par la CNAV). Seules les premières cartes *Vitale* émises comportaient encore quelques erreurs. Malgré les difficultés qui subsistent à identifier correctement les ayants droit et quelques cas d'usurpation d'identité (notamment à l'hôpital), les conditions sont réunies en France pour une très bonne qualité d'ensemble des données d'identification servant de base au chaînage des soins.

- Pour estimer l'impact d'une franchise de remboursement sur l'accès aux soins, il est essentiel de connaître les revenus des ménages.

Les limitations actuelles du SI par lui-même (hors possibilités d'appariement avec d'autres sources) découlent des sources administratives qui l'alimentent et sont bien connues :

- Le codage des diagnostics associé aux feuilles de soins n'a jamais été mis en œuvre. A la différence des bases de données issues de dossiers médicaux, le SI ne comporte que des informations diagnostiques partielles, dont les modes de recueil et de codage (en CIM-10) ne garantissent pas, en outre, une fiabilité parfaite :
 - Une part des diagnostics provient des résumés de sortie hospitaliers du PMSI ;
 - les autres viennent du service médical des caisses et concernent les affections de longue durée (ALD) et les maladies professionnelles indemnisées ;
- Le SI ne connaît pas non plus les causes de décès (l'appariement réalisé par la CNAMTS, la DREES et l'INSERM n'est encore qu'expérimental) ; on ne peut donc pas mesurer avec précision la mortalité associée à tel ou tel facteur de risque, ou consécutive à une hospitalisation récente²⁸ (la connaissance des causes permet d'écarter des décès manifestement sans lien avec le risque considéré) ;
- Les examens biologiques, anatomo-cyto-pathologiques ou d'imagerie sont mentionnés dans le SNIIRAM, mais sans leurs résultats ;
- Le SI ne fournit pas d'informations systématiques sur les facteurs de risque tels que la tension artérielle, les antécédents familiaux, les habitudes alimentaires, la consommation de tabac et d'alcool ou l'indice de masse corporelle... Il ne comporte pas non plus d'informations génétiques ou épigénétiques ni liens familiaux (pas même le lien mère-enfant) ;
- Le SI renseigne fort bien sur tous les médicaments et autres produits de santé achetés en pharmacie s'ils sont présentés au remboursement, mais seulement sur une partie des médicaments ou des dispositifs médicaux administrés ou implantés à l'hôpital (ceux qui sont facturables en sus du tarif forfaitaire de séjour). Les produits de santé non remboursables n'y sont pas, ni ceux achetés sans prescription médicale ;
- Le SI ne comporte pas de données socio-économiques (la seule indication sur les revenus du ménage est le bénéfice éventuel de la CMU complémentaire ou de l'aide pour une complémentaire santé) ni de données médico-sociales (handicap, dépendance...) hors de la notion d'hébergement en EHPAD ;
- Pour mémoire, une autre forme de limitation du SI est la profondeur historique de mise à disposition des données. De quatorze ans (voire vingt pour l'EGB), on a vu qu'elle est aujourd'hui suffisante -puisque'elle excède l'âge des plus anciennes données de la base- si les experts peuvent effectivement, en cas de besoin, accéder aux données archivées. Cette limite pourrait toutefois se révéler gênante à l'avenir s'il se confirme que le SI sera un terrain privilégié pour des études rétrospectives sur les effets des médicaments mais aussi sur les effets à long terme de facteurs environnementaux, qui peuvent exiger un recul historique important ;

²⁸ La mortalité post hospitalière ou post opératoire à 30 jours est un indicateur de qualité à manipuler avec précaution (l'ajustement au risque notamment est délicat) mais il est disponible dans d'autres pays, notamment au Royaume Uni. Le sujet a été évoqué notamment dans un rapport de l'IGAS de 2010 *L'information des usagers sur la qualité des prises en charge des établissements de santé*.

- Inversement, les données les plus récentes ne parviennent au SI qu’après quelques jours au mieux pour les médicaments télétransmis, voire quelques semaines pour les feuilles de soins saisies dans les caisses ou quelques mois pour les données hospitalières²⁹. Cela signifie qu’on ne peut utiliser le SI en temps réel pour surveiller l’état de santé de la population ou mesurer l’impact d’une campagne de communication. Pourtant les données sont en majorité télétransmises instantanément à l’assurance maladie. Les contrôles, corrections et décomptes puis la transmission par les caisses ou les hôpitaux prennent du temps mais rien n’empêcherait en principe de procéder ou de faire procéder à une centralisation et une exploitation de données provisoires en temps réel. C’est d’ailleurs ce que font déjà certaines entreprises à partir de données issues -directement ou indirectement- des feuilles de soins électroniques des pharmaciens et obtenues auprès des organismes concentrateurs techniques de cette profession ;
- Les possibilités de localisation géographique sont limitées à un niveau insuffisant pour l’analyse adéquate des inégalités sociales et territoriales ou de certains facteurs de risques liés à l’environnement.

Localisation géographique

Pour divers usages, comme l’étude des inégalités territoriales de santé ou de ressources de soins, de la santé environnementale ou de l’analyse d’agrégats spatio-temporels, on rencontre des limites imposées par l’imprécision de la domiciliation des personnes dans les bases de données nationales, la résolution spatiale la plus fine enregistrée étant la commune ou le code postal.

Concernant les professionnels et établissements de santé, les données initiales, dans les fichiers de l’assurance maladie, comportent des adresses précises qui sont désormais géocodées (coordonnées X/Y).

Pour les adresses des bénéficiaires des soins il serait envisageable qu’elles soient converties en code IRIS³⁰ et transférées sous cette forme au niveau national. Ce code pourrait être accessible et utilisable par les chercheurs sans rompre l’anonymat, à l’instar des autres données à caractère personnel contenues dans ces bases.

Une étape supplémentaire serait que le géocodage des adresses des bénéficiaires, au niveau de l’organisme de collecte soit réalisé au niveau le plus précis (coordonnées X/Y), ce qui autoriserait ultérieurement des agrégations en fonction des besoins, mais cela impliquerait de gérer ces informations à part pour des raisons de confidentialité, une adresse postale étant presque toujours directement identifiante.

²⁹ Les hôpitaux transmettant souvent avec un décalage de plusieurs mois les informations sur les séjours.

³⁰ Les Ilots Regroupés pour l’Information Statistique (IRIS) de l’INSEE sont de superficie variable et conçus pour inclure une population comprise généralement entre 1800 et 5000 habitants.

Pointer les limitations d'une base de données n'implique pas qu'il faille élargir indéfiniment son périmètre : le SI devra certainement être enrichi par le recueil de nouvelles catégories de données ou par l'adjonction permanente d'autres bases (comme c'est désormais le cas pour les résumés de sortie hospitaliers du PMSI et comme on l'envisage pour les causes de décès), mais il n'a pas vocation à réunir en son sein toutes les données nécessaires à toutes les études possibles dans le champ de la santé.

Pour chaque limite du SI, il faut mettre en regard le bénéfice attendu d'un enrichissement et le coût ou les risques de l'appariement ou de la collecte.

L'appariement avec des données recueillies par des enquêtes auprès d'un échantillon (y compris cohortes, prélèvements de tissus...) ou issues d'autres bases administratives³¹ peut être la solution la plus simple pour réunir les données nécessaires aux études ou aux recherches. Cependant :

- il se heurte à des difficultés juridiques et pratiques considérables lorsqu'il faut retrouver dans une base supposée anonyme les données relatives aux personnes ayant répondu à une enquête, comme nous le verrons au § 2.3 à propos des restrictions à l'usage du NIR ;
- et l'appariement avec d'autres bases administratives peut se heurter aux retards à constituer des SI nationaux dans certains domaines comme le médico-social³² (handicap, dépendance, établissements médico-sociaux...) ou la santé au travail et la santé scolaire.

Quant à un recueil exhaustif de données supplémentaires, il implique de prendre en compte, d'un côté :

- les progrès de l'informatique et la réduction qui en résulte des coûts de duplication et de transmission³³ ;
- l'élargissement des besoins en données pour la régulation du système. Il en a été ainsi pour le codage des prestations, plusieurs interlocuteurs de la mission étant d'avis que des transmissions d'informations supplémentaires, sur les diagnostics ou les résultats d'examens par exemple, pourraient devenir nécessaires, dans une logique de contrôle de la qualité des soins ou de paiement basé sur la qualité...

et d'un autre côté :

- les risques accrus au regard de la confidentialité ;
- les coûts éventuels pour les professionnels et les organismes à qui incomberait l'obligation de collecter ou transmettre les données. Il est sain que les administrations limitent les recueils exhaustifs aux données strictement nécessaires à l'accomplissement de leurs missions et qu'elles procèdent plutôt par enquêtes s'il s'agit seulement de répondre à un besoin de connaissance.

³¹ Notamment si les données relèvent d'un autre domaine que la santé (revenus, professions...).

³² La pauvreté des SI nationaux dans le domaine médico-social traduit sans doute moins le rôle prépondérant des collectivités territoriales en ce domaine que la difficulté des acteurs publics à définir ensemble un référentiel minimum commun pour les systèmes de gestion et pour les statistiques.

³³ Les résultats d'examens sont déjà tous enregistrés sous forme électronique et tous les médicaments administrés dans les établissements ont vocation à l'être aussi.

2 Les accès au SNIIRAM

Les restrictions juridiques mises à l'accès au SI se justifient essentiellement par un risque de ré-identification des personnes dans cette base pourtant réputée anonyme (2.1), toutefois la complexité des règles d'accès actuelles est inadaptée (2.2), notamment pour ce qui concerne les appariements avec d'autres données et l'emploi du NIR (2.3) ; en outre, la difficulté technique d'utilisation du SI (2.4) constitue un obstacle au moins aussi gênant que les barrières juridiques.

2.1 Les risques de réidentification

Les données constitutives du SI sont, à la source, des données individuelles nominatives collectées pour assurer le remboursement des soins ou la rémunération des établissements lors de chaque acte de soins en ville ou lors de chaque séjour en établissement de santé. Quand les décomptes sont transmis ensuite à la base centrale du SNIIRAM par les caisses, ces données individuelles nominatives sont anonymisées (on dit parfois – plus précisément – qu'elles sont « dé-identifiées »). Les noms, les adresses n'y figurent plus et le NIR est remplacé par un identifiant propre au SI, au moyen d'un procédé irréversible : il n'est pas possible de revenir au NIR à partir de l'identifiant SNIIRAM³⁴.

La même technique et les mêmes clés sont utilisées pour les résumés de sortie hospitaliers. Comme on l'a vu, les soins relatifs à une même personne peuvent être « chaînés », la personne étant théoriquement toujours désignée par le même numéro d'anonymat, sans que ni son nom ni son NIR n'apparaissent.

L'anonymat des personnes n'est pas pour autant garanti ; une réidentification indirecte est éventuellement possible et cela même si les textes affirment l'anonymat des bases de données considérées. Ainsi l'article L. 161-28-1 du CSS dispose que « Les données reçues et traitées par le SNIIRAM préservent l'anonymat des personnes ayant bénéficié des prestations de soins »³⁵.

Le caractère indirectement nominatif des données individuelles du SI résulte comme on le verra, de leur précision, de leur chaînage et de leur exhaustivité, toutes propriétés qui font justement leur intérêt, notamment pour un usage en épidémiologie, dans le cadre de la vigilance sanitaire, ou en économie de la santé pour le suivi des parcours des patients (pour ne citer que ces deux exemples).

Comment une base supposée anonyme peut-elle comporter des données personnelles ? Il ne suffit pas de remplacer le nom et le NIR des personnes par un identifiant chiffré ou aléatoire (le projet de règlement européen sur la protection des données personnelles qualifie cet identifiant de *pseudonyme*) pour garantir l'anonymat des données. Ainsi, les données individuelles *détaillées* du

³⁴ Dans le cas du SNIIRAM et du PMSI, le chiffrage utilise un algorithme appelé FOIN (fonction d'occultation des informations nominatives) recommandé par la CNIL (voir sur le site de la CNIL le *Guide de sécurité des données*, la fiche n° 17, pp. 42 et suivantes). L'algorithme est utilisé deux fois avec des clés différentes détenues par des personnes différentes, de sorte que nul ne puisse constituer une table de correspondance entre les identifiants de départ et les numéros d'anonymat.

³⁵ On notera que l'article L. 6113-8 du CSP, calqué sur l'article 63 de la LIL, est plus prudent s'agissant du PMSI : « *Les destinataires des informations [...] mettent en œuvre, sous le contrôle de l'Etat au plan national et des agences au plan régional, un système commun d'informations respectant l'anonymat des patients, ou, à défaut, ne comportant ni leur nom, ni leur prénom, ni leur numéro d'inscription au Répertoire national d'identification des personnes physiques, [...].* »

PMSI et du SNIIRAM sont-elles bien anonymes prises une par une, en ce sens qu'elles ne comportent pas l'identité des personnes³⁶, mais le SI ne peut pas être en accès libre parce qu'en croisant certaines informations qui y figurent, on peut identifier des personnes *connues par ailleurs* (des proches, des collègues ou des célébrités).

Par exemple, dès lors que le système rassemble tous les épisodes de soins d'une personne, si un tiers ayant accès à la base dispose d'une connaissance, acquise par ailleurs, de certains de ces épisodes et si ceux-ci sont suffisamment discriminants, il est possible de repérer la personne concernée dans l'ensemble des enregistrements et d'accéder ainsi à des informations supplémentaires sur sa santé (diagnostics, médicaments pris) au mépris du respect dû à la vie privée.

Ce risque est d'autant plus grand que les informations rassemblées sous l'identifiant commun sont nombreuses. Ainsi l'appariement soins de ville/hôpital augmente les risques. Les possibilités qu'un tiers ait connaissance de certains épisodes de soins significatifs augmentent lorsque que les épisodes de soins accessibles sous un même identifiant sont plus complets : plusieurs hommes nés le même mois de la même année ont pu être hospitalisés 3 jours en février 2010 à l'hôpital X ; mais il n'y en a sans doute qu'un seul qui ait été hospitalisé 6 mois plus tard pendant 10 jours à l'hôpital Y et ait dans l'intervalle consulté plusieurs fois le Dr. Z.. De même, plus la durée couverte par les données accessibles est grande, plus la probabilité est importante qu'un tiers puisse s'appuyer sur la connaissance d'épisodes de soins significatifs pour procéder à une réidentification.

Le risque de réidentification s'accroît avec la précision des données se rapportant à une même personne³⁷. Il est plus facile de repérer une personne dont on connaît certains épisodes de soins si les dates de soins sont exprimées en jour/mois/année que si les données accessibles se limitent à une date en mois/année ou en année. Le risque est diminué si l'on floute (rend moins précises) certaines données. C'est pourquoi la CNIL accorde une protection particulière à certaines données, dites « variables sensibles » : code postal du domicile, dates exactes (j/m/a) des soins, date (m/a) de naissance, date (j/m/a) de décès qui facilitent en effet la réidentification surtout lorsqu'il est possible de les croiser (i.e. en utiliser plusieurs à la fois). Un tel croisement doit faire l'objet d'une autorisation spéciale.

Si on prend l'exemple des séjours hospitaliers du PMSI (MCO), avec pour seules informations l'hôpital (le code FINESS), le code géographique du domicile³⁸, l'âge (en mois et année de naissance), le sexe, le mois de sortie et la durée du séjour, 89 % des personnes hospitalisées dans l'année 2008 (et 100 % des personnes hospitalisées deux fois cette année) sont théoriquement identifiables, à condition bien sûr de connaître pour chacune de ces personnes les informations en question³⁹.

Le risque diminue ou disparaît lorsque les données sont agrégées. Il n'y a plus alors que des groupes et non des individus. Dans ce cas, le risque d'une éventuelle réidentification dépend de la granularité de l'agrégation et des possibilités de recoupements. Si par exemple, on dispose des épisodes de soins

³⁶ Les feuilles de soins, étant destinées aux remboursements et aux contrôles, sont bien sûr nominatives et comportent le NIR de l'assuré.

³⁷ C'est pourquoi les données chaînées du SNIIRAM incluant le PMSI présentent un risque de réidentification plus grand que les données du PMSI seul, ou du SNIIRAM sans données hospitalières.

³⁸ C'est le code postal du domicile ou un code postal d'une commune proche, figurant dans une table de correspondance officielle, de sorte que chaque code corresponde à une zone géographique d'au moins 5000 habitants.

³⁹ Présentation de Dominique Blum et Gilles Trouessin le 27 janvier 2012 à l'Association française des correspondants à la protection des données à caractère personnel.

(actes, séjours hospitalier) sur un agrégat de 100 000 personnes, la possibilité d'identifier une personne qui a bénéficié de tel ou tel soin est nulle. Si on dispose des mêmes données sur un agrégat de 10 personnes (par exemple, 10 habitants d'une zone géographique), il est sûrement plus facile, dès lors que l'on a accès par ailleurs à d'autres informations sur ces personnes, d'identifier une des personnes ayant reçu tel ou tel soin.

Le risque de réidentification est d'autant plus grand que la base est exhaustive : a contrario, il est réduit si l'accès n'est donné qu'à un échantillon. Si je sais que M. Untel âgé de 70 ans est décédé tel jour à l'hôpital X et si dans une base exhaustive une seule personne est dans ce cas, cette personne est M. Untel. Je ne peux parvenir à cette conclusion si je n'ai accès qu'à un échantillon⁴⁰. Ainsi dans le cas de l'échantillon généraliste de bénéficiaires (EGB), la personne qu'un tiers souhaiterait identifier n'a qu'une chance sur 100 de se trouver dans la base. De plus, même si un tiers repère dans l'échantillon une personne présentant les caractéristiques (dates de séjours hospitaliers et de décès par exemple) qu'il reconnaît, il ne peut avoir l'assurance qu'il n'existe pas dans la base complète une autre personne présentant ces mêmes caractéristiques. Mais il faut prendre garde là aussi aux recoupements. Par exemple l'EGB seul ne permet pas de reconnaître des personnes mais si le tiers dispose par ailleurs d'une base PMSI avec l'intégralité des séjours, il peut éventuellement constater que M. Untel, présent dans les deux ensembles, est bien la seule personne ayant les caractéristiques connues (dates de séjours hospitaliers...) et en déduire toute son histoire médicale.

Enfin, plus le nombre de personnes accédant aux données est grand, plus grand est également le risque de réidentification : plus nombreuses sont les personnes qui peuvent en reconnaître d'autres dans la base et avoir un intérêt à leur nuire, plus nombreuses sont les personnes qui peuvent être reconnues et plus grands sont les risques de fuite et d'indiscrétion.

Cela étant, le risque de réidentification évoqué ci-dessus (un tiers ayant accès à la base y recherche, à partir d'informations qu'il connaît déjà, des données supplémentaires sur la santé d'un proche, d'un collègue ou d'une personnalité) reste un risque opportuniste, ponctuel, concernant à chaque fois un petit nombre de personnes. Il ne s'agit pas d'un risque de dévoilement systématique des identités des personnes présentes dans la base.

Une entreprise de réidentification systématique, industrielle, pour, par exemple, par rapprochement avec d'autres fichiers, constituer un fichier nominatif de personnes atteintes de telle ou telle maladie, est très peu probable :

- Techniquement, elle exigerait ou bien de récupérer les clés de chiffrement du NIR ou de disposer d'une base nominative regroupant elle-même des informations très détaillées relatives aux soins. Constituer un tel fichier et le rapprocher du SI supposerait des moyens que seule une organisation structurée peut rassembler ;
- Or, ces pratiques sont difficiles à garder secrètes et la crainte des sanctions pénales attachées à une telle entreprise semble suffisante pour qu'aucune institution ne prenne ce risque.

La menace de la sanction vaut aussi pour un opérateur ayant accès à la base qui procéderait à une réidentification opportuniste. Toutefois, dans ce cas, la dissuasion peut apparaître insuffisante (la réidentification opportuniste peut être discrète, difficile à détecter, ou couverte par le secret des

⁴⁰ C'est pourquoi l'accès à l'Échantillon Généraliste des Bénéficiaires (EGB) du SNIIRAM peut, sous certaines hypothèses, être classé avec les données agrégées, dans la catégorie des données anonymes.

sources dans le cas d'un journaliste). Par ailleurs, la valeur attachée au respect de la vie privée est telle que ce droit mérite d'être protégé a priori par une restriction à l'accès aux données et pas seulement a posteriori par une éventuelle sanction.

Même si le risque de réidentification ne doit pas être majoré, il est indéniable que ce risque existe et ne permet pas d'envisager un accès libre à l'ensemble des données du SI. Il ne fait aucun doute en particulier, au vu de ce qui précède et aux dires des experts rencontrés par la mission, que les données individuelles exhaustives du SI (le DCIR), associant soins hospitaliers et ambulatoires, permettent d'identifier une grande proportion des personnes présentes dans la base pour qui on connaît *même approximativement* l'âge, l'adresse, la nature et la date de certains soins et éventuellement la date du décès, de sorte qu'un tiers disposant de ces informations peut apprendre pour quelles maladies ces personnes ont reçu des soins. L'interdiction des requêtes dont le résultat comprendrait moins de N personnes (avec diverses valeurs possibles pour N) ne change pas non plus le problème si par hypothèse l'utilisateur de la base dispose d'un accès aux données individuelles.

Il convient de noter toutefois que ce risque diminue jusqu'à devenir très faible voire nul selon les données mises à disposition (occultation de certaines variables, floutage de certaines données, limitation des possibilités de chaînage entre secteurs ou dans le temps, présentation des données sous forme de tableau agrégé...).

Données personnelles et données personnelles de santé

« Constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres » (article 2 de la LIL).

En l'absence de définition légale, on considère généralement qu'une donnée de santé à caractère personnel est une donnée susceptible de révéler une maladie de la personne. La définition contenue dans le projet de règlement européen sur la protection des données est plus étendue : « *toute information relative à la santé physique ou mentale d'une personne, ou à la prestation de services de santé à cette personne* » (article 4 du projet de règlement).

Le principe

On entend dire parfois que la protection juridique des données personnelles de santé est excessive ou que c'est surtout un prétexte à l'accaparement de l'information par les administrations...

- puisqu'il n'y a pas eu en France, à ce jour et pour autant que l'on sache, de fuites en provenance des bases de données médico-administratives. *Mais ce constat ne justifie pas qu'on lève les protections qui ont contribué à ce résultat ;*
- parce que les malades seraient moins soucieux de secret, et plus nombreux à exposer publiquement sur Internet leurs maladies et leurs expériences des soins contribuant ainsi à améliorer les pratiques thérapeutiques. *Mais ce comportement n'autorise pas à disposer des données des personnes qui ne souhaitent pas dévoiler des informations sur leur santé ;*
- et parce que, enfin, l'intérêt collectif justifierait qu'on prenne le risque de dévoiler des informations de santé quand ce risque ne concerne qu'un petit nombre de personnes. *Cependant, sacrifier la vie privée de quelques-uns sur l'autel du plus grand bien, notamment du progrès économique, est une conception qui peut avoir des partisans minoritaires, mais n'est pas celle sur laquelle est fondé notre droit.*

Le droit

- En France et en Europe, les données relatives à la santé figurent parmi celles qu'il est *interdit*, sauf exceptions strictement définies, de collecter ou de traiter : « *origines raciales ou ethniques, opinions politiques, philosophiques ou religieuses ou appartenance syndicale des personnes, ou qui sont relatives à leur santé ou à leur vie sexuelle* » (LIL article 8).
- Ni en France ni en Europe, il n'est toléré que des informations sur la santé des personnes puissent être rendues publiques sans leur accord. Qu'elles résultent de la malveillance ou de la négligence, ces révélations constituent des menaces pour la vie privée, les intérêts matériels et les libertés. Elles sont passibles de sanctions pénales :
 - o articles 226-13 du Code pénal (secret professionnel : 1 an d'emprisonnement et 15 000€ d'amende) ;
 - o articles 226-16 à 226-24 du Code pénal (fichiers et informatique : 5 ans d'emprisonnement et 300 000€ d'amende).

Cela ne signifie pas que l'accès aux données personnelles de santé soit interdit par la Loi informatique et libertés mais (cf. notamment les articles 63 et 64) :

- qu'il doit être justifié,
- qu'il doit être limité à un petit nombre de personnes et entouré de garanties appropriées,
- et doit être autorisé par la CNIL (ou permis par la loi, ou autorisé par les intéressés).

Il existe toute une littérature scientifique sur la question de la *ré-identification*⁴¹ montrant que la frontière entre données personnelles et données anonymes est difficile à déterminer. La directive européenne sur la protection des données de 1995 ainsi que le projet de règlement européen posent en principe que la possibilité d'identification doit être appréciée en considérant « *les moyens raisonnablement susceptibles d'être utilisés par le responsable du traitement ou par toute autre personne*. Le législateur français est resté muet à ce sujet, laissant de fait à la CNIL et au juge le soin d'apprécier.

⁴¹ Voir notamment le Guide *Sécurité des données personnelles* de la CNIL ou le document de présentation du réseau *Quetelet* (http://www.reseau-quetelet.cnrs.fr/spip/article.php3?id_article=176), mais aussi des travaux comme ceux de Daniel C. Barth-Jones, un épidémiologiste américain, ou la revue de littérature de Prada, Sergio I & Gonzalez, Claudia & Borton, Joshua & Fernandes-Huessy, Johannes & Holden, Craig & Hair, Elizabeth & Mulcahy, Tim, 2011. "Avoiding disclosure of individually identifiable health information: a literature review", MPRA Paper 35463, University Library of Munich, Allemagne

2.2 Les accès au SI : état des lieux et difficultés

Les règles actuelles d'accès au SNIIRAM sont touffues et contestées. Au regard des données personnelles de santé, certaines dispositions peuvent paraître peu protectrices, d'autres au contraire paraissent élever des obstacles excessifs aux besoins de la recherche ou de l'information du public, d'autres enfin semblent relever surtout d'une logique de réciprocité entre les organismes concernés ou de méfiance à l'égard du mauvais usage qui pourrait être fait des données.

Le tout est régi par des dispositions spécifiques et par divers articles de la loi Informatique et Libertés (LIL), laquelle opère une distinction curieuse entre les « traitements de données de santé à caractère personnel ayant pour fin la recherche dans le domaine de la santé » (chapitre IX de la LIL) et les « traitements des données de santé à caractère personnel à des fins d'évaluation des pratiques ou des activités de soins et de prévention » qui font l'objet du chapitre X, lequel vise plus spécifiquement les accès au SNIIRAM et au PMSI, comme si ces bases, initialement destinées à l'évaluation qualitative et quantitative des activités de soins, ne pouvaient pas servir aussi à des travaux statistiques d'études et de recherche plus larges.

Dans cet ensemble de dispositions, il convient de distinguer la procédure par laquelle certains organismes obtiennent des droits d'accès permanents à certaines catégories de données du SNIIRAM (221) et les procédures donnant à certains organismes un droit d'accès ponctuel à certaines données du SI, qui s'appuient sur des dispositifs juridiques variés dont la caractéristique commune est que l'accès aux données doit être in fine autorisé par la CNIL (222).

A côté de ces règles d'accès aux données personnelles du SI, il n'est rien prévu de spécial pour favoriser la mise à disposition, que ce soit à titre gratuit ou onéreux, des données anonymes.

221 Les droits d'accès permanents et l'arrêté SNIIRAM

La liste des organismes disposant d'un accès permanent aux différentes catégories de données du SNIIRAM est fixée par un arrêté ministériel⁴² pris après un avis motivé de la CNIL, prévu à l'article L. 161-28-1 du CSS. Le but initial de cet arrêté était d'approuver un protocole passé entre les caisses nationales d'assurance maladie portant sur l'organisation du SNIIRAM. Le protocole a prévu un « Comité de pilotage inter-régimes » ou « COPIIR du SNIIRAM » de 23 membres, où siègent l'assurance maladie, les représentants des professionnels de santé et l'État (avec chacun un tiers des voix), les autres organismes – dont l'IDS – ayant un statut d'observateurs. Il se réunit deux fois par an en moyenne pour proposer des évolutions du protocole et une réponse aux demandes d'accès.

Le législateur est intervenu directement pour prévoir un accès des Agences régionales de santé au SNIIRAM⁴³. L'arrêté précise les modalités d'accès des ARS.

Ainsi les droits permanents d'accès au SNIIRAM sont fixés :

- par la loi pour les ARS et le GIP études en santé publique,
- par le ministre chargé de la sécurité sociale pour les autres organismes, sur proposition du COPIIR SNIIRAM (assurance maladie + professionnels de santé + Etat),

⁴² Cet arrêté est signé du ministre chargé de la sécurité sociale. Le ministre chargé de la santé (il peut s'agir de deux ministres différents) n'est pas mentionné.

⁴³ L'accès des ARS au SNIIRAM a été prévu par l'article L. 1435-6 CSP (loi HPST du 21/07/2009). On doit citer également parmi les lois prévoyant un accès au SNIIRAM l'article L. 5121-28 CSP (loi du 29/12/2011) créant un GIP « études en santé publique » pour lequel les décrets d'application n'ont pas été pris (voir chapitre 2 § 321).

- et avec comme condition l'accord⁴⁴ de la CNIL.

Les arrêtés relatifs à la mise en œuvre du SNIIRAM

Depuis 2002, année de parution du premier arrêté de mise en œuvre du SNIIRAM, six arrêtés successifs ont élargi le périmètre initial des données disponibles et la liste des organismes habilités. Le dernier arrêté (reproduit en Annexe 11) est celui du 19 juillet 2013.

- Le premier arrêté, du 11 avril 2002, prévoyait déjà qu'outre l'assurance maladie (avec des droits différents pour les médecins conseil et les personnels administratifs), les services de l'Etat, les ARH et les unions régionales de médecins libéraux avaient accès aux données, mais uniquement sous forme de statistiques agrégées pour ce qui concerne les bénéficiaires.
En ce qui concerne les données individuelles sur les bénéficiaires (anonymisés), tout accès à des données était conditionné dans cette première période par un accord spécifique de la CNIL (sauf pour les régimes d'assurance maladie).
- En 2005, un nouvel arrêté a élargi la liste des utilisateurs potentiels (agences sanitaires, chercheurs INSERM, CNRS, IRDES, Haut conseil pour l'avenir de l'assurance maladie, CNSA), pour l'accès non seulement aux données agrégées, mais aussi à des données individuelles (anonymes) sur un échantillon au centième de la population, dont la constitution a été autorisée par ce même arrêté, avec un historique de 20 ans (Echantillon Généraliste de Bénéficiaires, EGB).
- Ce même arrêté a pris en compte la création, par la loi d'août 2004, de l'Institut des données de santé (IDS), en lui conférant deux missions :
 - o l'approbation préalable de nouveaux échantillons, ce qui permet aux organismes ayant déjà accès à l'échantillon généraliste de faire également des demandes d'extractions ad hoc à partir de l'entrepôt,
 - o l'approbation de demandes d'accès formulées par des organismes de recherche, universités, écoles ou structures d'enseignement liés à la recherche autres que l'INSERM, le CNRS ou l'IRDES (par exemple des universités, des services de CHU,...).
- L'arrêté de 2005 a également ouvert des accès (aux données agrégées et à l'EGB) à tous les membres de l'IDS, qui incluent l'Union nationale des professions de santé (UNPS), le collectif inter-associatif sur la santé (CISS), l'Union des organismes d'assurance maladie complémentaire (UNOCAM) et les fédérations hospitalières (FHF, FHP, FEHAP, Unicancer).
Depuis, l'accès aux données agrégées a été élargi aux adhérents des membres de l'IDS : donc par exemple tous les hôpitaux et cliniques, toutes les associations membres du CISS, tous les organismes d'assurance complémentaire membres de l'UNOCAM (FFSA, CTIP et FNMF), tous les syndicats de professionnels libéraux au titre de leur appartenance à l'UNPS.
- L'arrêté de 2005 précise enfin qu'aucun organisme de recherche, université, école poursuivant un but lucratif ne peut accéder au SNIIRAM.
- Depuis, cinq nouveaux arrêtés ont été publiés en 2006, 2008, 2011, 2012 et 2013, qui ont progressivement élargi les accès.

⁴⁴ Le ministère qui pourrait théoriquement aller contre l'avis de la CNIL, n'a jamais fait usage de cette possibilité du fait du risque de censure de l'arrêté par le Conseil d'Etat.

Si on récapitule la situation actuelle des accès en application de l'arrêté de 2013 :

- o Accèdent aux données agrégées (pas de données individualisées par bénéficiaire) : les Unions régionales de professionnels de santé, les syndicats de professionnels libéraux membres de l'UNPS, les associations adhérentes au CISS, les établissements hospitaliers (tous statuts confondus), ainsi que quelques organismes complémentaires participant au projet MONACO, de faisabilité d'une base de données commune avec l'assurance maladie obligatoire⁴⁵ ;
- o Accèdent aux données agrégées et à l'EGB : le ministère de la santé, le ministère de l'économie et des finances, le ministère de l'agriculture, les ARS (tout le personnel), l'ATIH, l'ANSM, l'ABM, le HCAAM, l'IDS, l'UNPS, l'IRDES, le CETAF, l'INSERM, le CNRS, la HAS, l'INCA, le fonds CMU, l'OFDT, l'UNOCAM, les fédérations de complémentaires (FNMF, FFSA, CTIP), le CISS, les fédérations hospitalières et l'UNPS ;
- o Accèdent à l'ensemble des données du SNIIRAM (DCIR, avec la limitation sur le croisement des données sensibles pour tous ces organismes sauf pour des personnels placés sous l'autorité de médecins au sein du service médical des caisses et, à titre expérimental, à l'InVS) : les régimes de base d'assurance maladie, la CNSA, l'InVS, la HAS, l'ANSM et, sur leur champ de compétence régionale, les médecins des ARS.
- o Tous les organismes non lucratifs peuvent obtenir un accès temporaire aux données agrégées ou à l'EGB après approbation de l'IDS, ou obtenir des extractions ad hoc après approbation par l'IDS et accord de la CNIL⁴⁶.

S'agissant des accès aux données du PMSI (auprès de l'ATIH), les dispositions du chapitre X de la LIL s'appliquent directement (elles sont reprises mot pour mot dans le code de la santé publique) et seul l'accord de la CNIL est nécessaire :

- Les données de l'ATIH sont transmises automatiquement, chaque mois ou chaque année, aux « partenaires » (services ministériels, ARS, assurance maladie et fédérations d'établissements) ;
- Il n'existe pas de distinction selon la nature des organismes quant aux droits d'accès (tous sont soumis à l'accord préalable de la CNIL), la différence étant seulement que les organismes publics peuvent disposer des données gratuitement⁴⁷.

⁴⁵ Le projet *Méthodes Outils et Normes pour la mise en commun de données de l'Assurance maladie Complémentaire et Obligatoire* (MONACO) est mené par l'IDS avec comme maîtrise d'œuvre l'IRDES. C'est un test d'appariement de données entre AMO et AMC portant sur 2500 personnes environ. Les Organismes d'assurance maladie complémentaire participant au projet MONACO sont AG2R, AXA, EOVI, GROUPAMA, Malakoff-Médéric, mutuelle OCIANE, Swiss Life, GMA.

⁴⁶ L'historique des arrêtés SNIIRAM et ce récapitulatif sont repris d'une note de la CNAMTS.

⁴⁷ L'ATIH communique sous licence, pour une somme comprise entre 20 et 30 K€ (selon les options choisies) un disque comportant l'ensemble des données individuelles de séjours hospitaliers d'une année. L'acquisition de ces données est subordonnée à l'accord de la CNIL. Les données sont gratuites pour les administrations publiques et les chercheurs. Sans compter les demandes d'extractions de données, près de 300 copies de la base complète ont été diffusées en 2012, selon l'ATIH, dont un peu plus de la moitié aux « partenaires » et 139 à des demandeurs « externes » (c'est-à-dire privés).

Liste des organismes ayant accès aux bases de données du SNIIRAM (arrêté du 19 juillet 2013)				
	Suivi des dépenses	Suivi de l'activité par offreur de soins (1)	Echantillon généraliste des bénéficiaires	Données individuelles exhaustives (DCIR +PMSI)
Régimes de base d'assurance maladie	X	X	X	X
Ministères (santé, sécurité sociale, agriculture, finances)	X	X	X	
dont Direction de la Recherche des Etudes, de l'Evaluation et des Statistiques (DREES)	X	X	X	Attente accord CNIL et arrêté
Agences Régionales de Santé (ARS)	X	X	X	X (pour médecins et sur périmètre régional)
Haute autorité de santé (HAS)	X	X	X	X
Agence nationale de sécurité du médicament (ANSM)	X	X	X	X
Institut de veille sanitaire (InVS)	X	X	X	X
Agence technique de l'information sur l'hospitalisation (ATIH)	X	X	X	Attente accord CNIL et arrêté
Agence de la biomédecine	X	X	X	
Haut conseil pour l'avenir de l'assurance maladie (HCAAM)	X	X	X	
Institut national du cancer (INCa)	X	X	X	
Fonds CMU	X	X	X	
Institut des données de santé (IDS)	X	X	X	
Union nationale des professions de santé (UNPS)	X	X	X	
Syndicats de professionnels libéraux membres de l'UNPS	X	X		
Unions régionales des professions de santé (toutes professions)	X	X		
Fédérations hospitalières (FHF, FHP, FEHAP)	X	X	X	
Etablissements membres des fédérations hospitalières	X	X		
Union Nationale des Organismes Complémentaires d'Assurance Maladie (UNOCAM)	X	X	X	
Fédérations constitutives de l'UNOCAM (FNMF, CTIP, FFSa) signataires de la charte d'engagement entre l'assurance maladie obligatoire et les assureurs maladie complémentaires	X	X	X	
Organismes complémentaires contributeurs de données membres des fédérations constitutives de l'UNOCAM signataires de la charte d'engagement entre l'assurance maladie obligatoire et les assureurs maladie complémentaires	X	X	X	
Collectif interassociatif sur la santé CISS	X	X	X	
Associations adhérentes du CISS	X	X		
Institut national de la santé et de la recherche médicale (INSERM)	X	X	X	
Centre national de la recherche scientifique (CNRS)	X	X	X	
Institut de Recherche et Documentation en Economie de la Santé (IRDES)	X	X	X	
Centre technique d'appui et de formation des centres d'examen de santé (CETAF)	X	X	X	
Observatoire des drogues et toxicomanies	X	X	X	
(1) identification nominative des professionnels réservée aux régimes d'assurance maladie, géolocalisation à l'IRIS pour l'INVS et médecins des ARS (URPS en attente accord CNIL et arrêté), au département pour les autres organismes.				

Source CNAMTS

Ce dispositif est peu efficace et peu cohérent.

Il est peu efficace notamment parce que les procédures d'autorisation via l'arrêté SNIIRAM, faisant intervenir un grand nombre d'acteurs et d'instances (membres du COPIIR, CNIL, ministre chargé de la sécurité sociale), prennent un temps très long même en cas d'urgence sanitaire : il a fallu ainsi plus de deux ans pour que l'ANSM obtienne le droit d'accéder aux données du SNIIRAM, accès qui lui est pourtant indispensable pour évaluer les risques associés à l'usage en vie réelle de certains médicaments ou produits de santé ; dans l'intervalle, les services de la CNAMTS ont mené eux-mêmes les plus urgentes des études exigées mais cette lenteur démontre à elle seule que la procédure est inadéquate.

Il est peu cohérent parce que, théoriquement conçu pour protéger des données personnelles, il limite aussi l'accès aux données agrégées et à l'EGB, réputés anonymes, pour lesquels il n'a pas été envisagé jusqu'à présent qu'ils soient mis en accès libre, ni à titre gratuit ni à titre onéreux. La voie d'accès à ces données prévue via l'IDS exclut les acteurs privés à but lucratif, en dehors du cas des organismes d'assurance maladie complémentaire. Pourtant il n'est pas interdit à des organismes non lucratifs d'utiliser le SNIIRAM pour des études commanditées et financées par les laboratoires pharmaceutiques. En ce cas, il est seulement demandé que les finalités de l'étude et son financement par un industriel soient mentionnés dans le dossier de demande. Autant il paraît légitime et nécessaire que les accès aux données définies comme clairement anonymes soient facilités, autant l'accès permanent d'un nombre croissant d'organismes aux données indirectement nominatives du SI (par exemple si l'accès au DCIR était systématiquement accordé à tous les laboratoires des CHU), serait de nature à compromettre la confidentialité des données et la confiance du public dans leur protection.

Il est peu cohérent aussi parce que la CNIL et le ministère de la santé ont autorisé l'ATIH à pratiquer une politique sensiblement plus libérale quant à la diffusion des bases de données PMSI. Il est certes indiqué à l'article R. 6113-31 CSP que « *Les données concernant les personnes physiques ne peuvent être communiquées que sous forme de statistiques agrégées et d'informations constituées de telle sorte que ces personnes ne puissent être identifiées* » ; et en application des articles 62 et suivants de la Loi Informatique et Libertés (chapitre X de la LIL), les demandeurs publics ou privés d'une copie des bases PMSI doivent obtenir l'accord de la CNIL, mais cet accord semble donné systématiquement⁴⁸ et il a été rapporté à la mission que l'interdiction de réaliser et de diffuser des copies est peu respectée. Il est vrai que les dates précises du séjour hospitalier sont floutées (il reste le mois de sortie, la durée du séjour et l'intervalle éventuel, en jours, entre deux hospitalisations⁴⁹). La CNIL a parfois exigé que soient rendues imprécises d'autres données telles que l'adresse du malade et son âge (code postal→département, âge→tranche d'âges), ou que soit supprimée la mention du décès, mais cela reste peu fréquent. Il n'a jamais été envisagé de définir avec l'accord de la CNIL des bases anonymes qui pourraient être ensuite diffusées librement. On doit s'interroger en tout cas sur l'existence de deux dispositifs juridiques et de deux stratégies distinctes de diffusion des mêmes données par deux organismes publics.

La distinction entre les droits d'accès des médecins et des non médecins aux données du SNIIRAM pose également un problème de cohérence⁵⁰ : un médecin – ou un autre professionnel de santé – est tenu au secret médical à l'égard des malades qu'il soigne. Lorsque, en revanche, il s'agit de données personnelles auxquelles il accède dans le cadre d'une activité administrative ou statistique, il est tenu au même secret professionnel que les autres professionnels indépendamment du métier exercé et des diplômes antérieurs (médecin, statisticien, informaticien...).

⁴⁸ Le directeur de l'ATIH n'a pas connaissance de refus. Dans quelques cas, la CNIL a exigé un floutage supplémentaire.

⁴⁹ Comme on l'a vu, cela peut permettre quand même une ré-identification si on connaît la personne par ailleurs.

⁵⁰ Cette distinction figure cependant dans la loi pour les ARS. S'agissant des médecins conseil de l'assurance maladie, l'article L. 161-29 du CSS leur donne un privilège d'accès aux données nominatives associées à un éventuel codage des pathologies (qui n'a pas été mis en œuvre à ce jour) : le SNIIRAM n'est pas visé.

L'accès des ARS au SNIIRAM

Les Agences régionales de santé, qui regroupent des services provenant à la fois de l'État et de l'assurance maladie, se sont vu reconnaître par la loi un droit d'accès au SI mais se sont vu imposer aussi un régime spécial qui a suscité un vif mécontentement des statisticiens au sein des ARS. C'est l'article L.1435-6 du CSP, issu de la loi du 21 juillet 2009, qui a réservé à des médecins des ARS l'accès aux données personnelles de santé: « *Les agents de l'agence régionale de santé ayant la qualité de médecin n'ont accès aux données de santé à caractère personnel que si elles sont strictement nécessaires à l'accomplissement de leurs missions, dans le respect de l'article 226-13 du code pénal* ».

La CNIL a cependant considéré que les médecins des ARS n'avaient pas besoin (et n'auraient donc pas le droit) de croiser les données dites sensibles (dates de soins, code postal du domicile, date de naissance en mois et année, date de décès)⁵¹. Les statisticiens des ARS venus des anciennes Unions régionales des caisses d'assurance maladie notamment, qui sont tenus au secret professionnel au même titre que les médecins et qui avaient naguère un accès au SNIIRAM (et aux bases régionales de l'assurance maladie), l'ont ainsi perdu.

Pour mémoire, un autre problème de cohérence a été soulevé à propos de l'information du public sur les tarifs moyens par acte pratiqués par les médecins. Ce sujet est traité dans la seconde partie du rapport au §2.4.

2.2.2 Les accès ponctuels au SNIIRAM

Par accès ponctuel, ou accès à la demande, on désigne des accès à des datamarts par des organismes ne disposant pas d'un droit d'accès permanent, ou des demandes d'extractions de données, quel que soit l'organisme demandeur, pour des études utilisant ces données seules ou appariées à d'autres.

Comme on l'a vu, l'arrêté SNIIRAM exclut d'emblée les demandeurs privés à but lucratif (sauf les organismes d'assurance maladie complémentaire), ce qui n'empêche pas que les demandes commanditées par un laboratoire pharmaceutique mais présentées par un laboratoire de recherche public soient recevables.

Le pouvoir d'autorisation appartient toujours *in fine* à la CNIL mais les voies par lesquelles la demande est instruite sont multiples et imbriquées de sorte qu'il n'est pas facile de savoir où s'adresser. Cette complexité traduit selon la mission un défaut de doctrine claire et l'empilement des dispositifs juridiques.

Ainsi, pour les données du SNIIRAM, peuvent être chargés d'instruire les demandes d'accès ponctuel liées à un projet de recherche, d'étude ou de travaux statistiques précis :

⁵¹ Cela signifie qu'ils peuvent accéder à l'une ou l'autre de ces catégories de données mais à une seule à la fois. L'article 48 de la LFSS pour 2013 a permis l'accès de personnels non médecins des ARS aux mêmes données que les médecins sous la responsabilité desquels ils sont alors placés, mais seulement dans le cadre des expérimentations relatives aux « personnes âgées en risque de perte d'autonomie ». L'accès des ARS aux données d'identification des professionnels de santé fera aussi l'objet d'une expérimentation.

- Ou bien le Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé (CCTIRS⁵²) pour les recherches en santé (recherches clinique et épidémiologie notamment mais aussi d'autres disciplines, telles que l'économie et les sciences sociales). Ces recherches sont visées par le chapitre IX de la LIL. Il s'agit essentiellement de recueils de données avec l'accord des personnes mais cela inclut fréquemment aussi l'appariement avec des données du SNIIRAM (le formulaire du CCTIRS demande de préciser si on aura recours au SNIIRAM). Les demandes d'appariement ou d'extraction utilisant le NIR posent un problème examiné au § 2.3 ci après. Il faut noter en outre un problème de frontière entre les compétences du CCTIRS en matière de protection des données personnelles et les compétences des Comités de protection des personnes prévus par la loi Jardé du 5/03/2012 (pour la protection des personnes dans les essais thérapeutiques) ;
- Ou bien le Conseil national de l'information statistique (CNIS) pour les services statistiques de l'État et les études qu'ils réalisent ou font réaliser (7° du II de l'article 8 et article 25 de la LIL)⁵³ ;
- Ou les services de la CNIL directement pour des demandes d'extraits des bases, visées à l'article 25 de la LIL et au chapitre X (chapitre dont relèvent le PMSI et –normalement- le SNIIRAM) ;
- Ou bien, pour les demandes ponctuelles d'accès au SNIIRAM, l'Institut des données de santé à qui ce rôle a été reconnu dans un arrêté SNIIRAM de 2005. Cela pose là aussi un problème de moyens et de compétences, voire de doctrine. En effet, l'IDS a fait savoir récemment que son comité d'experts tenait toutes les données du DCIR (sans croisement des données sensibles) et du PMSI (sans le jour exact de sortie) pour anonymes. Pour sa part, la mission considère que cette affirmation n'est pas suffisamment étayée et propose, au chapitre 2, que les risques de réidentification soient audités.

Pour mémoire, la loi du 29/12/2011 créant un GIP « études en santé publique » a prévu aussi une que ce GIP (réunissant l'État, la CNAMTS, l'ANSM, l'InVS et la HAS) pourrait autoriser directement l'accès aux données du SNIIRAM pour des études de vigilance et d'épidémiologie impliquant notamment des médicaments. D'autres voies sont possibles mais il reste incontestablement pour les pouvoirs publics et les agences sanitaires un besoin, non satisfait aujourd'hui, de disposer d'une procédure d'urgence pour associer des chercheurs qualifiés à des études liées à une crise sanitaire ou visant à la prévenir.

L'augmentation du nombre des demandes, par les chercheurs en particulier, est de nature à engorger les services de la CNIL et les services chargés de l'instruction même si les projets de recherche ayant déjà fait l'objet d'une évaluation (cas des études demandées à des laboratoires pharmaceutiques par les agences sanitaires) ou semblables à des projets-types déjà acceptés peuvent bénéficier d'une procédure allégée. Ni le CCTIRS ni encore moins l'IDS ne disposent de moyens et de compétences suffisants pour faire face à ce défi. Il n'entre pas en outre dans leur mission actuelle de se prononcer sur l'intérêt des projets de recherche ou d'étude qui leur sont soumis, ce qui interdit normalement de rejeter des projets au protocole rigoureux mais à visée de

⁵² Le CCTIRS défini à l'article 64 de la LIL est rattaché au ministre de la Recherche. Il donne un avis sur la méthodologie de la recherche, la nécessité du recours à des données personnelles et la pertinence de ces données par rapport à l'objectif de la recherche.

⁵³ Un Comité du secret statistique, instauré en 2004, examine les demandes d'accès de chercheurs à des données personnelles détenues par l'INSEE ou les services statistiques ministériels.

promotion commerciale par exemple. Le dévouement et la qualité des personnes qui participent à l'évaluation des demandes ne sont pas en cause mais dans un article de 2009 publié dans la revue *Médecine/Sciences*, Mme Bonaïti-Pellié, ancienne présidente du CCTIRS, écrivait que si tous les projets devant être présentés au CCTIRS l'étaient vraiment, ce comité ne parviendrait à faire face à la demande en raison du faible nombre d'experts (15) et du peu de moyens mis à sa disposition, alors que le nombre de dossiers a doublé entre 2007 et 2012 passant de 482 à près de 1000.

Les demandes ponctuelles donnent lieu en outre, lorsqu'elles sont acceptées, à des extractions de données, parfois volumineuses, parfois même exhaustives dans leur champ. Comme pour les fichiers de l'ATIH et malgré les engagements pris, il existe un fort risque que ces données, une fois acquises par les demandeurs, donnent lieu à des copies sauvages et soient conservées sans limite de temps. S'agissant de données auxquelles l'accès par hypothèse doit être restreint, un tel dispositif n'est pas viable.

Les chercheurs dont l'intérêt pour le SI s'est développé depuis 2010 ont été les premiers à se plaindre des difficultés d'accès aux données individuelles du SNIIRAM et plus particulièrement de la difficulté à obtenir le droit de les apparier avec des données d'enquête recueillies par eux auprès de personnes ayant pourtant donné leur accord pour cet appariement. Cet obstacle a été souligné dans les deux rapports successifs du Haut Conseil de la Santé Publique de 2009 et 2012. La difficulté est principalement due à la protection spéciale accordée au NIR dans la LIL.

2.3 Les difficultés particulières des appariements de données et le problème du NIR

Pour réaliser une étude ou une recherche, il est souvent très utile de rapprocher – apparier – des données provenant de plusieurs sources et relatives aux mêmes personnes :

- soit des données issues d'une enquête – y compris les données des très grandes cohortes épidémiologiques en population comme CONSTANCES - avec celles d'une base administrative,
- soit des données de plusieurs bases administratives entre elles.

Par exemple pour étudier l'impact des carrières professionnelles et des conditions de travail qui y sont associées, on peut apparier le SI avec la base des carrières de l'assurance vieillesse. Ou bien, pour réaliser des études très détaillées sur la santé d'une catégorie de population, on peut constituer une cohorte de personnes volontaires, auprès desquelles on recueillera régulièrement des informations, en complétant ces réponses par des données du SI, si ces personnes ont donné leur accord : les données du SI couvriront un champ plus étroit que l'enquête (par exemple la consommation de médicaments et les arrêts de travail) mais seront plus précises et remonteront plus loin dans le temps. Elles concerneront les mêmes personnes mais aussi, si c'est possible, un échantillon témoin⁵⁴ : on parle en ce cas d'appariement entre une base médico-administrative et des données d'enquête.

⁵⁴ Un problème des enquêtes c'est que les personnes acceptant d'y répondre ne sont pas représentatives : une connaissance même partielle des « non répondants » est donc très utile pour redresser les biais.

Cela suppose toutefois :

- que ce soit autorisé (par la loi et s'il y a lieu par la CNIL...),
- que les sources ou bases de données en question utilisent directement ou indirectement des identifiants identiques, pour que les données puissent bien être rattachées à la personne concernée.

Le nom, le NIR et les rapprochements de fichiers

Le NIR est un identifiant équivalent à la chaîne « Nom-Prénoms-Sexe-Date et Lieu de naissance ». Il s'en distingue essentiellement⁵⁵ par le fait qu'il est plus court et de longueur fixe, que sa transcription occasionne moins de fautes d'orthographe et que son emploi évite donc des erreurs et des corrections, surtout si les NIR utilisés ont été préalablement certifiés (c'est-à-dire si on a vérifié qu'ils sont bien conformes au NIR délivré initialement à la personne).

Le rapprochement des données concernant une même personne est techniquement facile aujourd'hui quand les deux bases sont directement nominatives (nom et prénoms, voire date et lieu de naissance, ou adresse etc.). Il peut demeurer des incertitudes sur l'orthographe, sur les noms de naissance ou d'usage, ou sur les dates de naissance mais il n'est pas difficile avec des outils informatiques modernes de rapprocher deux fichiers nominatifs et, avec un peu de temps, de lever la plupart des incertitudes. Comme le savent les utilisateurs de moteurs de recherche sur Internet, un index numérique commun n'est pas du tout nécessaire pour croiser des fichiers nominatifs ou pour effectuer une recherche dans une base de données (comme c'était le cas dans les premiers temps de l'informatique).

L'interdiction d'utiliser le NIR dans des fichiers nominatifs hors de la sphère sociale et d'en faire un index commun aux bases administratives de l'État⁵⁶ a eu autrefois une forte valeur symbolique, exprimant une défiance à l'égard des administrations publiques et de leur propension supposée à croiser leurs fichiers pour exercer un contrôle plus étroit sur les citoyens ainsi exposés dans toutes leurs dimensions (en tant que contribuables, allocataires et assurés sociaux notamment). Pour empêcher que les administrations procèdent à de tels croisements sans autorisation, la loi a restreint l'usage de l'index commun, supposé indispensable à ces croisements illicites de fichiers. S'agissant des fichiers nominatifs comme le sont normalement les fichiers de gestion des administrations (fichiers fiscaux, fichiers de police, fichiers d'allocataires par exemple), cette protection contre les croisements illicites est aujourd'hui illusoire. En revanche elle est devenue un obstacle sérieux à l'appariement de fichiers licite, respectueux des droits des personnes et utile à l'intérêt collectif.

En effet, dans les bases de données telles que le SI, constituées non pour gérer mais pour aider à la décision, l'emploi du NIR a bien sûr été interdit, de même que la mention du nom, mais le NIR a été utilisé pour produire, via un chiffrage non réversible, un numéro d'anonymat. **Autrement dit, le NIR qui n'est plus nécessaire pour apparier des bases nominatives, est devenu indispensable pour apparier des bases anonymes quand le numéro d'anonymat est produit à partir du NIR.**

En l'absence de table de correspondance conservée en lieu sûr entre NIR et numéro d'anonymat, le moyen le moins malcommode de retrouver aujourd'hui une personne dans le SI est de chiffrer de

⁵⁵ Les NIR certifiés ne comportent pas non plus d'homonymes alors qu'il peut exister quelques personnes portant les mêmes nom et prénoms, nées le même jour au même endroit.

⁵⁶ La contestation du projet Safari qui prévoyait cet emploi généralisé du NIR a été largement à l'origine du vote de la loi Informatique et Libertés du 6 janvier 1978.

nouveau son NIR pour retrouver le numéro d'anonymat⁵⁷. Or, cette opération exige un décret en Conseil d'État, y compris lorsqu'il s'agit d'apparier des données d'enquête épidémiologique avec les données relatives aux mêmes personnes contenues dans le SNIIRAM et alors que ces personnes ont donné leur accord et que, par hypothèse, la CNIL a autorisé l'enquête en question. Les chercheurs ont fait valoir à juste titre que l'exigence d'un décret en Conseil d'État dans de tels cas est un obstacle quasiment infranchissable au développement de la recherche en santé et il est admis par tous –notamment par la CNIL - que cette disposition est déraisonnable.

Les modifications de la loi du 6 janvier 1978 intervenues en 2004 ont pris davantage en compte les menaces associées aux fichiers privés, mais il est resté acquis que le NIR était un identifiant dangereux et devrait rester cantonné à la sphère sociale. Les restrictions à l'emploi du NIR ont même été rendues plus sévères encore avec les articles 27 et 29 de la LIL dans leur rédaction de 2004 : un décret en Conseil d'État⁵⁸ est nécessaire pour tout traitement nouveau utilisant le NIR, y compris pour des organismes ayant déjà le droit de l'utiliser. Mais si la finalité de ce dispositif est de faire obstacle à des croisements illicites de fichiers, n'est-ce pas directement le régime juridique des croisements de fichiers qu'il faudrait mieux encadrer ?

2.4 Une complexité et des délais excessifs

Les droits d'accès cependant ne sont qu'un aspect du problème : le traitement des données se heurte à une réelle complexité et aux insuffisances du service aux utilisateurs.

Les demandes ponctuelles d'extraction d'un échantillon demandent en moyenne, selon l'IDS, un délai de 17 mois. La CNAMTS fait observer que ces 17 mois comprennent des délais administratifs, (instruction des dossiers, avis CNIL etc.) qui interviennent en amont de la CNAMTS. Le délai de cette dernière pour extraire les données serait d'environ 6 mois, y compris les aller retour avec les demandeurs⁵⁹, ce qui est encore beaucoup. Il reste que ces extractions de plus en plus nombreuses représentent une lourde charge et que la mission de contribuer aux études et à la recherche ne peut pas être au premier plan pour une institution comme l'assurance maladie.

En outre, même pour les données agrégées du SNIIRAM et pour l'EGB auxquels un accès permanent est assez largement ouvert (via l'IDS notamment), on observe que des organismes ayant le droit d'y accéder en profitent peu (et ne profiteraient sans doute pas davantage d'un accès à des données plus détaillées). Cela tient au fait que les données du SNIIRAM sont complexes et que malgré les progrès réalisés, la manière dont elles sont organisées et mises à disposition ne facilite pas leur compréhension ni leur utilisation.

⁵⁷ L'opération inverse est impossible de sorte que ces dispositions pourraient aussi empêcher que l'on alerte directement des personnes sur un risque de santé auquel elles seraient exposées sans qu'elles le sachent, risque mis en évidence par une étude épidémiologique sur données anonymes.

⁵⁸ Rappelons qu'un décret est un acte réglementaire du Premier ministre et qu'un décret en Conseil d'État est préalablement soumis à l'avis du Conseil d'État : c'est une procédure assez lourde et chronophage pour l'administration qui ne peut se prêter à cet exercice pour tous les projets de recherche en santé nécessitant un appariement de données.

⁵⁹ Exemple : deux nouveaux anticoagulants oraux (NACO) ont fait l'objet d'une demande d'étude par la commission de la transparence, qui semble avoir explicitement incité les industriels à travailler sur le SNIIRAM. La 1ère demande date de janvier 2009 mais le protocole a dû être revu, à la demande de la HAS et sa validation a été obtenue en décembre 2012. Le délai de 4 ans ne résulte donc pas de lenteurs liées aux diverses autorisations et à l'extraction des données mais bien à la conception de l'étude et à la validation du protocole.

Quelques chiffres donnent une idée de la complexité du SNIIR-AM⁶⁰.

En termes de volumétrie, près de 20 milliards de lignes de prestation au total sont disponibles ; de plus, les données brutes transférées sont extrêmement volumineuses (un enregistrement SNIIR-AM comporte plus de 1 000 variables), ce qui pose des problèmes pour le stockage et la manipulation des données, particulièrement pour les cohortes qui peuvent être suivies pendant des décennies.

En termes d'architecture, le SNIIR-AM comporte 7 dictionnaires avec 785 000 objets (tables, index, synonymes ...) ; la partie consommation de soins du SNIIR-AM comporte 12 tables et environ 300 variables ; on trouve 75 lignes par personne en moyenne dans la table centrale *Prestations*. De plus, pour la partie PMSI (uniquement MCO) le nombre de tables et de variables change chaque année (en 2007 : environ 25 tables et 370 variables ; en 2008 : environ 30 tables et 450 variables). Pour construire une variable « utile », il faut effectuer la jointure de plusieurs tables, qui ne sont pas les mêmes selon les variables d'intérêt. Par exemple pour identifier les femmes ayant eu un frottis du col utérin sur 3 ans, il faut savoir que sur la période récente, il existe 3 codifications différentes à cause de l'évolution des codifications : depuis 2010, les cabinets d'anatomo-pathologie doivent coder les frottis du col de l'utérus avec la CCAM ; par contre, les laboratoires continuent à utiliser le code de la NABM ; il faut aussi prendre en compte les actes codés en NGAP (actes réalisés en cabinet ou établissement). Au total, pour retrouver toutes les patientes ayant eu un frottis utérin sur trois ans, 5 tables sont à interroger, plus 2 pour les identifiants, avec 3 clés de jointures différentes et dans 2 environnements (PMSI et DCIR, qui restent pour l'instant distincts).

En termes de compréhension des données, il faut connaître de nombreux aspects avant même de formuler une demande et en analyser les résultats. Par exemple : nature précise des traitements, des actes, des examens correspondant à une pathologie d'intérêt ; modalités de facturation (inclus dans les séjours hospitaliers, facturés en sus des forfaits, rétrocedés, réalisés en hôpital de jour, et/ou en hospitalisation complète, en ville, facturation différente dans le privé et dans le public ...) ; spécificités locales (DOM, régime d'Alsace Moselle) ; modalités liées aux caractéristiques sociales des personnes (CMU et AME), parfois différentes au régime général, à la MSA, au RSI ; délais de transmission des données par les établissements et les professionnels, non homogènes sur l'ensemble du territoire. De plus, les réponses à ces questions sont évolutives en fonction d'une réglementation très changeante. Il faut aussi tenir compte du fait qu'il s'agit de données de gestion destinées au remboursement et à la facturation ; il existe donc diverses données « polluantes » qui peuvent être source d'erreurs dans l'interprétation (lignes servant à annuler un remboursement, distinction entre la date de la mise à disposition dans le système et la date de réalisation de l'acte, etc.).

Une base constituée de décomptes bruts est difficilement exploitable même pour des bons connaisseurs des bases de données et de l'assurance maladie. Par exemple⁶¹, un décompte comportant 3 lignes pour une consultation chez un médecin généraliste (la consultation, une majoration de coordination, la participation forfaitaire) correspond en réalité à un acte et non à trois ; inversement, il existe dans la base, comme on l'a vu, trois codes différents pour un frottis, selon la profession ou la situation d'exercice du professionnel qui l'a réalisé.

⁶⁰ Source : présentation du projet PLASTICO par le Pr. Marcel Goldberg citant lui-même une présentation de Dominique Polton et Philippe Ricordeau (CNAMTS) aux Assises du médicament en 2011

⁶¹ Exemple emprunté à D. Polton.

Une part de cette complexité est irréductible, un prix à payer pour la richesse de la base qui ne se laisse pas connaître sans efforts et justifie que les utilisateurs s'organisent autour de pôles de compétence comme ont commencé à le faire les épidémiologistes, avec le soutien financier de la CNAMTS.

Toutefois pour l'essentiel, la longueur des délais et la complexité technique ne semblent pas des fatalités mais plutôt ce que d'aucuns qualifieront de lenteurs et d'autres de défauts de jeunesse. L'explicitation des concepts et la réorganisation de la base afin de s'affranchir des finalités et du jargon administratifs, faciliter les requêtes, l'amélioration du service de formation, de conseil et d'assistance aux utilisateurs, la mise à disposition d'outils de requêtes, un effort pour améliorer l'ergonomie du portail, tout cela relève de remèdes classiques que les services de la CNAMTS ont commencé à mettre en œuvre pour l'EGB. De même, l'industrialisation de certaines requêtes permettant la sélection et la mise à disposition sur le portail de lots de données clairement anonymes allégerait la charge et raccourcirait les délais. Mais ces évolutions supposent du temps et des moyens compte tenu de la taille et de la complexité du SI. Leur nécessité et l'investissement qu'ils exigent doivent être mieux reconnus.

Les interlocuteurs rencontrés lors de la mission s'accordent pour estimer que la richesse du SI est actuellement sous-utilisée. Cette sous-exploitation – liée aux obstacles juridiques mais aussi conceptuels et techniques - est mentionnée par les autorités chargées de l'organisation du système de soins mais également, et c'est plus problématique, par les autorités chargées de la vigilance et de la sécurité sanitaire qui considèrent que ces obstacles à l'accès aux données les empêchent d'explorer certaines hypothèses sur des dangers ou des risques sanitaires potentiels.

Chapitre 2 Les propositions

1 Les enjeux de l'accès aux données de santé

On a évoqué au chapitre 1 les risques de réidentification des personnes qui conduisent à prévoir des limitations d'accès au SI. Pour autant, l'accès aux données présente des bénéfices potentiels importants (1.1) même si on ne peut exclure des possibilités de mésusage (1.2).

1.1 Les bénéfices potentiels

En regard du risque de réidentification, il faut prendre en compte les bénéfices potentiels que l'on peut attendre d'un accès au SI. Ces bénéfices sont de plusieurs ordres : démocratique (121), sanitaire (122) et économique (123).

111 Bénéfice démocratique

Les données du SI dès lors qu'elles sont accessibles et mises en forme peuvent alimenter le débat public sur la santé. Il serait vain de prétendre énumérer l'ensemble des débats qui peuvent être éclairés par des recours aux données du SI. A titre d'exemple, le SI peut servir à documenter les inégalités territoriales de recours aux soins, les inégalités territoriales d'offre de soins, les pratiques de dépassements⁶². Au-delà de la statistique descriptive, les données du SNIIRAM constituent également le support d'études qui éclairent les controverses sur les politiques de santé.

Ces données peuvent aussi participer de la transparence de l'action publique en permettant une évaluation des résultats des politiques mises en œuvre. Encore faut-il que les politiques se formulent sous la forme d'objectifs chiffrés contrôlables, ce qui reste peu fréquent. On peut néanmoins noter que le SNIIRAM permet de constituer certains des indicateurs inclus dans les programmes de qualité et d'efficience annexés à la Loi de financement de la sécurité sociale (LFSS) et qu'il a servi également de source pour documenter certains des indicateurs associés à la loi de santé publique de 2004.

112 Bénéfice sanitaire

Dès 2006, il était reconnu que le SI « représente indiscutablement un outil au potentiel extrêmement fort dans les domaines de la pharmaco-épidémiologie et de l'évaluation post-AMM » même s'il était pointé la difficulté de se prononcer sur « la valeur réelle du système » en raison du

⁶² Cf. CISS, 60 Millions de consommateurs, Santéclair : Création de l'observatoire citoyen des restes à charge en santé. Dossier de presse 27 mai 2013. Les créateurs de l'observatoire présentent une première étude effectuée à partir de l'accès au SI dont bénéficie le CISS en tant que membre de l'IDS mais soulignent qu'ils se heurtent à la limitation des données accessibles, « propriété quasi exclusive de l'assurance maladie ».

« manque d'expérience » et en « l'absence de toute collaboration avec des équipes ou des personnes compétentes » en ces domaines⁶³.

L'étude relative au Médiateur a montré que le système d'information permet de tester certaines hypothèses sur les effets en vie réelle des médicaments⁶⁴. Le nombre des observations mobilisées dans le SI autorise la détection d'événements graves mais de très faible fréquence, de ce fait non décelables à partir des essais cliniques. Le SI peut ainsi servir de support, seul ou en lien avec des études observationnelles classiques, aux études post-inscription demandées par l'ANSM lors de la mise sur le marché des médicaments.

Si les premières études réalisées à partir du SI concernent le médicament, la démarche d'évaluation peut être élargie à d'autres domaines (matériels médicaux, efficacité des actes...) et aux risques environnementaux.

Si ce type d'études n'est possible que depuis l'appariement entre données de ville et données issues du PMSI, soit depuis 2009, le SI peut maintenant apporter une contribution majeure aux politiques de vigilance et de sécurité sanitaire et de ce fait sauver des vies⁶⁵.

Au-delà de la vigilance-sécurité le SI peut aussi servir de support à des études de santé publique : évaluation des états de santé, des facteurs de risque, des inégalités de santé.

113 Bénéfice économique

Le premier gain économique que l'on peut attendre du SNIIRAM est directement lié au précédent : délivrer des soins sûrs et de qualité est certainement la meilleure manière de maîtriser les dépenses de santé.

Au delà, le SI peut contribuer à réaliser des études de comparaison des pratiques non seulement en termes d'efficacité sanitaire mais aussi en termes de coûts économiques. Ces études d'efficience médico-économique sont encore peu développées en France. Dans le cadre de la loi Obama, elles font l'objet d'un fort investissement aux Etats-Unis où l'on place de grands espoirs dans les gains d'efficience qu'elles permettront de réaliser : un budget de 1,1 milliard de dollars a été réservé pour financer les *comparative effectiveness research*.

Le SI peut aussi contribuer à documenter les décisions sur l'organisation de notre système de soins. Ainsi s'il est admis qu'il faut organiser une médecine de parcours notamment pour les patients chroniques, l'on manque encore d'études permettant d'analyser leurs trajectoires actuelles au sein du système de soins. Le SI peut certainement contribuer à décrire les parcours réels et éclairer les voies et moyens pour les améliorer.

Le recours au SI peut s'avérer un moyen économique pour établir certains résultats qui ne peuvent aujourd'hui être approchés qu'en organisant des études en vie réelle, lourdes et coûteuses. Il peut

⁶³ Groupe de travail coordonné par B. Bégaud et D. Costigliola. La pharmaco-épidémiologie en France. Evaluation des médicaments après leur mise sur le marché. Etat des lieux et propositions. Janvier 2006.

⁶⁴ Weill A, Païta M, Tuppin P, Fagot J-P, Neumann A, et al. (2010) Benfluorex and valvular heart disease: a cohort study of a million people with diabetes mellitus. *Pharmacoepidemiol Drug Saf* 2010 19: 1256–1262

⁶⁵ Martin-Latry K, Bégaud B. Pharmacoepidemiological research using French reimbursement databases: yes we can ! *Pharmacoepidemiology and drug safety* 2010; 19: 256–265

permettre aux laboratoires pharmaceutiques de fournir à un moindre coût certaines études que les autorités publiques (DGS, ANSM, HAS et CEPS) leur demandent de réaliser⁶⁶.

Les laboratoires pharmaceutiques peuvent aussi être amenés à utiliser le SNIIRAM pour réaliser de leur propre initiative des études afin de vérifier certaines propriétés de leurs produits que seule une étude sur des grands nombres, en vie réelle peut permettre de révéler. En ce sens, la disponibilité d'un système d'information comme le SNIIRAM peut servir à développer des capacités d'étude et de recherche (pharmacologie, épidémiologie...) qui peuvent favoriser l'attractivité du territoire pour la localisation des capacités de recherche.

La disponibilité d'une telle base dans la mesure où elle peut favoriser le développement d'une recherche publique ou privée en santé (au sens large) peut constituer un atout dans la compétition internationale en matière de recherche.

L'existence d'une base de données à disposition des chercheurs peut favoriser, notamment en facilitant l'accès des doctorants à des données déjà constituées, le développement d'une communauté académique dans les domaines de l'épidémiologie, pharmacologie, santé publique, économie de la santé.

La prise de conscience des bénéfices d'une meilleure utilisation des bases de données en santé est d'ailleurs un phénomène international : « *health data constitutes a significant resource in most OECD countries and it makes economic and ethical sense to use this data as much as possible* »⁶⁷. Au Royaume-Uni, la réorganisation annoncée du dispositif d'accès aux données de santé (création du *clinical research data link*) est partie intégrante d'un *growth plan* destiné à rendre l'économie britannique plus compétitive, qui parmi les diverses mesures prévoit : « le soutien aux projets de croisement des données électroniques de santé de manière à placer le Royaume-Uni en position mondiale stratégique pour la recherche en santé »⁶⁸.

1.2 Le mésusage éventuel des données

La disponibilité d'informations constitue a priori une richesse mais on ne peut exclure que cette richesse soit galvaudée. Il existe un risque de mésusage des données sachant que par mésusage, on entend la diffusion d'informations erronées, biaisées et non le risque de ré-identification évoqué au chapitre 1.

Bien évidemment les données ne valent que par l'interprétation qui en est fournie, il ne faut pas succomber à la fascination de la donnée brute qui parlerait d'elle-même. Pour prendre des exemples évidents : une différence dans le niveau de consommation de soins entre deux zones géographiques ne signale un problème d'équité que si les données sont redressées pour tenir compte des caractéristiques de la population (âge, morbidité...) ; une différence entre deux zones géographiques en matière de densité médicale libérale doit, au moins, être pondérée par le niveau d'activité des praticiens pour apprécier le niveau de service de soins dont bénéficient les résidents de la zone. Les

⁶⁶ Bien sûr, la possibilité de mieux utiliser le système d'information pour réaliser de telles études ne devraient pas aboutir à multiplier les demandes d'études adressées aux laboratoires pharmaceutiques. Le travail entrepris pour rationaliser les demandes d'études doit en tout état de cause être poursuivi.

⁶⁷ OCDE. *Strengthening health information infrastructure for health care quality governance*. OECD Health policy Studies 2013. p. 13

⁶⁸ Department of Health. Launch of the clinical practice research data link. DOH, déc. 2001.

données ne pourront servir de support à un débat éclairé que dans la mesure où elles feront l'objet d'un traitement adéquat. La richesse potentielle du SI ne se révèle pas par le seul fait d'accéder aux données. Contempler les données du SNIIRAM n'apporte pas en soi d'informations. Au-delà des applications les plus simples de statistiques descriptives, les données du SNIIRAM ne parlent que si on leur pose de bonnes questions, si on les aborde à partir d'un protocole scientifiquement construit⁶⁹. Mettre à disposition les données n'en reste pas moins un facteur de transparence dans la mesure où cette mise à disposition favorise le pluralisme des traitements et des interprétations.

Par ailleurs, on a souligné en première partie la complexité du SI, complexité qui en elle-même peut conduire à produire des résultats erronés.

Au-delà de ces mésusages involontaires, il faut également tenir compte des mésusages intentionnels ou du moins de ceux liés aux intérêts engagés dans le résultat des traitements. La question se pose avec acuité dans la perspective d'une utilisation du SI par les industries de santé. Une littérature abondante a démontré que les études cliniques, alors même qu'elles étaient publiées dans des revues à comité de lecture, pouvaient être affectées de biais divers lorsqu'elles étaient financées par les laboratoires. Ces biais, parfois difficilement détectables, sont mis en lumière par des travaux qui montrent que les études ainsi financées sont systématiquement plus favorables, à la fois dans leurs résultats et dans leurs conclusions, aux produits étudiés que les études sans financement industriel⁷⁰.

On retiendra de ce paragraphe que les données du SI constituent une richesse tant en termes de transparence démocratique, de sécurité sanitaire, d'amélioration de la qualité des soins, de pilotage des politiques de santé que d'efficacité économique. Pour autant, exploiter cette richesse expose à des risques au regard de la protection de la vie privée. Un mésusage de cette richesse ne peut par ailleurs être exclu.

⁶⁹ Sauf à ce qu'il soit démontré que le *data mining* permet en ce domaine aussi de faire émerger du rapprochement des données des résultats pertinents.

⁷⁰ Pour une revue récente de la littérature sur ce sujet et une analyse des divers biais qui entachent les études financées par les laboratoires voir : G. Schott, H. Pachi, U. Limbach, U. Gundert-Remy, W. Ludwig, K. Lieb : The financing of drug trials by pharmaceutical companies and its consequences. *Deutsch Arztebl Int* 2010; 107(16): 279–85

2 Les principes

Les propositions concrètes relatives à l'accès au système d'information doivent s'articuler autour de quelques principes. Aussi cette partie revient sur les finalités du SI (11), sur sa « propriété » (12), sur les conditions de l'accès aux données (13) et sur la protection des données relatives aux professionnels de santé (14).

2.1 Les finalités du SI

Comme on l'a vu au chapitre 1, les finalités du SI ont été envisagées, à l'origine, principalement sous l'angle de la maîtrise des dépenses.

Les objectifs poursuivis conditionnent la perspective dans laquelle sont abordés la gestion du SI et l'accès aux données. Les objectifs assignés à l'origine en 1998, soit il y maintenant plus de 15 ans, méritent aujourd'hui d'être reformulés.

Les finalités du SNIIRAM pourraient aujourd'hui, en accord avec la pratique, être définies de manière plus large. A cet égard, il serait certainement pertinent d'affirmer que le SNIIRAM contribue :

- à mettre à disposition du public des informations pertinentes sur les politiques de santé et d'assurance maladie ;
- à améliorer la sécurité, la qualité de soins, le pilotage des politiques de santé publique par des études relatives à la vigilance sanitaire, à la pharmaco-épidémiologie, à la médico-économie, à l'évaluation des pratiques et de l'organisation des soins, à l'évaluation de l'état de santé de la population ;
- à servir de support à des recherches dans le domaine de la santé.

La mise en exergue de ces nouvelles finalités dans la loi (modification de l'article L 162-28-1) nous paraît correspondre à ce qu'il est légitime d'attendre aujourd'hui du SI et fournir un cadre plus pertinent sur les réflexions relatives à l'accès aux données.

On notera que le même article L 162-28-1 précise dans son alinéa final que « les données reçues et traitées par le SNIIRAM préservent l'anonymat des personnes ayant bénéficié des prestations de soins ». Cet alinéa est rédigé comme si le SNIIRAM désignait un système d'information (un ensemble de données) mais aussi l'opérateur d'éventuels traitements. Il est donc important que le législateur prévoie que les données rassemblées au sein du SNIIRAM peuvent dans « des conditions préservant l'anonymat des patients » et pour les finalités évoquées ci-dessus faire l'objet de traitement par des personnes dûment habilitées.

2.2 La propriété de la base

L'article L 161-28-1 du CSS créant le SNIIRAM s'insère dans une section du Code de la sécurité sociale qui porte pour titre « Système d'information de l'assurance maladie et cartes de santé ». Il prévoit que le SNIIRAM est « mis en place par les organismes d'assurance maladie » et que ses « modalités de gestion et de renseignement » sont définies conjointement par un protocole passé entre les trois principales caisses d'assurance maladie, protocole approuvé par arrêté du ministre chargé de la sécurité sociale.

De fait, le système d'information est principalement constitué à partir de l'activité des professionnels de santé et de l'activité de gestion des caisses d'assurance maladie même si, depuis qu'il intègre le PMSI, les établissements de santé et l'ATIH contribuent à son alimentation. La CNAMTS gère en pratique le SI qui est placé sous sa responsabilité.

Même si le protocole rappelle qu'il s'agit d'une « base nationale » et se borne à en confier la gestion technique à la CNAMTS, il est clair qu'une confusion s'est installée entre la gestion et la propriété de la base⁷¹.

On peut voir un exemple de cette logique propriétaire dans l'institution du COPIIR SNIIRAM qui régule les accès au SI sur la base d'un partage des voix par tiers entre Etat/assurance maladie/professionnels de santé. On peine à décrypter la logique qui préside à cette pondération des voix sauf à considérer que les propriétaires des données (l'assurance maladie qui les gère et les professionnels de santé qui les fournissent) partagent avec l'Etat le soin de réguler les accès à la base.

Il est donc nécessaire de rappeler que le SI est un « bien public » constitué en vue des bénéfices collectifs que la collectivité peut attendre de son exploitation. Il doit à ce titre être clairement placé sous l'autorité de la ministre de la santé sous le contrôle du Parlement. Bien évidemment, cette autorité ne peut s'exercer qu'en concertation avec les parties prenantes fort diverses intéressées à la gestion et au développement du système (usagers et gestionnaire du SI mais aussi représentants des diverses composantes de notre système de santé et d'assurance maladie). Aucune de ces parties prenantes ne peut revendiquer de prééminence, quelles que soient par ailleurs les responsabilités qu'elle assume dans la gestion du SI ou dans le système de santé.

Cette réaffirmation est d'autant plus justifiée que les finalités du SI dépassent désormais largement, on vient de l'indiquer, celles envisagées à l'origine (la maîtrise des dépenses) qui le situait dans le champ des partenaires conventionnels (caisses d'assurance maladie et professionnels de santé).

D'ailleurs, pour lever toute ambiguïté, le SNIIRAM pourrait être rebaptisé Système National d'Information de Santé.

⁷¹ Arrêté du 12 juillet 2012 relatif à la mise en œuvre du SNIIRAM

2.3 Les conditions d'accès au SI

Caractériser le SI comme un bien public, comme le « bien de tous » ne signifie pas pour autant que tous doivent pouvoir accéder aux données. Les conditions d'accès à la base doivent prendre en compte le risque de réidentification des patients et des professionnels, le risque de mésusage mais également les bénéfices publics -démocratiques, sanitaires et économiques- liés à l'accès à la base.

Il paraît nécessaire de revenir sur ces bénéfices et ces risques notamment dans la mesure où s'exprime parfois le soupçon que le risque de réidentification soit invoqué pour restreindre l'accès aux données alors que ce risque serait mis en avant par crainte exagérée d'un mésusage ou pour protéger les intérêts stratégiques propres des acteurs « propriétaires » des données.

La réidentification d'un patient constituerait la mise en cause d'un droit fondamental (respect de la vie privée, de l'intimité). Personne ne se situe dans une perspective utilitariste où le bénéfice collectif justifierait une atteinte – un risque d'atteinte en l'occurrence – à un intérêt privé considéré comme fondamental.

Toutefois, il est clair que des sensibilités différentes peuvent s'exprimer au plan des principes quant à l'équilibre à viser entre le risque d'atteinte à un droit fondamental et la perspective d'un bénéfice collectif. Pour certains un bénéfice collectif justifie une prise de risque maîtrisée, pour d'autres aucun bénéfice collectif ne peut justifier le moindre risque. Différences que l'on retrouve au plan international, l'OCDE note des différences entre pays « dans l'application du principe de respect de la vie privée » ainsi que dans la conception de « l'équilibre entre les droits individuels à la vie privée et les droits collectifs à des soins sûrs et efficaces et à un système de santé efficient »⁷².

Ce débat sur les valeurs est redoublé par un débat de fait sur la portée réelle du risque et sur l'étendue des bénéfices. Ni l'étendue du risque, ni les bénéfices attendus ne sont, ne peuvent être, précisément caractérisés.

En général, les arguments de valeur et de fait sont mêlés : ceux enclins à valoriser la perspective de bénéfices collectifs par rapport aux risques d'atteinte à un droit individuel seront aussi ceux qui minimiseront l'ampleur du risque et souligneront l'importance des bénéfices collectifs. Inversement, ceux qui, au plan des principes, considèrent que les droits individuels doivent l'emporter sur la perspective de bénéfices collectifs seront les plus enclins à considérer que les risques sont importants et les bénéfices limités.

Ce n'est pas l'objet de ce rapport, bien sûr, de prétendre trancher entre des positions de principe. Il s'agit simplement de tenter de proposer une organisation de l'accès qui puisse refléter un point d'équilibre correspondant à l'appréciation commune sur l'équilibre à instaurer entre préservation d'un droit individuel fondamental et bénéfices publics.

A cet égard, il semble largement admis que le risque de réidentification dès lors qu'il est avéré justifie des restrictions a priori à l'accès aux données. Ce principe n'est d'ailleurs remis en cause par aucun de nos interlocuteurs.

Le risque évoqué jusqu'à présent est relatif à la vie privée des patients, le risque quant à la protection des données relatives aux professionnels n'est pas apprécié dans les mêmes termes. Au contraire, au

⁷² OCDE. *Strengthening health information infrastructure for health care quality governance*. OECD Health policy Studies 2013, p. 14

cours de la mission, certains de nos interlocuteurs ont revendiqué une transparence complète sur les données individuelles d'activité des professionnels nominativement identifiés (cette question est évoquée infra au paragraphe 24).

Si donc le risque de mettre en cause la vie privée des patients justifie des restrictions à l'accès aux données, de telles restrictions ne peuvent être justifiées par le risque de mésusage. Le risque d'un mésusage ne doit pas faire obstacle au principe de transparence de la donnée publique. Limiter l'accès aux données au titre d'un mésusage potentiel revient à ériger les pouvoirs publics en juge a priori de la capacité des acteurs de la société civile à faire un usage pertinent des données. Les organisations publiques ne peuvent revendiquer un monopole de l'information dès lors qu'aucun droit fondamental (ou aucun intérêt public majeur) n'est mis en cause par la publicité de l'information. Limiter l'accès à des données par crainte d'un éventuel mésusage aboutit à se priver de bénéfices éventuels que la collectivité pourrait tirer de l'usage.

Dès lors, il semble que les principes suivants peuvent guider la réflexion sur les conditions d'accès :

Si les données présentent un risque de réidentification : il est légitime de restreindre les accès et de contrôler les garanties apportées par les personnes qui accèdent, les finalités de l'accès (le risque pris ne se justifie que si l'intérêt public est démontré) et la pertinence des traitements envisagés au regard de la finalité.

Si les données ne sont pas indirectement nominatives : les données doivent être communicables à tous ou rendues publiques (*open data*). Le mésusage éventuel des données ne justifie pas à lui seul une restriction à l'accès.

Bien évidemment, si ces principes peuvent s'énoncer simplement, leur mise en œuvre est délicate car l'appréciation sur le caractère indirectement nominatif d'un ensemble de données peut prêter à discussion (cf. chapitre 1). Par ailleurs, il faut rappeler, comme le souligne à juste titre l'OCDE, que l'enjeu, en matière d'accès aux données de santé n'est pas d'éviter tous les risques mais de prendre des décisions informées dans un contexte d'incertitude. Si cette incertitude est inévitable, il faut prendre en compte qu'éviter de décider, ou repousser les décisions comporte également des risques en termes de sécurité des patients ou de qualité des soins⁷³.

2.4 La protection des données relatives aux professionnels de santé

Si la nécessité de protéger les données relatives aux patients ne fait pas débat, certains intervenants rencontrés dans le cadre de la mission ont revendiqué de pouvoir accéder à des données nominatives pour ce qui concerne la pratique des professionnels de santé et plus particulièrement des médecins.

La demande s'exprime ainsi : il est légitime que les patients puissent accéder à des données individuelles sur les professionnels de santé pour pouvoir faire des choix éclairés lors du recours aux soins. Par exemple, le patient doit pouvoir dans la perspective d'une prothèse de hanche accéder à l'historique du profil d'activité des chirurgiens pour pouvoir choisir celui qui pratique couramment cet acte ; accéder au profil des prescriptions des praticiens permettrait aux patients de choisir ceux

⁷³ OCDE *Ib*.p 88

dont le comportement correspond à leur souhait (gros prescripteur/petit prescripteur, gros prescripteur d'antibiotiques/petit prescripteur d'antibiotiques...).

Un parallèle est établi avec le secteur hospitalier où il est admis que des données caractérisant l'activité ou les performances de tel ou tel établissement peuvent légitimement être diffusées. Bien évidemment, ce parallèle ignore la différence fondamentale entre une institution et une personne physique.

Cette demande s'inscrit dans une tendance générale à ne pas s'en remettre aux « autorités » constituées (pouvoirs publics, ordre des médecins) censées assurer la qualité de la médecine mais à en appeler à l'intervention du consommateur, à son choix éclairé. La diffusion de telles données serait complémentaire aux modalités actuelles d'acquisition d'information pour faire des choix (relation dans le monde de santé, réseau social) basés sur la « réputation ». Il est possible de défendre que cette diffusion d'information renforce l'égalité devant les soins dans la mesure où le mode actuel d'acquisition repose sur des réseaux différenciés socialement.

Il faut souligner toutefois que la plupart des études consacrées à cette problématique soulignent un paradoxe : une forte demande d'informations comparatives mais un faible impact de ces données sur les choix effectifs des patients.

Il s'agit de savoir si un bénéfice collectif potentiel (meilleure information des patients, choix plus avisés...) justifie la transparence sur des données personnelles des professionnels de santé.

Ces données relatives à l'activité n'ont pas la même sensibilité que des données personnelles relatives à la santé : le respect de la vie privée n'englobe pas les informations relatives à la vie professionnelle. Toutefois, une transparence sur la pratique permettrait de reconstituer des données qui relèvent de la vie privée (ex : une description de toute l'activité couplée avec le montant des dépassements permet de reconstituer assez précisément, si ce n'est le montant du revenu, du moins le total des honoraires de chaque praticien).

En tout état de cause, il s'agit de données personnelles détenues par une administration publique dont l'article 13 de la loi 17 juillet 1978 fixe des conditions pour une réutilisation : « soit lorsque la personne y a consenti, soit si l'autorité détentrice est en mesure de les rendre anonymes ou, à défaut d'anonymisation, si une base législative ou réglementaire le permet ».

L'information sur les cabinets de médecine générale au Royaume-Uni

Le NHS était traditionnellement très contraignant pour les patients : les patients devaient s'inscrire dans un cabinet de médecine générale en fonction de leur résidence, l'orientation vers tel ou tel services hospitalier était décidée par le médecin généraliste.

Depuis les années 2000, le NHS a mené une politique active pour ouvrir des possibilités de choix aux patients. Il s'agit d'améliorer la satisfaction des patients et d'introduire une concurrence entre les producteurs de soins pour favoriser des progrès en matière de qualité.

En cohérence avec cette volonté de favoriser le choix des patients, le NHS diffuse des informations sur les producteurs de soins. Le site *NHS Choice* est le vecteur principal de cette politique d'information.

Pour ce qui concerne les cabinets de médecine générale, l'utilisateur britannique peut disposer à partir

de ce site, d'informations pratiques sur chaque cabinet (nom des GPs, heures d'ouverture, possibilité ou non d'accueillir de nouveaux patients, possibilité de bénéficier du service en ligne et de la prescription électronique, possibilité d'accéder en ligne au dossier médical et aux résultats d'examens...) mais aussi d'éléments sur la performance comparée des cabinets. Pour l'essentiel, il est possible de savoir comment le cabinet se situe :

- au regard des principales rubriques d'un sondage annuel qui mesure le retour d'expérience des usagers du cabinet. Ces rubriques permettent notamment de consulter l'appréciation portée par les patients sur les temps d'attente mais aussi sur les consultations tant des médecins que des infirmières ;

- au regard des principales rubriques du dispositif de paiement à la performance qui représente une part substantielle de la rémunération des cabinets. Le niveau de performance est présenté globalement pour les indicateurs relatifs à l'organisation du cabinet et pour les performances cliniques. Les performances sont détaillées pour certains domaines cliniques (asthme, diabète...). Il est vrai que cette information reste peu significative, sauf pour de rares cabinets, car le système de paiement à la performance britannique discrimine peu entre les cabinets.

Le site permet aux internautes de porter une appréciation (classement de une à cinq étoiles) et de publier des commentaires sur le cabinet.

Par ailleurs, il est possible d'accéder, sur d'autres sites, au détail des résultats de chaque cabinet pour chacun des nombreux indicateurs du paiement à la performance, ainsi que de connaître mois par mois les prescriptions cabinet par cabinet.

L'information ainsi dispensée n'est pas individuelle, le cabinet (et non le médecin) est considéré comme l'unité qui assure la responsabilité de la prise en charge des patients. Les informations ainsi dispensées concernent donc, en moyenne, un groupe d'environ cinq médecins. Mais elles peuvent être individuelles dans la mesure où il existe -même s'ils sont rares, environ 5% des GPs- des cabinets où n'exerce qu'un seul médecin.

D'autres considérations font obstacle au dévoilement de ces données personnelles. Il serait contraire au principe déontologique (« la médecine ne doit pas être pratiquée comme un commerce »⁷⁴) dont découle l'interdiction de « tous procédés directs ou indirects de publicité »⁷⁵. Cette interdiction de publicité qui s'applique aux médecins vaut, sauf à être vidée de son contenu, également pour les tiers.

Même si « la médecine ne doit pas être pratiquée comme un commerce », le secret industriel et commercial est aussi invoqué pour restreindre l'accès aux informations relatives aux médecins. Ainsi, dans le cadre d'une demande d'accès au SNIIRAM, la CADA estime après avoir noté que les données relatives aux professionnels sont indirectement nominatives, que « l'accès d'un tiers à cette base serait susceptible de porter atteinte au secret en matière commerciale ou industrielle, qui protège l'activité libérale des professionnels concernés »⁷⁶.

⁷⁴ Art. 19 du code de déontologie

⁷⁵ *Ib.*

⁷⁶ Avis rendu par la CADA en date du 19 avril 2012 sur la demande de Monsieur Jacques Servier.

On ne sait donc toujours pas si la médecine est ou non un « commerce » mais en tout état de cause, les données personnelles relatives à la pratique font l'objet, hors disposition législative spécifique, de protections multiples.

Le législateur est toutefois déjà intervenu pour permettre la publicité de certaines informations nominatives sur les professionnels de santé. L'article L 162-1-11 CSS prévoit en effet que « les caisses peuvent mettre en œuvre des services de conseil administratif ou d'orientation ». Ces services « peuvent fournir des informations [...] sur la participation des professionnels de santé à la formation continue, à la coordination des soins et à la démarche d'évaluation de la qualité professionnelle »⁷⁷ et « fournissent tous éléments d'information sur les tarifs d'honoraires habituellement demandés et toutes informations utiles à la bonne orientation du patient dans le système de soins ».

Le législateur en utilisant une formulation générale « toutes informations utiles à la bonne orientation du patient » a ouvert assez largement le champ des informations que les caisses pourraient diffuser et a fait prévaloir l'intérêt du patient sur des considérations relatives au maintien de la confidentialité des données relatives au professionnel.

Cet article constitue la base juridique des pages du site AMELI par lesquelles la CNAMTS diffuse des données nominatives sur les professionnels de santé notamment la situation conventionnelle (secteur1/secteur 2) et des données sur les tarifs moyens pratiqués.

Cet article n'ouvre qu'aux caisses le droit de diffuser de telles informations et ne se prononce pas sur les initiatives que pourraient prendre d'autres acteurs en ce domaine. Cette exclusivité pose question ; des opérateurs privés qui voulaient relayer l'information du site AMELI sur les tarifs se sont vu opposer par la CADA l'article 13 précité de la loi du 17 juillet 1978⁷⁸.

Dès lors qu'autorisée par la loi pour les caisses, la diffusion de données nominatives relatives aux dépassements ne pose plus une question de principe ; il serait donc logique et souhaitable que toute institution intéressée ait accès aux données sources sur les tarifs moyens et puisse, sous la forme qu'elle souhaite, diffuser l'information. Elargir les canaux d'information est cohérent avec l'objectif d'assurer une pleine information des patients en ce domaine.

Une disposition législative devrait donc prévoir que la CNAMTS fournit les données sources nominatives sur les tarifs pratiqués par les professionnels de santé et que les informations issues de ces données peuvent être librement réutilisées. Bien évidemment, les personnes qui réutiliseront ces données devront veiller à ce que « ces dernières ne soient pas altérées, que leur sens ne soit pas dénaturé et que leurs sources et la date de dernière mise à jour soit mentionnées »⁷⁹. Les médecins devraient pouvoir exercer un droit de rectification sur les données sources.

Faut-il pour autant accepter que d'autres informations relatives à la pratique soient dévoilées ?

On peut écarter certaines objections :

- Le dévoilement de ces données faciliterait le marketing des industriels du médicament et des matériels médicaux. Les interlocuteurs rencontrés lors de la mission considèrent unanimement

⁷⁷ On notera que ces références sont aujourd'hui datées depuis la mise en place du DPC.

⁷⁸ Avis rendu par la CADA dans sa séance du 19 avril 2012 sur la demande du conseil de la société *Le Guide Santé*.

⁷⁹ Cf. article 12 de la loi du 17 juillet 1978 relatif aux conditions de réutilisation des informations publiques.

que d'ores et déjà sur la base des informations actuellement collectées (données issues des pharmacies, panels volontaires...) et en s'appuyant sur les données recueillies par les visiteurs médicaux, il est possible d'établir des profils relativement précis des médecins. On peut d'ailleurs s'étonner de cette unanimité dès lors que le législateur est intervenu par des dispositions inscrites dans le code pénal pour restreindre les possibilités de profilage des médecins⁸⁰. Une étude spécifique sur les techniques de ciblage des médecins et sur les données qui leur servent de support serait dans ce contexte certainement utile pour apprécier si les inquiétudes exprimées par certains de nos interlocuteurs, notamment les représentants des médecins, sont fondées ;

- Le dévoilement de ces données permettrait aux organismes complémentaires de discriminer entre les professionnels de santé (conventionnement sélectif, taux de remboursement différencié en fonction de la pratique du médecin consulté...). A supposer que ce risque soit effectif, il est toujours possible si on veut y faire obstacle de prohiber par la loi de telles pratiques.

D'autres objections paraissent plus pertinentes :

- Le risque de mésusage des données. Par exemple, il n'est certainement pas souhaitable que se développe un classement des médecins basé sur leur propension à prescrire des arrêts de travail ; la même inquiétude peut être évoquée pour ce qui concerne la propension à prescrire certains médicaments recherchés pour d'autres motifs que des raisons exclusivement sanitaires.

- La faible pertinence des données qui pourraient être diffusées. Les données brutes ont peu de sens : par exemple, un profil de prescription ne commence à devenir significatif que s'il est ajusté en fonction des caractéristiques de la patientèle. Il est sûrement très optimiste de penser que le jeu de la concurrence ou les exigences du public suffiront pour que les diffuseurs de l'information consentent aux investissements nécessaires pour présenter des informations pertinentes⁸¹.

Enfin, laisser se développer une large publicité des données nominatives sur la pratique des professionnels de santé engage, on l'a vu, une position de principe sur l'exercice de la médecine : il sera difficile de permettre à des tiers de diffuser toutes informations sur les médecins sans autoriser les médecins eux-mêmes à communiquer sur leur pratique.

Les objections évoquées valent pour certaines données mais ne sont pas pertinentes pour d'autres.

Parmi les données qui pourraient faire l'objet d'une telle ouverture, on peut évoquer les indicateurs relatifs au « paiement sur objectif de santé publique ». Ce sont des indicateurs négociés qui ont recueilli l'approbation des professionnels concernés à travers leurs représentants et qui ont paru suffisamment pertinents pour justifier une rémunération particulière des médecins. On peut donc considérer qu'il serait légitime de les porter à la connaissance des patients. Encore faut-il préciser que ces indicateurs dont le spectre est restreint n'éclairent qu'une faible partie de la pratique des

⁸⁰ L'article L 4113-7 du code pénal prohibe « la constitution et l'utilisation à des fins de prospection ou de promotion commerciale de fichiers composés à partir de données issues directement ou indirectement des prescriptions médicales ou des informations mentionnées à l'article L 162-29 du code de la sécurité sociale, dès lors que ces fichiers permettent d'identifier directement ou indirectement le professionnel prescripteur ».

⁸¹ Dans le cas des établissements de santé, les indicateurs construits à partir d'initiatives privées sont restés frustes ; les pouvoirs publics ont entrepris pour leur part mais avec un retard certain, d'établir des indicateurs de résultats (taux de mortalité...) ajustés sur le risque. Cf. PL Bras, P. Gohet, A. Lopez L'information des usagers sur la qualité dans les établissements de santé. Rapport IGAS 2010

médecins et qu'ils ne sont pas ajustés en fonction des caractéristiques de la patientèle (par exemple, le taux de recours à la mammographie, très sensible au statut socio-économique des patientes, ne peut être valablement comparé entre médecins sans ajustement préalable).

Au total, si on ne peut exclure, à terme, la publication de certaines données nominatives sur la pratique, ce dévoilement ne peut s'envisager que dans un cadre concerté avec les professionnels concernés.

On retiendra de cette partie que le SI a des finalités très larges et qu'il constitue donc un bien public qui ne peut être approprié par aucun des acteurs du système. Il doit être administré dans l'intérêt commun par une autorité légitime en concertation avec l'ensemble des parties prenantes. Dès lors que les données présentent un risque de réidentification des patients, l'accès doit être restreint, mais les lots de données qui ne présentent pas un tel risque de réidentification doivent être communicables à tous ou rendus publics. Le débat relatif à la transparence sur des données nominatives des professionnels mérite d'être ouvert dans un cadre concerté avec les intéressés.

3 Les actions à conduire

A partir des principes qui viennent d'être dégagés, des propositions sont formulées sous forme d'actions à conduire. Les objectifs qui sous tendent ces propositions d'action sont :

- *Permettre un accès public à des lots de données issus du SNIIRAM qui ne présentent pas de risques d'atteinte à la vie privée (31) ;*
- *Organiser des accès contrôlés pour exploiter pleinement le potentiel du SI (32) ;*
- *Rendre la base plus accessible et plus riche (33)*

3.1 Permettre un accès public à des lots de données qui ne présentent pas de risques d'atteinte à la vie privée

3.1.1 Distinguer les données indirectement nominatives des lots de données véritablement anonymes.

Le SI est constitué à partir de données personnelles de santé considérées par la LIL comme devant être particulièrement protégées. Accéder à l'ensemble des données permet la réidentification. Cette possibilité de réidentification conduit à réglementer les accès au SI.

Il est toutefois possible de constituer des lots de données qui ne permettent pas ces réidentifications (informations floutées, chaînage temporel limité, chaînage entre domaines limités) ou qui raisonnablement rendent ce risque très improbable. Ainsi, l'accès à certains espaces du SI (EGB, Datamarts de données agrégées⁸²) a été ouvert progressivement.

Comme on l'a vu, cette ouverture a concerné dans un premier temps les ex Unions régionales de médecins libéraux à travers des transmissions de données de la CNAMTS à l'isPL. Les structures administratives (administrations, agences), les organismes de recherche (INSERM, IRDES, CNRS) ont également obtenu des droits d'accès. Puis, sous l'égide de l'IDS, d'autres institutions relevant de la société civile ont bénéficié de droits d'accès (CISS, fédérations d'établissements de santé, fédérations d'organismes complémentaires, UNPS). Les membres de l'IDS ont accès à l'EGB et aux Datamarts. Les adhérents des membres de l'UNOCAM (les fédérations d'organismes complémentaires) ont accès à l'EGB et aux Datamarts s'ils adhèrent à un dispositif d'échange de données. Les adhérents des membres de l'IDS ont accès aux Datamarts ainsi que les organismes complémentaires, membres des fédérations constitutives de l'UNOCAM, qui fournissent leurs propres données dans le cadre du projet MONACO.

Par ailleurs, la base PMSI fait elle-aussi, comme on l'a vu dans la première partie, l'objet d'une diffusion large (environ 290 bases fournies par l'ATIH) après autorisation préalable de la CNIL. Il semble que les demandes ne soient jamais refusées même si la CNIL exige dans certains cas le floutage de certaines données (âge en classes quinquennales, code résidence limité au département, remplacement du mode de sortie décès par un mode de sortie domicile). Il ne nous a pas été possible

⁸² Dans la suite, la mention des Datamarts du SNIIRAM sans autre précision renvoie aux données agrégées définies en première partie au § 1.3, à l'exclusion donc des données individuelles du Datamart des consommations inter-régimes (DCIR) et du PMSI notamment.

de reconstituer la logique qui préside aux décisions de la CNIL et il n'y pas, à notre connaissance, de « doctrine » accessibles sur les exigences de floutage. Il serait certainement plus efficient et plus pertinent, après expertise des possibilités de réidentification, de définir des lots de données floutées réputés anonymes et de ce fait accessibles sans contrainte ; seules les demandes d'accès à des données plus riches feraient l'objet d'une autorisation de la CNIL.

Cette extension progressive de l'accès à l'EGB et aux Datamarts de données agrégées, et la communication large des données du PMSI s'avèrent, à l'expérience, porteuses d'une contradiction. C'est trop ou pas assez, car de deux choses l'une :

- soit ces données sont indirectement nominatives et, dans ce cas, l'accès peut paraître trop large ;
- soit l'accès aux données ne présente pas de risque de réidentification et dans ce cas l'accès aux données est trop restreint, il mériterait d'être autorisé pour tous.

On peut d'autant plus s'interroger sur la cohérence de la politique d'ouverture des données qu'il est difficile de reconstituer la logique qui préside à la sélection des organismes ayant vocation à accéder à certains espaces du système.

Pourquoi un organisme d'assurance maladie complémentaire pourrait-il disposer de données inaccessibles pour une association de consommateurs ou pour un centre de recherche universitaire qui ne relève pas du CNRS et de l'INSERM ?

Par ailleurs, des accès sont octroyés au nom de la réciprocité à certains organismes parce qu'ils apportent eux-mêmes des données (cas des fédérations membres de l'UNOCAM qui ont accès à l'EGB et aux Datamarts et des organismes complémentaires qui ont accès aux Datamarts dès lors qu'ils contribuent au projet MONACO). Or, rien au plan des principes, ne justifie que les autorisations d'accès soient l'objet d'un « échange » : on ne « marchande » pas l'accès à des données publiques si cet accès présente un risque pour la vie privée ; si les données sont anonymes rien ne justifie que l'on exige d'autres données en échange. Le fait de détenir soi-même des données ne peut conférer un privilège d'accès à des données publiques.

Le principe d'une liste limitative, mais en extension, pour l'accès à certains espaces du SNIIRAM dès lors qu'aucun critère⁸³ ferme n'en définit les contours, ne peut que nourrir chez ceux qui en sont exclus le sentiment que des données publiques sont confinées sans raisons impérieuses de sécurité.

Après le mouvement d'ouverture progressive des accès à certains espaces du système d'information, il faut donc se donner les moyens de lever les ambiguïtés issues de la situation actuelle. On doit se fonder sur une appréciation documentée du risque de réidentification pour réguler les accès au SI. Il convient de distinguer des lots de données qui ne présentent pas de risques de réidentification et qui ont vocation à être communiqués ou diffusés à tous (par exemple, des données individuelles rendues imprécises, non chaînées ville/hôpital, avec une profondeur temporelle limitée mais aussi des données agrégées sur un groupe d'individus en nombre suffisant). Les accès à des niveaux de données plus riches doivent faire l'objet de dispositifs spécifiques d'autorisation.

S'il semble déjà acquis que l'accès à certains Datamarts n'entraîne pas de risque de réidentification, une expertise spécifique est nécessaire pour mesurer les risques de réidentification que présentent les différents types de données.

⁸³ Ce critère aurait pu être la nature publique ou privée, le caractère lucratif ou non de l'organisme. Aucun de ces critères ne permet de rendre compte de la liste des institutions autorisées.

L'expertise doit être globale. Il n'est pas possible, dans certains cas, de se prononcer sur les risques de réidentification à partir d'un lot de données pris isolément. Le rapprochement de lots de données peut rendre indirectement nominatif un lot qui, pris individuellement, est parfaitement anonyme. Il faut donc pour apprécier le risque prendre en compte dans leur globalité les données qui seraient accessibles sans contraintes.

Le mécanisme actuel des autorisations d'accès, le plus souvent au coup par coup, rend d'ailleurs difficile cette vue globale et peut aboutir à une prudence excessive : savoir que l'on court un risque potentiel sans être en mesure de l'objectiver peut entraîner des réactions de précaution excessives.

Cette expertise devrait être conduite sous l'égide du Ministère de la Santé en lien avec la CNIL et le ministère de la Recherche, et en faisant appel aux compétences notamment de l'INSEE, de l'ANSSI, de la CNAMTS et d'universitaires. Cet audit des risques devrait avoir pour objectif immédiat de spécifier les lots de données qui pourraient, sans risque, être communiqués ou publiés. Une attention particulière devrait dans le cadre de cette étude être portée à la granularité de présentation des données agrégées. En effet, le plus souvent les données agrégées sont présentées au niveau départemental alors même que des niveaux plus fins peuvent être utiles. La possibilité de réidentification augmente avec la finesse du découpage territorial mais il est certainement possible de proposer, pour bien des données agrégées, un niveau infra départemental.

En l'absence d'expertise systématique et publique sur les risques de réidentification, on ne peut écarter ni le soupçon que l'ouverture des données soit trop restrictive (soupçon qui s'exprime fortement dans le contexte actuel), ni la crainte d'un accès trop large.

Une telle étude fournira autant que possible les bases « objectives et scientifiques » d'un débat sur le périmètre des données du SI qui pourraient être accessibles à tous. L'objectif serait, à la suite de ce débat, de définir des types de données dont la mise à disposition tout public aurait vocation à être organisée. La définition de ces types de données s'effectuerait bien évidemment sous le contrôle de la CNIL.

Il n'est pas possible de préjuger des résultats de cette démarche. Toutefois, on peut espérer qu'elle aboutira à mettre à la disposition du public des lots de données permettant de satisfaire largement les besoins d'information ou d'études en matière de santé. Elle pourrait également satisfaire une part des besoins des chercheurs sachant que pour des recherches spécifiques, l'accès à des données plus riches restera nécessaire.

A l'inverse, l'expertise identifiera les types de données avec possibilité avérée de réidentification (et tranchera notamment le cas du DCIR dans sa forme actuelle). Pour le reste (échantillons, cohortes, données partielles ou floutées,...) il peut être difficile de trancher mais il s'agira de dégager des règles et de proposer des procédures pour faire basculer clairement les lots de données du côté de l'open data ou du côté de l'accès restreint.

312 Permettre l'accès effectif du public aux données anonymes

La définition de types de données qui pourraient être rendues publiques est un préalable, encore faut-il que les données soient aisément accessibles : l'accès au SI restera une opération complexe. Ouvrir l'accès n'est donc pas suffisant, il faut mettre en œuvre une politique de mise en ligne/publication.

Cette action de mise en ligne/publication est déjà assurée par la CNAMTS qui publie sur son site AMELI des données relatives aux soins de ville et par l'ATIH qui met elle aussi en ligne des données issues du PMSI. Elle est aussi assurée, pour le compte de ses membres, par l'IDS qui publie des indicateurs sous forme d'une trentaine de « tableaux de bord ». Ces tableaux de bord sont destinés à environ 180 organismes (membres de l'IDS, adhérents des membres de l'IDS, membres associés...). L'ISPL également diffuse des tableaux de bord en ligne tirés du SI, à l'usage des médecins.

Par ailleurs, l'IRDES assure de longue date la publication/diffusion de données de santé à travers la base Eco-Santé accessible gratuitement à tous, base alimentée à partir du SI mais aussi d'autres sources. Il faut aussi citer la FNORS qui maintient une base Score-Santé plus orientée que les initiatives précédentes vers la santé publique mais qui met aussi en ligne des données issues du SI.

Il faut également citer l'action de la DREES qui sur son site met aussi en ligne des données de santé, notamment la base SAE, complémentaire pour ce qui concerne les établissements de santé, des données issues du PMSI.

Le pluralisme des équipes qui assurent la fonction de mise en ligne est en soi une richesse (diversité des formats, des axes de présentation des données...). Dans la mesure où il s'agit d'initiatives financées sur fonds publics, il convient d'éviter les redondances et d'assurer la cohérence. La publication/mise en ligne des données sur le SI mérite d'être pilotée en veillant à la convergence des efforts des diverses équipes impliquées.

313 Fournir des données et des tableaux de bord à la demande

Même si des données sont ainsi mises en ligne, l'information diffusée restera générale et ne pourra pas satisfaire les besoins d'informations les plus spécifiques.

Par exemple, si dans le cadre de la mise à disposition publique, des données générales sur l'évolution de dépenses de biologie ou de radiologie sont mises en ligne, il est probable que cette information sera considérée comme insuffisante par des institutions (organisations syndicales, entreprises, organismes d'étude...) spécifiquement intéressées par ce domaine d'activité qui voudront des données plus détaillées (par type d'acte, par localisation géographique affinée).

Si ce besoin d'information peut être satisfait à partir du SI sans risque de réidentification, les données doivent être fournies ou des tableaux de bord constitués à la demande contre une rémunération destinée à couvrir les coûts de confection/mise à disposition. L'IDS, avec l'appui de la CNAMTS, confectionne aujourd'hui pour ses membres des tableaux de bord à la demande qui visent à répondre à des besoins spécifiques ; il s'agirait de se mettre en position de fournir cette prestation ou des extractions à tout demandeur.

Dès lors que l'information diffusée ne fait pas courir de risque d'atteinte à la vie privée, il n'y a pas de raison de ne pas la fournir mais si la production de l'information répond à un besoin spécifique, il n'y a pas de raison que le coût de production de cette information soit supporté par la collectivité.

Cette proposition fait apparaître une distinction entre un niveau d'information accessible gratuitement et un niveau d'information payant. La frontière entre les deux domaines est bien évidemment difficile à tracer. Elle devra se construire de manière pragmatique en distinguant les exigences d'une information générale sur le fonctionnement du système de santé (l'information de service public) et les besoins plus spécifiques.

3.2 Organiser des accès contrôlés pour exploiter pleinement le potentiel actuel et futur, en termes sanitaire et économique

Les actions proposées dans le chapitre précédent visent à définir des types de données pour lesquelles un accès public pourrait être prévu, mais ces données ne seront pas suffisantes pour permettre certaines recherches ou certaines études dont on peut attendre un bénéfice important en matière sanitaire ou économique.

Ce chapitre est consacré à des propositions visant à organiser les accès à des données plus riches pour exploiter le potentiel du SI tout en limitant strictement les risques d'atteinte au respect de la vie privée. Il se structure autour de quatre propositions :

- Se donner les moyens d'exploiter la richesse de la base de données au niveau des administrations publiques (321)
- Simplifier les procédures d'accès aux données à des fins de recherche, études et travaux statistiques (322)
- Faciliter les appariements (323)
- Mettre en œuvre des outils pour sécuriser et faciliter les accès (324)
- Prévenir les mésusages (325)

321 Se donner les moyens d'exploiter la richesse de la base de données au niveau des administrations publiques

Le potentiel de la base est actuellement sous exploité par les administrations publiques. Il faut donc confier au ministre chargé de la santé le soin de décider des accès permanent des institutions placées sous son autorité.

La loi devrait prévoir que le ministre chargé de la santé peut décider, après avis de la CNIL, des droits d'accès permanents aux données protégées, pour les services administratifs ou pour les organismes placés sous sa tutelle⁸⁴. Il revient en effet au ministre au regard des missions conférées à chacune des institutions d'apprécier si ces missions et les bénéfices attendus de l'accès au système justifient le risque qu'implique toute ouverture des accès. Rien ne justifie, comme c'est le cas actuellement que l'accès de certaines structures (agences, administrations centrales) soit subordonné à un avis du COPIIR SNIIRAM⁸⁵ et que pour d'autres (les ARS), les droits d'accès soient fixés par la loi.

Bien évidemment, limiter les risques exige de limiter les accès aux seules institutions pour lesquelles l'accès permanent est essentiel pour l'exercice d'une mission fondamentale en termes de sécurité et de qualité des soins.

Si ces institutions devraient pouvoir accéder aux données sans contrôle a priori, elles ne devraient exercer ce droit que dans des conditions de sécurité strictes : gestion des habilitations sous l'autorité du directeur de l'organisme, limitation des habilitations à un nombre restreint d'agents, contrôle

⁸⁴ Certains établissements publics à caractère scientifique et technologique bénéficient aujourd'hui d'un accès permanent, mais il est limité aux datamarts de données agrégées et –surtout– à l'EGB ; aucun n'a obtenu d'accès permanent aux données individuelles exhaustives du DCIR auxquelles leurs équipes ne peuvent accéder que sur demande et après accord de la CNIL.

⁸⁵ La procédure actuelle qui fait dépendre ces accès d'un avis du COPIIR SNIIRAM où les voix sont partagées à égalité entre les représentants de l'État, les représentants des caisses et les représentants des professionnels de santé ne nous paraît pas pertinente. Apprécier l'équilibre entre risque et bénéfices en ce domaine relève de la responsabilité du ministre.

interne sur les accès, traçabilité des requêtes. Les institutions concernées doivent rendre des comptes a posteriori sur l'utilisation de la base : publicité des requêtes effectuées, présentation des résultats obtenus. Un examen spécifique de l'activité en ce domaine (gestion des habilitations, bilan des requêtes, résultats...) devrait être effectué dans une instance appropriée qui peut varier selon les institutions (conseil d'administration, conseil de surveillance, comité d'éthique...).

A cet égard, le premier impératif est de permettre aux institutions publiques de tirer le meilleur parti du SI afin d'assurer la sécurité sanitaire de nos concitoyens. Après le cas *Médiateur*, il ne serait plus admissible que des signaux d'alerte soient négligés alors qu'ils pourraient être expertisés avec le SI.

Or, les interlocuteurs rencontrés au cours de la mission, chargés de la sécurité sanitaire, soulignent qu'ils ne peuvent traiter toutes les questions auxquelles ils sont confrontés. La réalisation sur le SI des études à visée de vigilance et de sécurité sanitaire repose actuellement principalement sur la CNAMTS dont ce n'est pas la mission : l'ANSM, la HAS et dans une moindre mesure l'InVS ne disposent pas des autorisations d'accès et des moyens adéquats.

Pour améliorer l'utilisation du système d'information dans cette perspective, il convient d'ouvrir à ces institutions des accès au cœur du système d'information sans qu'elles aient à justifier d'une autorisation a priori pour chaque étude. Cela implique un accès permanent à l'ensemble des données sur le mode de celui dont bénéficie la CNAMTS, au moins pour l'ANSM, la HAS et l'InVS.

Concrètement, la procédure d'élargissement de l'accès permanent est en cours pour l'InVS (qui disposait déjà d'un accès permanent sans possibilité de croiser les données sensibles) mais aussi pour l'ANSM et la HAS⁸⁶.

Le fait d'autoriser certaines institutions à accéder au SI ne suffit pas, encore faut-il qu'elles disposent des moyens pour en exploiter les potentialités. L'accès à la base est complexe, il n'est effectif que si l'institution est en capacité de constituer une équipe ayant une taille critique (au moins cinq personnes). Par ailleurs, les compétences propres à l'exploitation du système d'information doivent être articulées avec des compétences pharmaco-épidémiologiques et compétences spécifiques à la problématique étudiée (ex : une étude sur les effets d'un médicament cardio-vasculaire suppose de réunir un spécialiste des maladies cardio-vasculaires capable de construire la problématique de l'étude, un pharmaco-épidémiologiste et un bon connaisseur de la base qui sait formuler les requêtes pertinentes au regard de la problématique).

Dans un cadre contraint en termes d'emploi pour le ministère de la santé et ses opérateurs, les enjeux en termes de sécurité sanitaire sont tels que la constitution de ces équipes devrait constituer une priorité. Pour faire face à ce besoin de mener des études de vigilance et de pharmaco-épidémiologie, la loi du 29 décembre 2011 avait prévu la création d'un GIP entre l'Etat, la CNAMTS, l'ANSM, la HAS et l'InVS⁸⁷.

⁸⁶ Cf. arrêté SNIIRAM du 19 juillet 2013. Le COPIIR SNIIRAM s'était prononcé le 12 juin 2012 sur la demande d'accès de l'ANSM et de la HAS et sur l'extension aux données sensibles des possibilités d'accès de l'InVS. En revanche la CNIL ne s'est pas encore prononcée –à la date de remise du présent rapport- sur la demande d'accès de l'ATIH et de la DREES, acceptée par le COPIIR le 18 avril 2013.

⁸⁷ Art. L 5121-28 du CSP issu de la loi du 29/12/2011, figurant en annexe 10.

Avantages :

- La création de ce GIP permettrait par mutualisation de moyens de constituer rapidement une équipe ayant la taille et la gamme de compétences suffisantes, que chacun des organismes concernés pourrait avoir du mal à réunir ;
- Le texte donne déjà à ce GIP l'accès aux données du SI et la possibilité d'ouvrir cet accès à d'autres équipes (laboratoires universitaires) susceptibles d'être appelées en renfort en cas de besoin urgent ;
- Le fait que la loi existe permettrait d'aller vite.

Inconvénients :

- Le GIP fait courir le risque de conflits d'objectifs entre les diverses parties prenantes et de comportement de « cavalier seul » de chaque institution pour la mise à disposition de personnels ;
- La création d'un GIP externe aux institutions créerait une distance entre l'équipe dédiée à l'étude de la base et les spécialistes de chacune des problématiques au sein des institutions ;
- En l'absence de GIP, les moyens humains resteront dispersés mais la responsabilité de chaque directeur d'institution sera plus clairement affirmée : il lui appartient de se donner les moyens nécessaires à l'exercice de ses missions. Dès lors que des équipes seraient effectivement constituées, on peut espérer que la richesse du SI serait mieux exploitée : intégration de l'utilisation de la base dans le fonctionnement courant de l'organisme, dialogue de proximité entre les personnes compétentes en matière de SI et les spécialistes internes des problématiques traitées.

De fait, plus d'un an et demi après la loi créant le GIP, le décret d'application n'a pas été pris. En substitution, une convention de partenariat a été passée entre la CNAMTS et l'ANSM. Il importe de statuer sur l'avenir de ce GIP. Mais en tout état de cause il est urgent de dégager au sein de l'ANSM et de la HAS des moyens pour réaliser des études de vigilance et de pharmaco-épidémiologie.

Il faut aussi prévoir que les institutions concernées (ANSM, HAS, InVS) puissent s'adjoindre des équipes externes (équipes de laboratoires publics) qui bénéficieraient, pour la réalisation d'études urgentes de sécurité sanitaire, de leurs autorisations d'accès. Au-delà, on peut envisager que certains laboratoires publics qui, de fait, seraient des auxiliaires habituels de la politique de vigilances dès lors qu'ils apportent des garanties suffisantes en matière de préservation de la confidentialité des données soient dotés des mêmes droits d'accès que les institutions avec lesquelles ils collaborent.

Au-delà des droits d'accès justifiés par des considérations de sécurité sanitaire, il convient aussi, même si le problème est moins sensible, que les administrations publiques dont c'est une mission puissent exploiter la base pour produire des études et recherches sur la santé et le fonctionnement du système de soins. A cet égard, la CNAMTS est déjà habilitée et une procédure est en cours pour autoriser la DREES à accéder aux données individuelles de la base.

En ce domaine, il est plus délicat qu'en matière de sécurité sanitaire de circonscrire les institutions qui devraient avoir un accès profond à la base. L'admission de nouvelles institutions doit reposer sur une appréciation de la mission de l'organisme, de sa capacité à tirer parti de l'accès permanent et des bénéfices que la collectivité peut attendre des études et recherches conduites par l'institution.

L'exemple de l'ATIH permet d'illustrer cette problématique : il est indéniable que l'institution est en capacité d'exploiter la base mais ses missions fondamentales, centrées aujourd'hui sur l'hôpital et la tarification à l'activité, ne justifient pas elles seules un accès permanent. Pour autant dès lors qu'il est prévu dans sa convention d'objectif que l'ATIH contribue à des études sur les parcours de soins, il pourrait être pertinent de lui permettre un tel accès.

C'est la même démarche qui devrait être entreprise pour les ARS dans l'hypothèse où une nouvelle loi confierait à la ministre comme proposé supra de décider des accès des administrations placées sous sa tutelle.

Par ailleurs et comme on l'a vu dans la première partie, les ARS illustrent un cas où l'accès est différencié selon que l'opérateur est médecin ou non médecin ou selon qu'il est placé ou non sous l'autorité d'un médecin. Attacher des droits d'accès à la qualité de médecin ou au fait d'être placé sous l'autorité d'un médecin est une manière de réduire le risque en confinant les accès. Pour autant, ce mode de confinement est peu pertinent. Dès lors que l'on accepte qu'une institution accède à des données, restreindre l'accès aux seuls médecins peut rendre cette ouverture inopérante (les médecins peuvent ne pas être disponibles pour réaliser les études et requêtes, ils ne sont pas toujours les plus aptes à utiliser le système d'information). Le fait que le médecin soit astreint au secret médical dans la relation de soins n'apporte aucune garantie particulière lorsque le médecin accède à des informations hors de la relation de soins. Un agent non médecin est tenu comme ses collègues médecins au secret. Confiner l'accès aux médecins est un moyen terme qui constitue une fausse sécurité. Dès lors que l'on admet qu'une institution doit disposer de certaines données, il faut admettre qu'il appartient au responsable d'habiliter les personnes pour l'accès à ces données indépendamment du statut médecin/non médecin.

322 Simplifier les procédures d'accès aux données à des fins de recherche, études et travaux statistiques

Au-delà des accès des administrations publiques, il faut plus largement favoriser la réalisation de recherches et d'études à partir du SI et donc simplifier les procédures d'accès aux données protégées du SI.

L'autorisation d'accès à des données potentiellement réidentifiantes ne doit être possible que :

- si l'étude ou la recherche envisagée est susceptible d'apporter un bénéfice collectif en matière scientifique ou de santé publique. Il convient donc de prendre en compte la finalité du traitement, sa pertinence au regard de la finalité affichée et le besoin d'accéder effectivement aux données. Une étude qui ne viserait que des intérêts particuliers (par exemple : marketing santé...) ne saurait donner lieu à autorisation.
- si elle est réalisée par une institution qui apporte des garanties suffisantes quant à ses compétences et à la préservation de la confidentialité des données qui lui seront accessibles.

Aussi, l'accès aux données du SI dès lors qu'il s'agit de mobiliser des données potentiellement réidentifiantes doit se faire sur la base d'une autorisation préalable. Il importe de simplifier le dispositif et d'en raccourcir dans la mesure du possible les délais. (cf. chapitre I sur les difficultés d'accès aux données).

Il importe notamment de définir une seule voie d'accès pour en finir avec la complexité et la perplexité qu'entraîne aujourd'hui l'existence de plusieurs circuits d'autorisation selon que l'on

relève de l'article 25, du chapitre IX (traitement de données à caractère personnel ayant pour la recherche dans le domaine de la santé) ou du chapitre X (traitement des données à caractère personnel à des fins d'évaluation ou d'analyse des pratiques ou des activités de soins et de prévention) de la loi informatique et liberté (cf. chapitre 1).

Le processus d'autorisation doit faire intervenir la CNIL dans son rôle de protection des données personnelles. Actuellement la CNIL s'appuie dans le cadre du chapitre IX sur un avis préalable du Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé (CCTIRS) et dans le cadre du chapitre X sur un avis de l'IDS.

Dès lors que l'on souhaite unifier les circuits d'accès, il semble préférable de privilégier l'avis préalable d'une instance mobilisant des experts plutôt qu'une instance composée de représentants de la société civile. Juger si une étude est utile ou si des traitements informatiques/statistiques sont pertinents suppose une connaissance scientifique du domaine concerné par l'étude.

Il conviendrait donc de reconfigurer le CCTIRS pour qu'il puisse englober l'ensemble du champ potentiellement concerné par les études faisant appel aux données du SI : pharmacologie, épidémiologie, génétique mais aussi santé publique, économie de la santé, recherche en service de santé⁸⁸. Actuellement, le CCTIRS est composé de personnes compétentes « dans le domaine de la santé, d'épidémiologie, de génétique et de bio-statistique »⁸⁹.

Le CCTIRS selon la loi ne se prononce aujourd'hui que sur « la méthodologie de la recherche » et sur « la nécessité du recours à des données à caractère personnel et la pertinence de celles-ci »⁹⁰. A la différence de la procédure actuelle, le CCTIRS, pour éclairer la CNIL, devrait en outre se prononcer, explicitement sur la finalité de l'étude et sur sa pertinence au regard de ses finalités : est-elle susceptible d'apporter un plus en termes de connaissance ou de santé publique ?

Pour raccourcir les délais, on peut envisager que la procédure d'instruction aujourd'hui séquentielle (CCTIRS puis CNIL) soit simultanée. Un dossier commun serait déposé simultanément devant les deux institutions. Toute demande de complément d'information adressée par l'une des institutions serait portée à la connaissance de l'autre. La décision de la CNIL devrait alors suivre dans un délai rapproché l'avis du CCTIRS.

Bien évidemment, si comme on peut le souhaiter le nombre des demandes d'autorisation augmente, les moyens du CCTIRS devront être renforcés notamment pour ce qui concerne le vivier des rapporteurs. Dans cette perspective, la rémunération de cette fonction devra être revue.

Il est souhaitable que des recherches se développent à l'initiative des laboratoires de recherche pour des projets à visée académique mais elles émaneront aussi des industries de santé et notamment des laboratoires pharmaceutiques.

Lorsque l'étude est réalisée de sa propre initiative par un industriel de la santé, la CNAMTS, la HAS et l'ANSM devraient être systématiquement appelées à faire valoir leur position dès le dépôt du dossier

⁸⁸ Actuellement la répartition des compétences du CCTIRS et des CPP dans le cadre de la loi Jardé pose des problèmes non résolus mais qu'il n'est pas possible de traiter dans le cadre du présent rapport.

⁸⁹ Art 54 de la loi Informatique et Libertés. Si le terme « santé » est large, les scientifiques qui composent le comité sont effectivement compétents en épidémiologie, génétique et biostatistiques.

⁹⁰ *Idem*. Sur cette question voir C. Bonaïti-Pellié, P. Arveux, A. Billete de Villeneuve et al. Traitement de l'information en matière de recherche dans le domaine de la santé : Nul n'est censé ignorer la loi ? *Medecine/Science* n°1, vol.25, Janvier 2009 où s'exprime le regret que les scientifiques qui composent le CCTIRS ne puissent pas se prononcer sur l'intérêt de la recherche.

auprès de la CNIL et du CCTIRS. Ces agences de régulation sont les mieux à même d’alerter sur une étude dont on peut douter de la finalité affichée (étude à visée promotionnelle sans perspectives d’apport de connaissance) ou dont le protocole serait biaisé en fonction des intérêts du promoteur de l’étude.

En revanche, la procédure serait allégée lorsqu’une recherche est réalisée pour le compte d’un industriel de la santé mais à la demande d’une autorité publique (DGS, ANSM, HAS, CEPS). Dès lors que cette autorité publique a validé le protocole d’étude, on peut se satisfaire d’un simple avis de la CNIL sans avis préalable du CCTIRS.

Actuellement, les intervenants privés lorsqu’ils veulent réaliser une étude à partir des données du SI doivent passer par l’intermédiaire d’un laboratoire public de recherche. L’exigence de mobiliser un intermédiaire doit, nous semble-t-il être maintenue. Même s’il a été amplement démontré que les résultats des études et recherches sont influencés par leur mode de financement, la présence d’un intermédiaire apporte une garantie quant à la qualité des travaux. Pour autant, des organismes d’études privés, et pas seulement des laboratoires publics, sont aussi susceptibles de servir d’intermédiaire. Admettre des intervenants privés permettrait de mieux répondre aux demandes des autorités publiques adressées aux entreprises pharmaceutiques et d’éviter que certains laboratoires publics soient en position de quasi monopole. Ils pourraient donc être admis comme intermédiaires dès lors qu’ils peuvent justifier de leur compétences en matière de recherche, que leur ressources sont diversifiées (gage d’indépendance vis à vis du laboratoire commanditaire)⁹¹ et qu’ils présentent des garanties adéquates en matière de préservation de la confidentialité et de séparation entre les activités d’études et les autres fonctions (conseil, marketing...).

La procédure spécifique pour les services de la statistique publique qui s’appuie sur le CNIS serait par ailleurs maintenue.

323 Faciliter les appariements

Dans le chapitre 1, il a été fait état des obstacles qui rendent particulièrement difficiles les appariements entre données. Or, si des études sont possibles en s’appuyant sur les seules données du SI, son potentiel ne peut être pleinement exploité que si ses données peuvent être rapprochées d’autres sources (base de données déjà constituées, données collectées dans le cadre d’une enquête spécifique).

Il est donc proposé de modifier les articles 27 et 29 de la LIL et de ne plus exiger un décret en conseil d’Etat pour les traitements de données personnelles parmi lesquelles figure le NIR dès lors que le consentement des personnes a été recueilli. Les appariements seraient autorisés selon la procédure pour l’accès aux données personnelles de santé décrite au paragraphe précédent. Cette procédure nous paraît fournir des garanties suffisantes sans que soit imposé aux équipes de recherche de devoir solliciter que l’administration prépare un décret en Conseil d’Etat.

En l’attente de cette modification de la loi, il pourrait être envisagé de prendre un décret cadre. Cette possibilité initialement suggérée par la CNIL, a été explorée au ministère de la santé. Mais les services du ministère et la CNIL ont formulé des doutes juridiques sur la possibilité d’aboutir, concluant qu’une modification législative est nécessaire

⁹¹ La même exigence –débouchant sur un label- devrait d’ailleurs valoir pour les laboratoires publics.

Lorsque l'identifiant au sein du SI est construit par chiffrement irréversible du NIR, l'appariement avec des données extérieures au SI s'effectue en faisant subir au NIR associé aux données extérieures les mêmes opérations de chiffrement. Si le chiffrement irréversible a pu être considéré comme une sécurité, confier à un tiers de confiance une table de correspondance entre NIR et l'identifiant SNIIRAM permettrait de simplifier la procédure avec un degré similaire de sécurité⁹².

Les données ainsi appariées sont très sensibles et méritent d'être hébergées dans un environnement particulièrement sécurisé. Un tel environnement sécurisé pourrait être fourni aux équipes de recherche qui ne sont pas en mesure de garantir elles-mêmes la sécurité physique des bases ainsi constituées (cf. infra 324).

Les appariements de bases de données administratives entre elles –qui peuvent exiger également un passage par le NIR- posent un problème différent, de choix des priorités notamment. Ce point est évoqué ci-après au § 333.

324 Mettre en œuvre des outils pour faciliter et sécuriser les accès

Dans d'autres pays (Danemark, Canada), les plates-formes d'accès à des données confidentielles ont développé des techniques pour concilier les finalités contradictoires de la mise à disposition et de la protection.

Pour rendre accessibles des données personnelles tout en réduisant le risque de dissémination, on commence bien entendu, comme on l'a vu aux § 321 et 322, par trier sur le volet les organismes et les personnes habilités à accéder aux données, en fonction de leurs missions (pour les institutions qui disposent d'un accès permanent) ou de la nature du projet et de la qualité du demandeur (pour les organismes de recherche et d'études).

Mais on va généralement plus loin :

- d'abord, il peut être demandé aux personnes habilitées de signer une charte et certifier qu'elles ont été informées de leur obligation de secret et des sanctions dont elles sont passibles en cas d'indiscrétion ;
- ensuite, il y a des cas où la consultation et le traitement des données ne peuvent être réalisés que dans un lieu et sur des machines mis à disposition par la plateforme dont c'est la mission ; des dispositions sont prises pour que les utilisateurs de ces guichets ne puissent pas copier les données ; ils ne peuvent faire sortir que leurs résultats, sous le contrôle des agents de la plateforme ;
- enfin, des statisticiens attachés à la plateforme peuvent être chargés de réaliser eux-mêmes les études simples qui leur seraient commandées par les « clients » ce qui évidemment est supposé supprimer le besoin des demandeurs d'accéder eux-mêmes aux données identifiantes.

Il est évident que ces deux dernières procédures (le guichet physique unique où il faut se déplacer et la réalisation des études pour le compte de tiers) sont malcommodes, voire impraticables ou très coûteuses. Toutefois, il existe aujourd'hui des équivalents informatiques de ces procédures qui en ont les qualités sans en avoir les inconvénients.

⁹² Cela ne préjuge pas des modalités techniques de conservation (ou de reconstitution « à la volée ») de la table de correspondance (chiffrement...) ; l'important est que cette table puisse être utilisée pour des appariements autorisés.

En premier lieu, une base bien organisée et bien documentée peut offrir à ses utilisateurs qu'ils spécifient précisément leurs requêtes à l'aide d'outils en ligne fournis par la plateforme : ils obtiendront ainsi les résultats sans avoir « vu » eux-mêmes les données. Cette technique, dite de *remote execution*, permet de faire bouger la frontière entre études exigeant l'accès à des données personnelles et études ne permettant pas de réidentifier les patients : le besoin d'accéder aux données personnelles est réduit. Cependant, ce procédé – qui revient à faire réaliser l'étude par la plateforme- n'est applicable qu'à des études assez simples et sauf exception ne pourra pas servir aux chercheurs.

En second lieu, des dispositifs d'accès sécurisés à distance ont été conçus pour éviter que les données puissent être copiées et disséminées. Il se trouve que l'INSEE -ou plus précisément le Groupe des écoles nationales d'économie et de statistiques (GENES), ancienne direction de l'INSEE devenue un établissement public- a développé et mis en service un tel dispositif qui est aujourd'hui utilisé par plusieurs centaines de chercheurs en France. Le principe du Centre d'accès sécurisé distant (CASD) est de permettre à des personnes habilitées et dûment authentifiées, d'accéder depuis leur lieu de travail, à des données protégées et à des outils de traitement mis en ligne. Cela ne supprime pas le risque de dissémination mais le réduit très sensiblement. Il serait sans doute opportun que ce modèle soit utilisé pour l'accès au SI.

On pourrait s'inspirer ainsi de l'article 104 -sur l'accès aux données fiscales- de la loi sur l'enseignement supérieur et la recherche du 22 juillet 2013, qui prévoit l'accès de chercheurs habilités aux données fiscales personnelles et impose que « l'accès aux informations s'effectue dans des conditions préservant la confidentialité des données ».

325 Prévenir les mésusages.

Dans la partie relative aux principes sur l'accès aux données (paragraphe 23), nous avons souligné d'une part que le risque de mésusage ne pouvait à lui seul constituer une justification pour faire obstacle à un accès à des données qui ne présenterait pas de risque de réidentification ; d'autre part que l'accès à des données présentant un risque de réidentification devait faire l'objet d'un contrôle a priori.

Pour autant on ne peut négliger la crainte que les nouvelles possibilités d'étude ouvertes aux laboratoires soient exploitées dans une visée stratégique pour flatter les résultats de leurs produits. L'asymétrie entre les moyens que les laboratoires peuvent consacrer à de telles études et ceux que peuvent dégager les autorités publiques renforce cette inquiétude qui a été exprimée avec force par certains interlocuteurs de la mission.

A ce stade des propositions, il est prévu que des accès larges soient ouverts à des lots de données non réidentifiantes (paragraphe 31) et que des accès contrôlés soient réorganisés pour l'accès à des données avec risque de réidentification (paragraphe 32).

Il est peu probable que les données en accès public permettront de conduire des recherches approfondies telles que celles qui sont nécessaires pour faire émerger les propriétés d'un médicament. Les accès des industries de santé au SI resteront donc contrôlés a priori pour l'essentiel.

Ce contrôle a priori, dans lequel il est proposé une intervention des autorités de régulation (ANSM, HAS, CNAMTS) (cf. 322) permettra de limiter les risques de mésusage.

Bien évidemment, dès lors qu'un accès au SI aura été autorisé pour une recherche émanant d'un laboratoire, un compte rendu des résultats devrait, à défaut de publication, être exigé à une date préfixée et rendu public pour réduire le biais de publication.

Par ailleurs, le SI sera utilisé pour une grande part pour des études et recherches demandées par les autorités sanitaires. Développer l'utilisation du SI à des fins d'étude de vigilance peut permettre aux autorités sanitaires d'être plus exigeantes vis-à-vis des laboratoires : les données étant accessibles, certains résultats pourront être produits plus aisément et plus rapidement.

De plus, il est impératif que les autorités sanitaires disposent en interne de moyens d'études propres (cf. 321) et de contre expertise en proportion du développement éventuel des études réalisées par les laboratoires. On pourrait envisager à cet égard que les tarifs demandés aux laboratoires lors de l'accès aux données comprennent une quote-part destinée à financer, à l'initiative des autorités sanitaires, des revues critiques des études financées à l'initiative des laboratoires.

Enfin il ne faut pas négliger le fait qu'un SI accessible peut aussi servir de support à des études académiques et favoriser ainsi le développement d'une communauté scientifique capable de nourrir des controverses sur les méthodologies et les résultats. Un espace scientifique pluraliste reste la meilleure garantie contre les tentations de manipulation des résultats et des conclusions.

Au total, le risque ne doit pas être sous-estimé mais la manière de le combattre est la critique des études et non l'interdiction d'accéder aux données et la stérilisation des possibilités de recherche.

D'ailleurs, le simple fait qu'il soit possible d'accéder à une base riche constitue aussi un défi pour certaines stratégies de laboratoires (il sera plus difficile à un laboratoire de se réfugier derrière les difficultés et les délais pour mettre en place une étude observationnelle en vie réelle ; de même la disponibilité de la base peut relativiser la portée d'études construites par le laboratoire en fonction de ses intérêts)⁹³. Les avertissements des lanceurs d'alerte ne pourront plus être négligés dès lors qu'ils pourront être confirmés ou invalidés par le recours au SI.

Il faut souligner aussi que limiter a priori l'accès des laboratoires pharmaceutiques à la base contribuerait à les exonérer à bon compte de leur devoir de vigilance à l'égard de leurs produits. Si le SI est accessible, un laboratoire qui, alerté sur d'éventuels indésirables, n'utiliserait pas la base pour vérifier l'ampleur du problème pourrait être tenu pour responsable de cette omission.

Enfin, restreindre l'accès à la base française n'empêchera pas la production d'études à partir de données étrangères, études dont il est encore plus difficile pour les autorités françaises d'apprécier la pertinence. Les controverses sur l'efficacité des médicaments ont une dimension internationale ; la fonction critique à l'égard des études réalisées par les laboratoires ne relèvera pas des seuls pouvoirs publics français mais mobilisera de fait l'ensemble des autorités publiques, des chercheurs spécialisés, des institutions de vigilance (cf. *Prescrire* en France) au niveau mondial.

⁹³ Cf. l'étude récente de la CNAMTS qui établit l'absence de différence significative sur la morbidimortalité entre la rosuvastatine et la simvastatine, le fait même qu'une telle étude soit possible dévalorise les stratégies de comparaison qui utilisent un indicateur clinique intermédiaire pour comparer deux produits.

3.3 Rendre le SI plus accessible.

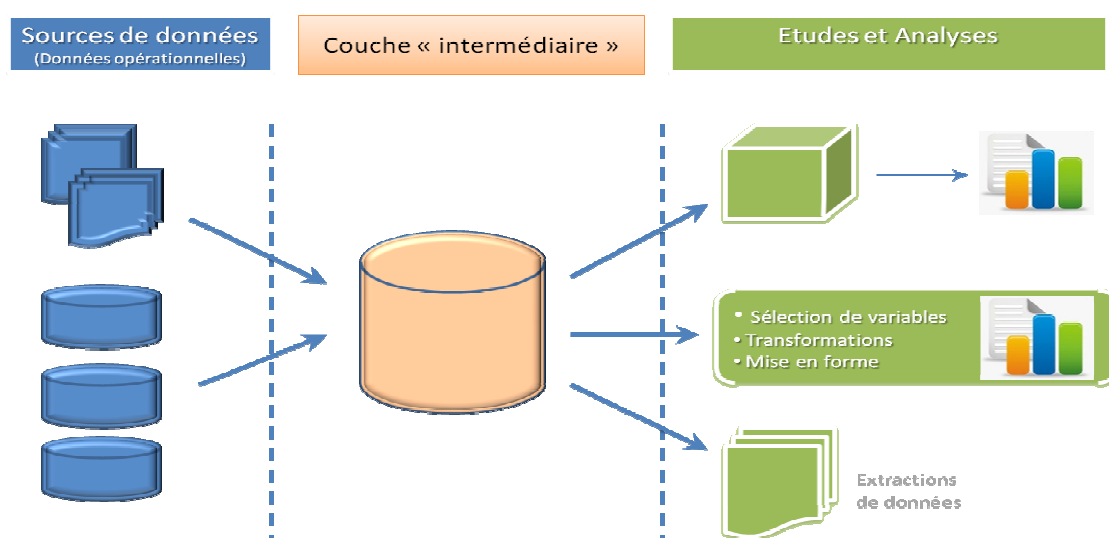
Les difficultés d'accès au SI ne sont pas simplement d'ordre juridique mais tiennent à la complexité du système d'information qui, constitué à partir des informations de gestion, n'est pas structuré à des fins de recherche et d'étude (cf. chapitre 1). Il convient donc de faciliter la compréhension des données et leur(s) traitement(s). Pour cela, il est nécessaire de :

- Urbaniser le SI et documenter ses composantes (331);
- Constituer une structure de service autour du SI (332) ;
- Prévoir l'enrichissement du SI (333).

331 Urbaniser le SI et documenter ses composantes

L'enjeu en ce domaine est de retraiter les données sources pour les présenter de la manière la plus proche du vocabulaire et des problématiques à traiter par les utilisateurs.

Pour le SI, cela signifie un travail de restructuration des données qui peut commencer dès l'entrepôt, et doit se poursuivre jusqu'aux Datamarts ou « cubes de données⁹⁴ » mis à disposition des utilisateurs.



L'information doit être préparée pour être exploitée de manière directe en suivant les principales étapes de constitution / simplification suivantes :

- Filtrage : Certaines données n'ont pas d'utilité à être présentées directement à l'utilisateur ; d'autres ne doivent pas l'être pour des raisons de périmètre de visibilité (données sensibles, profondeur d'historique limitée...) ;
- Traduction : Les données (ou sous-ensembles de données) doivent être « renommées » pour revêtir une sémantique plus proche du contexte d'analyse de l'utilisateur ;
- Réorganisation : La réorganisation des données doit permettre de constituer des indicateurs dérivés des données brutes, qui seront plus adaptés aux demandes des utilisateurs, en constituant des agrégations ou en permettant des analyses d'évolutions dans le temps directement.

⁹⁴ Structure des données multidimensionnelle offrant des possibilités de « navigation » dans les différentes dimensions des données.

Les travaux menés actuellement par la CNAMTS dans le cadre du projet « EGB Simplifié » sont un bon exemple de réorganisation et retraduction pour que les données soient moins connotées « données de remboursement ».

En prolongement de cette réorganisation des données, la mise en place d'une gestion des métadonnées permet de dispenser les utilisateurs de la complexité technique et de mettre à leur disposition un environnement dans lequel l'appréhension, l'accès et le partage de l'information sont plus aisés. Dans un tel contexte, les utilisateurs doivent être en mesure de saisir :

- Ce que chaque information qui leur est présentée recouvre,
- Comment chaque information / indicateur a été construit (règles métiers, sources de données...).

La constitution de dictionnaires de données est donc à systématiser en la complétant si possible d'informations sur la qualité de la donnée et des différentes transformations réalisées dans le processus de regroupement et de simplification.

332 Constituer une structure de service autour du SI.

La réorganisation de la base de la base pour la rendre plus accessible doit être prolongée par la création d'une structure de service tournée vers les utilisateurs. La création de cette « plate-forme » est nécessaire pour réduire les obstacles à l'accès aux données évoquées au chapitre 1.

Elle aurait vocation à assurer l'hébergement de la base ou à en faire assurer l'hébergement, dans des conditions qui en assurent la sécurité et la disponibilité⁹⁵. Elle aurait aussi vocation à organiser les contrôles d'accès (habilitations, authentification...).

Cette plate-forme de services assurerait les fonctions évoquées au chapitre 31 : faciliter les accès aux espaces ouverts du SI, réaliser et maintenir les tableaux de bord qui permettent d'accéder aux informations publiques extraites du SI et/ou alimenter les services qui mettent à disposition les données (Eco-santé...), fournir des tableaux de bord à la demande.

Elle aurait aussi vocation à servir d'appui aux équipes de recherche qui ont vocation à accéder à des données protégées : contribuer à l'expression du besoin sur les données à extraire, fournir les extractions après autorisation, répondre aux interrogations sur la nature des données, gérer les conventions avec les utilisateurs, tarifier les services, former les utilisateurs. En ce domaine, fournir les données autorisées dans un délai raisonnable constituerait un objectif central de la structure de service (cf. chapitre 1...).

Il s'agit de fonctions qui sont assurées actuellement par l'ATIH pour les données hospitalières et par la CNAMTS dès lors que des données ville-hôpital sont nécessaires. Mais elles constituent des missions annexes par rapport aux objectifs principaux de l'ATIH et la CNAMTS. L'IDS assume aussi une partie de ces fonctions pour ses membres.

Cette fonction de service est bien évidemment liée à celle d'urbanisation du SI et de documentation des données.

Dans cette fonction d'aide aux utilisateurs, la plate-forme pourra compter sur les initiatives des utilisateurs/chercheurs eux-mêmes soucieux de constituer des réseaux pour capitaliser des

⁹⁵ Elle pourrait également offrir un service d'hébergement des bases appariées (cf. 323)

connaissances, des expériences et des savoir-faire relatifs à l'utilisation du SI⁹⁶. La plate-forme de service pourrait fournir un support à ces initiatives.

Il n'est pas possible de prévoir aujourd'hui le dimensionnement de la plate-forme qui doit pouvoir s'adapter à la demande et qui doit donc être financé par des contributions des utilisateurs (cf. 41).

333 Prévoir l'enrichissement du SI

Le mouvement d'enrichissement du SI engagé depuis l'origine est appelé à se poursuivre.

Une expérimentation a permis de tester l'appariement avec le CépdiDC, la base gérée par l'INSERM qui recense les causes de décès. Cet appariement accroît la pertinence du SI pour les études de santé publique. L'industrialisation de cet appariement suppose notamment de modifier l'article L 2223-42 du code général des collectivités territoriales qui énumère les destinataires de ces informations.

Pour autant, même après ce développement, le SI sera encore amené à évoluer et de nombreuses questions sont posées sur son enrichissement.

Par exemple, le SI pourrait être enrichi si le codage des diagnostics en ville, prévu de longue date, permettait d'intégrer cette information. De même si le SI enregistre les actes de biologie, il n'intègre pas les résultats dont on pourrait envisager qu'ils soient codés et viennent également enrichir le SI. Si le SI intègre les médicaments de la liste en sus dispensés pendant les hospitalisations, il gagnerait à englober également les médicaments courants dispensés au cours de l'hospitalisation.

Pour des études plus fines sur les inégalités territoriales et sociales en matière de santé, l'adresse des personnes pourrait être géocodée, convertie en code IRIS et transférée dans le SI.

Un premier test dans le cadre du projet MONACO a permis d'apparier pour 2500 personnes les données de remboursement des régimes obligatoires avec ceux d'organismes complémentaires, il convient d'apprécier si cette expérimentation mérite, afin de mieux connaître les restes à charge après AMO et AMC, d'être étendue ou si au contraire le modèle de microsimulation INES-OMAR de la DREES fournit en ce domaine des informations suffisantes.

Dès lors qu'un système d'information sur les prises en charge médico-sociale se développerait, son intégration dans le SI serait également susceptible d'enrichir les possibilités d'étude.

Sans prétendre à l'exhaustivité, cette évocation des perspectives de développement du SI vise à montrer que de nouveaux investissements et donc des choix d'orientation sont nécessaires. Il conviendra d'arbitrer entre les projets ou du moins de les prioriser. Ces choix devront être construits en association avec les parties prenantes du système de santé (représentants de patients/usagers, des professionnels de santé, des établissements de santé, des régimes obligatoires, des organismes complémentaires, des industries de santé) et avec les représentants des institutions et organismes de recherche qui utilisent le SI (cf. § 4.3 Quelle gouvernance ?)

⁹⁶ Cf. le projet de réseau REDSIAM (Réseau Données SNIIRAM) en cours de constitution. Le besoin d'un pôle de compétences commun -que la DREES devrait animer- pour les utilisateurs du côté des services de l'Etat est bien identifié également.

4 Le dispositif économique et institutionnel

Le paragraphe 3 était consacré aux actions qu'il est proposé de conduire ; ce paragraphe vise à préciser le modèle économique du SI (41), fixer les termes du débat sur l'institution qui devrait les mettre en œuvre (42) et proposer une organisation pour la gouvernance (43).

4.1 Quel modèle économique ?

Les propositions d'action présentées dans la partie précédente supposent tout d'abord de constituer des équipes et mutualiser des compétences au sein des agences et administrations afin d'exploiter les richesses du système d'information notamment dans une perspective d'accroître les études en matière de sécurité sanitaire. Les contraintes sur les emplois dans le périmètre santé/social (ministère/opérateurs) sont fortes. Dans le cadre de ce rapport, on ne peut que se borner à souligner que la constitution de ces équipes devrait bénéficier d'une priorité.

Au-delà de cette action urgente, les propositions présentées aboutissent à prévoir un investissement pour urbaniser la base (proposition 331). Il n'a pas été possible à ce stade d'en estimer le montant. Il s'agit d'un investissement relativement conséquent dont on ne peut indiquer à ce stade qu'un ordre de grandeur d'environ 10 millions d'euros comprenant l'hébergement, la sécurité et les études initiales. Mais ce montant reste modeste si cet investissement est considéré comme un investissement d'avenir ; un système urbanisé constituerait une infrastructure de recherche.

A la suite de cet investissement initial, dans un fonctionnement en routine, la gestion du système et les fonctions de service associées (hébergement des données, documentation, fourniture de lots de données, services aux utilisateurs) entraîneront des coûts de fonctionnement dont certains sont d'ailleurs d'ores et déjà supportés par la CNAMTS, l'ATIH ou même l'IDS.

Parmi ces coûts de fonctionnement, certains renverront à des missions de service public (fourniture de données publiques, service aux administrations et agences publiques utilisatrices), d'autres pourront être reportés vers les utilisateurs à travers une tarification éventuellement différenciée selon la nature du commanditaire (entreprise privée ou laboratoire académique). La tarification des services permet de garantir que les moyens consacrés à la fourniture de ces services évoluent en fonction de la demande ; il ne serait pas opportun que les recherches à partir du système d'information soient limitées simplement parce que les moyens affectés aux services aux utilisateurs sont contingentés. Un modèle économique fondé sur une tarification des services doit permettre à l'offre de services de suivre les besoins d'études.

Même s'il faut tenir compte des différences de contexte qui ne permettent pas bien évidemment de transposer terme à terme le modèle anglais, on notera que pour une fonction similaire, le Royaume-Uni a créé une institution (le CPRD), adossée au service statistique santé/social, qui assume des fonctions similaires à celles envisagées ici et dont il envisage qu'elle soit équilibrée financièrement.

Il n'a pas été possible de chiffrer plus précisément les ressources et dépenses de la plate-forme. Les interlocuteurs de la mission considèrent que le SI pourrait être largement sollicité pour des études et recherches mais l'ampleur de la demande et de son évolution ne peut être précisément quantifiée. Le modèle devra être construit de manière pragmatique en fonction de l'expérience mais le principe d'une tarification permettant d'assurer une évolution des moyens consacrés à la base en fonction des demandes devrait être privilégié.

Par ailleurs, dans le cadre du CISS, le LEEM soucieux de voir se constituer une plate-forme permettant l'exploitation des données du SI a proposé de contribuer sous forme d'un fonds participatif à l'investissement initial d'urbanisation de la base. Cette hypothèse pourrait être explorée.

Cette proposition et le fait que la plate-forme de service du SI soit financée par des contributions dont certaines proviendront de l'industrie pharmaceutique peuvent faire craindre une perte d'indépendance du SI « soumis aux intérêts de l'industrie pharmaceutique » (le fait que l'AFSSAPS reçoive des redevances des laboratoires a pu engendrer le soupçon que l'agence aurait aliéné son indépendance). Pour autant, il serait aussi problématique que les entreprises pharmaceutiques bénéficient d'un service gratuit. Le fait que l'institution chargée de la plate-forme n'ait pas vocation à prendre des décisions à l'égard des entreprises pharmaceutiques devrait apaiser ces craintes.

Chaque année, des centaines de millions d'euros sont déplacés entre l'industrie pharmaceutique et les régimes d'assurance-maladie à travers les décisions prises en matière de taxe et surtout de prix des médicaments pour assurer le respect de l'ONDAM. Au regard de ces montants, les sommes nécessaires pour l'investissement et le fonctionnement du SI sont faibles. Il est donc possible de faire contribuer l'industrie pharmaceutique par ce biais sans affectation des fonds. Il nous semble plus pertinent toutefois d'envisager des contributions et redevances affectées (participation à l'investissement initial, paiement lors de l'accès aux données).

4.2 Quelle organisation pour gérer la base ?

Les actions évoquées au chapitre 3 (diffusion de l'information publique, urbanisation et documentation des données, constitution d'une plate-forme d'aide aux activités de recherche réalisées à partir de ces bases de données, enrichissement par d'autres sources, mise en place d'un environnement technique sécurisé) constituent un ensemble de chantiers ambitieux, au service d'une perspective volontariste d'amélioration de l'accessibilité et de développement des usages de ces données.

Ces chantiers s'inscrivent dans le prolongement d'un certain nombre d'actions mises en œuvre par la CNAMTS dans les années récentes. La base de données a évolué et s'est enrichie en permanence (création de l'EGB et des Datamarts, appariement avec le PMSI, mise à disposition d'un EGB simplifié, appariement avec les causes de décès...). La CNAMTS répond à des besoins externes (formations, support aux utilisateurs, collaboration avec l'IDS pour la confection de tableaux de bord, extraction de données pour des projets de recherche, contribution à la constitution de grandes cohortes et à leur appariement avec le SI...). Elle assure la diffusion de certaines des données sur AMELI, alimente Eco-santé et Score santé. Elle a accompagné et soutenu financièrement la création d'une plateforme mutualisée au service des équipes de recherche et a développé une compétence sur les études pharmaco-épidémiologiques sur le SI pour répondre à la demande des pouvoirs publics et de l'Agence nationale de sécurité du médicament.

La mise en œuvre de ces chantiers de modernisation de la base de données et sa gestion au quotidien nécessitent, compte tenu de sa taille et de sa complexité, des moyens en compétences et en infrastructures informatiques qui ne peuvent être développés que par une organisation ayant une surface suffisante.

En théorie, la création d'une institution nouvelle spécifiquement dédiée à la gestion du SI et chargée de mettre en œuvre les actions évoquées précédemment pourrait constituer une solution. Elle présente en effet les avantages inhérents à la création d'une structure à la vocation unique et resserrée : pas de conflit d'objectifs au sein de la structure, focalisation de la direction sur les objectifs assignés à l'institution, possibilité d'imputation des résultats obtenus, clarté du modèle économique, possibilité d'une plus grande souplesse par rapport à un positionnement au sein d'une administration centrale. La base serait ainsi gérée par une institution neutre dont la seule fonction serait le service au grand public d'une part, et aux divers utilisateurs d'autre part (administrations, équipes de recherche...).

Néanmoins, en pratique, la création d'une nouvelle structure publique dédiée paraît aujourd'hui peu opportune, dans un contexte où la multiplication des agences et la rareté des moyens amènent plutôt à envisager des rationalisations.

Dès lors, deux options apparaissent possibles :

- continuer à confier cette responsabilité à la CNAMTS,
- faire prendre en charge la base par la DREES.

Chacune de ces options a des avantages et des inconvénients.

La première option (CNAMTS) aurait des avantages certains : on ne s'expose pas aux risques inhérents à tout changement institutionnel, un lien étroit est maintenu entre la production des données (du moins pour la partie soins de ville, de loin la plus complexe) et l'administration du SI.

Elle présente aussi des limites :

- Le SI regroupe des données extérieures à la sphère de gestion et de production de données de la CNAMTS. Au-delà des soins de ville, il englobe les données de l'hôpital et, demain certainement, les causes de décès. Cette tendance à élargir le SI devrait se confirmer (lien avec des données médico-sociales, appariements avec des cohortes...) ;
- Les finalités du SI excèdent aussi la sphère des missions de la CNAMTS. Le SI constitue désormais un outil de la politique de sécurité sanitaire, un vecteur d'information du public sur les politiques de santé, un support pour la recherche en pharmacologie, en santé publique, en économie de la santé... ;
- Dès lors que le SI est géré par la CNAMTS, une des parties prenantes du système de santé, on ne peut dissiper le soupçon qu'elle gère un bien commun en fonction de sa perspective propre. Ce soupçon perdure malgré la création de l'IDS et se manifeste notamment dans les relations tendues entre les deux institutions ;
- La mission première de la CNAMTS n'est pas de servir la recherche, de mettre en forme des données. Les moyens de la CNAMTS sont contingentés par sa COG, il est normal qu'elle privilégie ses objectifs principaux (qualité et efficience du remboursement, gestion du risque et maîtrise médicalisée des dépenses) sur le service aux utilisateurs du SI. Les investissements à consentir (urbanisation de la base...) sont des investissements d'intérêt commun et ne relèvent pas de la CNAMTS.

En tout état de cause si la CNAMTS devait continuer à assurer cette fonction, il faudrait identifier dans sa Convention d'Objectifs et de Gestion les moyens d'investissement nécessaires à l'urbanisation du SI et à ses développements ultérieurs, isoler la fonction plate-forme de service et la doter d'un budget propre alimenté notamment par la tarification des services.

La deuxième option (DREES) a l'avantage de positionner la base comme un bien commun qui ne relève pas d'une parmi les parties prenantes du système de soins. Elle présente également certains avantages : garanties inhérentes aux administrations statistiques (rigueur, indépendance, secret...), adossement à une structure déjà constituée ayant une expérience de la gestion des données, proximité de la base et de certains des utilisateurs de données, intégration du SI dans les projets stratégiques du ministère en matière d'études et de statistiques, synergie avec des fonctions déjà assurées par la DREES notamment pour la diffusion d'informations publiques et le lien avec les chercheurs.

Elle soulève aussi des interrogations :

- La première est la question des moyens et de la capacité, dans la situation actuelle des finances publiques, de dégager des ressources dans cet objectif. Ce sont en effet des fonctions nouvelles et ambitieuses que la DREES ne peut pas assurer à moyens constants. De ce point de vue, l'assurance maladie peut être plus en situation, à court terme, de dégager des moyens par redéploiements, même s'il ne faut pas sous-estimer les contraintes fortes que fera sans doute peser la prochaine COG sur les gains de productivité de la branche.
- Il y a les risques inhérents à toute migration (constitution d'une équipe nouvelle...) ainsi que celui de créer une distance entre la plate-forme et le gestionnaire/producteur de la majorité des informations de la base (celles relatives aux soins de ville). A cet égard la poursuite de la gestion par la CNAMTS, qui a développé un savoir-faire sur ces bases, garantit une certaine sécurité.
- Comme pour la CNAMTS, ce positionnement de la base conduirait la DREES à piloter un outil servant de support à des recherches situées au delà de son domaine de compétence (vigilance sanitaire, pharmaco-épidémiologie...), à exercer des fonctions de service qu'elle n'exerce pas actuellement (plate-forme d'aide à la recherche, extraction de données ou construction de tableaux de bord à la demande...). La base serait ainsi gérée par une institution qui a vocation à l'utiliser (avantage : retour direct d'utilisateur sur la qualité de service, inconvénient : biais éventuel dans le pilotage de la base vers les besoins de la statistique publique...).

Le choix entre ces hypothèses suppose en premier lieu une expertise technique sur la séparation des fonctions de production de données brutes et de gestion du SI aujourd'hui imbriquées au sein de la CNAMTS pour ce qui concerne les soins de ville (analyse des solutions techniques, de leurs coûts et de leurs risques). Sous réserve des conclusions de cette expertise, et même s'il n'y a pas de solution qui s'impose sans discussion, les auteurs de ce rapport en pondérant les divers arguments privilégient l'option d'adossement à la DREES.

En tout état de cause, quelle que soit l'institution qui gère la base, elle doit le faire au service de tous, au nom de principes clairs et dans la plus grande transparence. C'est pourquoi, au-delà des aspects techniques de gestion de la base qui font l'objet de ce chapitre, le pilotage des actions qui seront mises en œuvre par cette organisation doit faire l'objet d'une gouvernance particulière, qui implique à la fois les utilisateurs, experts de ces données et l'ensemble des acteurs intéressés à l'amélioration et à l'usage de ce bien commun.

4.3 Quelle gouvernance ?

Il est proposé de créer un Haut conseil des données de santé, chargé de proposer aux Ministres chargés de la sécurité sociale et de la santé les choix d'orientation et les priorités à opérer dans les programmes d'amélioration des bases de données, de garantir qu'elles soient bien au service du public et de tous les utilisateurs, et à qui l'institution chargée de la gestion rendrait compte périodiquement des progrès accomplis.

Les décisions de gestion et de développement seraient prises par les ministres chargés de la santé et de la sécurité sociale après consultation de ce Haut Conseil, qui comprendrait l'ensemble des acteurs du système de santé concernés par les données regroupées dans le système d'information : représentants des patients/usagers, des professionnels de santé, des établissements de santé, des caisses de sécurité sociale et des partenaires sociaux, des organismes complémentaires, des industries de santé.

Une partie de ces acteurs est représentée aujourd'hui au sein de l'IDS, qui a conduit sur les dernières années un travail important d'information et d'élaboration d'une expertise partagée. La proposition est d'aller plus loin en confiant à une formation représentant toutes les composantes et parties prenantes intéressées un rôle plus concret d'éclairage des décisions à prendre pour faire évoluer le système d'information et optimiser ses usages.

Il est proposé également que ce Haut Conseil pour les données de santé soit doté d'un Conseil scientifique, rassemblant les experts et les utilisateurs, confrontés dans la pratique au système d'information et à même d'exprimer des besoins d'amélioration concrets. Dans cette instance pourraient être représentés les diverses disciplines scientifiques susceptibles de bénéficier du système d'information (pharmacologie, épidémiologie, santé publique, économie et gestion de la santé...) ainsi que d'autres utilisateurs experts du SI (caisses, agences, administrations, INSERM...).

Là encore, cette consultation des experts et utilisateurs existe déjà de manière non formalisée, par exemple avec les clubs d'utilisateurs animés par la CNAMTS : c'est à leur demande par exemple (et notamment à la demande des chercheurs) qu'a été développé l'EGB simplifié. Il s'agirait d'organiser cette expression des besoins pour qu'elle contribue à la fonction d'orientation confiée au Haut Conseil qui, avec les experts, rassemblerait l'ensemble des acteurs du débat démocratique.

Ces deux instances de réflexion et de surveillance devraient communiquer et échanger au sein du Haut Conseil sous des formes à définir.

Quelle que soit l'hypothèse retenue pour la gestion de la base (adossement à la CNAMTS ou à la DREES), l'institution responsable devra rendre compte régulièrement au Haut Conseil de son activité, des résultats de son action (accès du public aux données, développement de la recherche...), de l'atteinte des objectifs fixés (sécurité des données, qualité et accessibilité des données qui ont vocation à être rendues publiques, qualité du travail d'urbanisation et de documentation de la base, qualité de l'aide apportée aux chercheurs...) et lui soumettre les projets structurants pour l'évolution du SI.

Ce Haut Conseil serait placé auprès du ministre chargé de la santé et de la sécurité sociale et son secrétariat pourrait être assuré par la DREES, de la même façon que l'INSEE assure le secrétariat du Conseil national de l'information statistique (CNIS) et de l'Autorité de la statistique publique. Elle pourrait ainsi mettre au service du Haut conseil ses compétences, sa connaissance des

problématiques liées à la statistique publique et son expérience des données, en apportant une garantie de neutralité par rapport aux différentes parties prenantes.

ANNEXES

Annexe 1 – Liste des personnalités rencontrées

CNAMTS F. Van Rookeghem Directeur général, D. Polton Directrice des études et de la stratégie, P. Ricordeau, H. Caillol, S. Martinon, C. Gissot

DSS T. Fatome Directeur, F. Godineau Chef de service, A. Henrion

SG affaires sociales D. Piveteau Secrétaire général, P. Burnel (DSSIS)

DGOS : Y. Le Guen sous-directeur

DREES : F. von Lennep Directeur, N. Fourcade sous-directrice

DGS J-Y Grall Directeur, C. Poiret Chef de service, C Choma Sous-directrice

ARS IdF : Claude Evin Directeur général, Patrick Olivier Directeur de la stratégie

HCAAM : A. Morel présidente, M. Reynaud

Etalab H. Verdier Directeur, R. Lacombe Chargé de l'innovation et du développement, L. Lucchesi Directrice adjointe

Cabinet de la ministre déléguée auprès du ministre du Redressement Productif, chargée des PME, de l'Innovation et de l'Economie numérique V. Pailhes, *Conseiller technique usages du numérique* P-V Lobadowsky Chargé de mission

IDS C. Babusiaux Président, R. Decottignies Directeur

CEPS : D.Giorgi Président, J-Y Fagon

InVS A. Bruant-Bisson, Directrice générale adjointe, C Desenclos Directeur scientifique adjoint, B. Xerri : Directeur de la direction scientifique et de la qualité, A. Debever : chargé de mission, C. Hachin, juriste, A. Doussin, responsable de mission BDMA et SIP

HCSP (Grp de travail SI) : M. Goldberg président, C. Cases (Directrice générale INED), Y.Charpak (EFS), A. Doussin (InVS), H. Caillol (CNAMTS), A. Trugeon (OR2S de Picardie), S. Rey (DREES et CCTIRS)

CCTIRS : J-L. Serre (président), J. Fresson

INSEE: M. Isnard,

GENES A. Frachot Directeur general, K. Gaddouche

ATIH H. Holla Directeur, M. Bensadon Directeur adjoint

NHS England E. Doyle Head of Open Data and Transparency Patients and Information Directorate

CSMF M. Chassang Président, J-F Rey Président UMESPE

Le Bloc J. Marty Président

MG France : C. Leicher Président, G. Urbejtél

isPL : J-P. Binon Président, P. Boutin

UNPS P. Leportier Vice-Président, W. Joubert Secrétaire général, P. Corne

UFC Que Choisir M. Escot Journaliste

OR2S Picardie A Trugeon Directeur

IRDES Y Bourgueil Directeur, Z. Or Directrice de recherche, P. Dourgnon Maître de recherche

ANSM D. Maraninchi Directeur général, M. Zureik Chef de pôle épidémiologie des produits de santé

HAS JL. Harousseau Président du collège, D. Maigne Directeur, T. Le Ludec Directeur délégué, C. Rumeau- Pichon Directrice adjointe, C Grenier, Sabine Cohen-Hygounenc

Conseil national de l'Ordre des pharmaciens I.Adenot Présidente

CNIL E. Geffray Secrétaire général, H. Machi Directeur juridique, S. Nerbonne Adjointe du directeur juridique, F. Lesaulnier Responsable du pôle santé

LEEM P. Lamoureux Directeur Général, P. Zagamé Président Novartis, C. Lassalle Directrice des affaires scientifiques, C. Jouan- Flahaut Directeur médical

Celtipharm P. Guerin Président

CEGEDIM J-C Labrune Président, P. Morgon Chief Marketing Officer

IMS Health France : R. Chu Président, D. Morlet-Vigier Senior Principal health economics & outcome research

Santéclair M. Binst Directrice Générale

FNMF : Agnès Bocognano (directrice déléguée), Christophe Lapierre, Mathieu Cousineau,

FFSA A. Rouché, directeur santé

Santé-Citoyenneté (AMC) : P. Kieck et O. Kusnik (AG2R-La Mondiale), C. Cadion et N. Thoul (Allianz), A. Guillot et J. Boucher (France Mutuelles), P. Drapier (Swiss Life), J. Borella (Médéric), J-F. Tripodi (Carte Blanche), P. Delemarre

Revue **Prescrire** : Drs Schilliger et Toussaint

IRES : J-P. Moatti Président, Alfred Spira

Pr. Didier Sicard ex président du CCNE et président du comité d'experts de l'IDS

N. Renaudin

Prs. B. Bégaud et D. Costagliola

J. de Kervasdoué CNAM

C. Saout Ancien Président du collectif interassociatif sur la santé (CISS)

A. Chicoye Directeur exécutif, Institut Economie et Management de la Santé, ESSEC Business School

E. Le Fur Journaliste

R. Pellet

Annexe 2 : Glossaire

ALD Affection de Longue Durée
AMC Assurance Maladie Complémentaire
AMM Autorisation de Mise sur le Marché
AMO Assurance Maladie Obligatoire
ANSM Agence Nationale de Sécurité du Médicament et des produits de santé (ex AFSSAP)
ARS Agence Régionale de Santé
ATIH Agence Technique de l'Information sur l'Hospitalisation
BDMA Base(s) de Données Médico-administratives
CADIS Commission pour l'Analyse et la Diffusion d'Informations Statistiques (de l'IDS)
CASD Centre d'Accès Sécurisé Distant aux données confidentielles (de l'INSEE)
CCAM Classification Commune des Actes Médicaux
CCMSA Caisse Centrale de la Mutualité Sociale Agricole
CCNE Comité Consultatif National d'Éthique
CCTIRS Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé
CE Conseil d'État
CépiDc Centre d'épidémiologie sur les causes médicales de Décès (INSERM)
CEPS Comité Économique des Produits de Santé
CHU Centre Hospitalier Universitaire
CIM Classification Internationale des Maladies (de l'Organisation Mondiale de la Santé)
CISS Collectif Interassociatif Sur la Santé (réunit les associations d'usagers)
CMUC Couverture Maladie Universelle Complémentaire (couverture complémentaire gratuite)
CNAMTS Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés
CNAV Caisse Nationale d'Assurance Vieillesse
CNIL Commission Nationale de l'Informatique et des Libertés
CNIS Conseil National de l'Information Statistique
CNRS Centre National de la Recherche Scientifique
CNS Conférence Nationale de Santé
CNSA Caisse Nationale de Solidarité pour l'Autonomie
COG Convention d'Objectifs et de Gestion (entre les caisses nationales de sécurité sociale et l'État)
COPIIR Comité de Pilotage Inter Régimes (du SNIIRAM)
COTSAM Comité pour la Transparence des Statistiques de l'Assurance Maladie
CPP Comité de Protection des Personnes
CSP Code de la Santé Publique
CSS Code de la Sécurité Sociale
DAJ Délégation aux Affaires Juridiques (au ministère des affaires sociales et de la santé)
DCIR Données de Consommations Inter Régimes (du SNIIRAM)
DGFIP Direction Générale des Finances Publiques
DGOS Direction Générale de l'Offre de Soins
DGS Direction Générale de la Santé (au ministère des affaires sociales et de la santé)
DIM Département d'Information Médicale (dans un hôpital)
DREES Direction de la Recherche, des Études, de l'Évaluation et des Statistiques (au ministère des affaires sociales et de la santé)
DSSIS Délégation à la Stratégie des Systèmes d'Information de Santé (au ministère des affaires sociales et de la santé)
DSS Direction de la Sécurité Sociale (au ministère des affaires sociales et de la santé)
EDP Échantillon Démographique Permanent (de l'INSEE)
EGB Échantillon Généraliste de Bénéficiaires (du SNIIRAM)
ESP Établissement de Santé Public
ESPS Enquête Santé Protection Sociale

FNORS Fédération Nationale des Observatoires Régionaux de Santé
 FOIN Fonction d'Occultation des Informations Nominatives (procédé d'anonymisation)
 GHS Groupe Homogène de Séjours hospitaliers
 GIP Groupement d'Intérêt Public
 HAD Hospitalisation A Domicile
 HAS Haute Autorité de Santé
 HCAAM Haut Conseil pour l'Avenir de l'Assurance Maladie
 HCSP Haut Conseil de la Santé Publique
 IDS Institut des Données de Santé
 IGAS Inspection Générale des Affaires Sociales
 INCa Institut National du Cancer
 INS Identifiant National de Santé
 INSEE Institut National de la Statistique et des Études Économiques
 INSERM Institut National de la Santé Et de la Recherche Médicale
 InVS Institut de Veille Sanitaire
 IRDES Institut de Recherche et Documentation en Économie de la Santé
 IRESP Institut de Recherche En Santé Publique
 LFSS Loi de Financement de la Sécurité Sociale
 LIL Loi du 6 janvier 1978 relative à l'Informatique, aux fichiers et aux Libertés
 LPP Liste des Produits et des Prestations
 MONACO Méthodes, Outils, Normes pour la mise en commun de données entre les Assureurs
 maladie Complémentaires et Obligatoires
 NIR Numéro d'Inscription au Répertoire national d'identification des personnes physiques
 ONDAM Objectif National d'Évolution des Dépenses de l'Assurance Maladie
 ORS Observatoire Régional de la Santé
 PLASTICO PLATeforme Scientifique et Technique d'aide à la gestion de Cohortes et de grandes
 enquêtes de santé publique (projet INSERM)
 PMSI Programme de Médicalisation des Systèmes d'Information
 PS Professionnel de Santé
 RAC Reste A Charge (part non remboursée)
 RNIPP Répertoire National d'Identification des Personnes Physiques
 RNIAM Répertoire National d'Identification des bénéficiaires de l'Assurance Maladie
 RSI Régime Social des Indépendants
 SipaPH Système d'information partagé pour l'autonomie des Personnes Handicapées (CNSA...)
 SLM Section Locale Mutualiste
 SNAT IH Système NATIONAL d'Information sur l'Hospitalisation (de l'ATIH)
 SNDM Système National des Données Mutualistes
 SNGC Système National de Gestion des Carrières (de la CNAV)
 SNGI Système national de Gestion des Identifiants (de la CNAV, pour toutes les branches de la
 sécurité sociale)
 SNIIRAM Système National d'Information Inter Régimes de l'Assurance Maladie
 SSR Soins de Suite et de Réadaptation (à l'hôpital)
 T2A Tarification A l'Activité (des hôpitaux)
 UNCAM Union Nationale des Caisses d'Assurance Maladie
 UNOCAM Union Nationale des Organismes Complémentaires d'Assurance Maladie
 UNPS Union Nationale des Professionnels de Santé

Annexe 3 : Des liens utiles

Les systèmes d'information pour la santé publique - rapport du HCSP de novembre 2009

<http://www.hcsp.fr/explore.cgi/avisrapportsdomaine?clefr=114>

comportant une importante annexe descriptive sur les bases de données

http://www.hcsp.fr/Explore.cgi/Telecharger?NomFichier=hcspr20091111_sisp_src.pdf

Rapport Goldberg HCSP 2012 Pour une meilleure utilisation des bases de données médico-administratives http://www.hcsp.fr/explore.cgi/hcspr20120309_bddadmination.pdf

Site de l'Institut des données de santé <http://www.institut-des-donnees-de-sante.fr/>

Le secret statistique et la protection des données, sur le site de l'INSEE (<http://www.insee.fr/fr/insee-statistique-publique/default.asp?page=statistique-publique/secret-statistique.htm>)

Géolocalisation et codes IRIS :

<http://www.insee.fr/fr/methodes/default.asp?page=definitions/iris.htm>

Informations relatives à la réutilisation des données publiques (« open data ») sur le site de la CADA (<http://www.cada.fr/la-reutilisation-des-informations-publiques,2.html>).

Présentation du CCTIRS par sa présidente Catherine Bonaïti-Pellié (2009)

(<http://fr.scribd.com/doc/31033917/CCTIRS-MS-2009>)

Site de la cohorte CONSTANCES :

<http://www.constances.fr/fr/>

Annexe 4 : Principales sources de données en santé

Extrait de l'annexe du rapport de 2009 du HCSP sur les systèmes d'information pour la santé publique

Les sources à vocation exhaustive	Base de données de l'évaluation des programmes de dépistage organisés des cancers du sein, du colo-rectum et du col de l'utérus
	Base de données du réseau français des registres de cancer FRANCIM
	Bulletin d'interruption volontaire de grossesse (BIG)
	Centre d'épidémiologie sur les causes médicales de décès (CépiDc-INSERM)
	Certificats de santé au 8 ^{ème} jour, 9 ^{ème} mois et 24 ^{ème} mois
	Dispositif de surveillance des maladies à déclaration obligatoire
	Données de remboursement de l'assurance maladie (SNIIR-AM)
	Données des registres qualifiés
	Données issues des déclarations de vente effectuées auprès de l'AFSSAPS - Base « TAXE » (Suivi des ventes de médicaments à usage humain en France)
	Enquêtes des centres régionaux de pharmacovigilance - agence française de sécurité sanitaire des produits de santé - (AFSSAPS)
	Fichier national des accidents corporels (ONISR)
	Groupement pour l'Élaboration et la Réalisation de Statistiques (GERS)
	Observatoire des caisses de l'assurance maladie des professions indépendantes (OCAPI)
	Programme de médicalisation des systèmes d'information pour l'activité hospitalisation à domicile (PMSI HAD)
	Programme de médicalisation des systèmes d'information pour les soins de courte durée en médecine, chirurgie, obstétrique et odontologie (PMSI MCO)
	Programme de médicalisation des systèmes d'information pour l'activité soins de suite et de réadaptation (PMSI SSR)
	Rapports d'activité en psychiatrie (RAPSy)
	Recueil d'information médicalisée en psychiatrie (RIM-P)
	Registre des Victimes Corporelles d'Accidents de la Circulation Routière dans le département du Rhône.
	Réseau FranceCoag (RFC)
	Statistique annuelle des établissements de santé (SAE)
	Ventes d'alcool
Les réseaux de surveillance à partir d'échantillon	Base de données sur la qualité de l'air (BDQA) – ADEME
	Observatoire de la qualité de l'air intérieur (OQAI)
	Réseaux de surveillance du RAISIN (réseau d'alerte d'investigation et de surveillance des infections nosocomiales)
	Réseaux de surveillance en Santé Travail de l'InVS (TMS, Samotrace, MCP...)
	Système de surveillance des intoxications au CO (SIROCO)
	Système national de surveillance des plombémies chez l'enfant (SNSPE)

Les grandes enquêtes nationales en population générale déclaratives	Alimentation et état nutritionnel des bénéficiaires de l'aide alimentaire (ABENA)
	Alimentation, état nutritionnel et santé mentale des personnes âgées en institutions (ANAIIS)
	Baromètre cancer
	Baromètre santé
	Baromètre santé environnement
	Baromètre santé nutrition
	Changement organisationnel et informatisation
	Cohorte des consultants des Centres d'exams de santé (Constances)
	Cohorte santé, inégalités et ruptures sociales dans l'agglomération parisienne (SIRS)
	Contexte de la sexualité en France (CSF)
	Echantillon national témoin représentatif des personnes diabétiques (ENTRED)
	Enquête entrée en vie adulte
	Enquête fécondité, contraception et dysfonctions sexuelles (fecond)
	Enquête handicap santé (HSM et HSI)
	Enquête handicap santé aidants informels (HSA)
	Enquête IVG
	Enquête nationale auprès des usagers de centres d'accueil et d'accompagnement à la réduction des risques (Ena-CAARUD)
	Enquête nationale périnatale (ENP)
	Enquête santé et protection sociale (ESPS)
	Enquête sur la santé et les comportements lors de l'appel de la préparation à la défense (ESCAPAD)
	Enquête sur les conditions de travail
	Epidémiologie en France de l'alimentation et de l'état nutritionnel des enfants (Epifane)
	Etude de Suivi d'une Population d'Accidentés de la Route dans le Rhône (ESPARR)
	Etude longitudinale française depuis l'enfance (ELFE)
	European school survey project on Alcohol and other drugs (ESPAD)
	Recueil commun sur les addictions et les prises en charge (RECAP)
	Santé et itinéraire professionnel (SIP)
	Surveillance médicale des expositions aux risques professionnels (SUMER)
	Trajectoires et origines (TeO)
	VIH : enquête sur les personnes atteintes (VESPA)
Avec examen de santé ou prélèvements biologiques	Corpulence des enfants de 7-9 ans scolarisés en CE1 et CE2
	Enquête épidémiologique nationale de prévalence de la carie dentaire des enfants de 6 à 12 ans
	Enquête nationale de prévalence du saturnisme et de séroprévalence des maladies infectieuses chez les enfants de 6 mois à 6 ans
	Enquêtes triennales sur la santé des enfants et des adolescents scolarisés
	Etude nationale nutrition santé (ENNS)
	Etudes individuelles et nationales sur les consommations alimentaires (INCA)
Source de données échantillonnée et permanente	Echantillon généraliste de bénéficiaires (EGB)
Les grandes enquêtes nationales auprès des professionnels et structures de soins	Baromètre santé «médecins généralistes»
	Enquête nationale de prévalence des infections nosocomiales
	Enquête nationale sur les événements indésirables graves associés aux soins, observés en établissement de santé (ENEIS)
	Enquête permanente sur les accidents de la vie courante (EPAC)
	Enquête sur les opinions et pratiques des médecins du travail en matière de prévention
	Enquête sur les opinions et pratiques des médecins généralistes en matière de santé environnement et de santé au travail

Annexe 5 : Rapport OCDE – Strengthening health information infrastructure for Health Care quality governance, *Good Practices, New Opportunities and Data Privacy Protection Challenges*

Cette annexe constitue une courte synthèse d'une version préliminaire publiée le 2 avril 2013 d'un rapport de l'OCDE conçu comme support aux discussions dans les différents pays concernant le futur règlement communautaire sur la protection des données. La version finale du rapport de l'OCDE sera publiée début juin.

En 2011/12, vingt-cinq pays ont participé à cette étude de l'OCDE pour comprendre l'état d'avancement de ces pays dans le développement et l'analyse des données de santé personnelles ainsi que les raisons pour lesquelles l'utilisation de ces données peut être problématique dans certains pays. Le champ de l'enquête couvre les registres, les bases de données administratives, les enquêtes sur la population, les sondages auprès des patients et les dossiers cliniques.

Des améliorations dans les systèmes de santé et des évolutions dans la gouvernance engagées dans la majorité des pays depuis 5 ans

Pour la majorité des pays, l'infrastructure nationale mise en place pour gérer ces données s'améliore et devrait continuer d'évoluer encore dans les 5 ans à venir :

- De nombreux pays réalisent des analyses et des croisements sur des données de santé à caractère personnel pour les mettre en relation avec la qualité et le ratio coût-efficacité des traitements, pour identifier les sous-utilisations, l'abus et le mauvais usage de thérapies ou protocoles, réduire les différences dans les pratiques de soins, d'évaluer et réviser les lignes directrices des soins cliniques, et suivre les dépenses de santé.
- Certains pays prévoient que les données des systèmes de dossiers électroniques de santé soient également utilisées pour la surveillance de la qualité des soins.

Il existe des différences considérables entre les pays liées principalement à l'équilibre choisi entre : droits individuels et respect de la vie privée d'une part, droits collectifs à des soins de santé sûrs et efficaces avec un système de santé très performant d'autre part. Un des rôles que souhaite jouer l'OCDE dans les années à venir est de continuer à aider les pays à atteindre l'objectif d'amélioration et de renforcement de l'infrastructure d'information sanitaire de sorte que les analyses et les utilisations des données privées sur la santé, la qualité des soins et la surveillance du rendement du système de santé et la recherche deviennent plus répandues et plus régulières. Cela nécessitera le suivi des progrès au niveau de chaque pays et permettra ainsi la mise en place d'un référentiel de comparaison.

Cette étude indique clairement que les infrastructures nationales des données s'améliorent dans les pays auditionnés et que les capacités techniques pour analyser et collecter des données de santé personnelles sont plus importantes aujourd'hui qu'elles ne l'étaient il y a cinq ans. On peut noter cependant un problème évoqué de limitation des ressources tant du côté des organismes qui approuvent les propositions de projets que des organismes qui effectuent les analyses et les croisements de données pour le compte d'autrui.

Les cinq prochaines années semblent prometteuses, tant en termes de nombre de pays qui envisagent de mettre en œuvre des systèmes de dossiers nationaux de santé et le nombre de pays qui estiment probable que les données de ces systèmes seront utilisées pour enrichir des bases de données existantes de santé et pour les croisements avec de multiples sources de données

La France bien positionnée pour davantage capitaliser sur son patrimoine de données en santé

En comparaison des 19 autres pays, la France est bien positionnée en ce qui concerne la disponibilité des bases de données individuelles de santé à l'échelle nationale et les utilisations régulières des données liées à la santé publique et à la recherche.

Table D.1. Data available at a national level

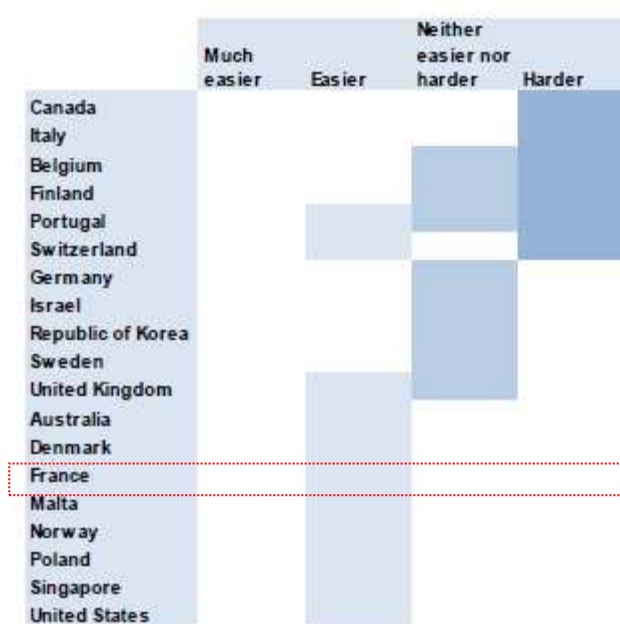
	Hospital in-patient data	Primary care data	Cancer registry data	Prescription medicines data	Mortality data	Formal long-term care data	Patient experiences survey data	Mental hospital in-patient data	Population health survey data	Population census or registry data
Australia	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Belgium	Yes	Yes	Yes	Yes	Yes	Yes	nr	Yes	Yes	Yes
Canada	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Denmark	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Finland	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
France	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Germany	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Israel	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Japan	Yes	Yes	No	Yes	Yes	Yes	nr	nr	Yes	Yes
Korea	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Malta	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Norway	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Poland	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Portugal	Yes	Yes	Yes	Yes	Yes	nr	nr	Yes	Yes	Yes
Singapore	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Sweden	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Switzerland	Yes	No	No	No	Yes	Yes	No	Yes	Yes	Yes
United Kingdom	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
United States	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Total Yes	19	16	17	14	19	16	11	17	19	19

Note: The data custodian should be a national authority and data should be included even when it does not cover 100% of the nation.
nr: no response.

Source: OECD HCQI Questionnaire, Secondary Use of Health Data, 2011/12.

La France est reconnue comme disposant de processus en place depuis de nombreuses années pour mettre à disposition des données pour la recherche et les statistiques. Les dispositifs utilisés en France offrent un niveau élevé de sécurité pour les données et l'application de ces dispositifs de sécurité a été plus répandue en France que dans d'autres pays. La France travaille à faire en sorte que les données soient traitées d'une manière systématique et cohérente pour les statistiques la recherche et la santé publique.

Figure 8.1. Over the past five years, has it become easier or harder to use personal health data to monitor health and health care quality?



Note: In some countries there were differences of opinion and all responses are reported. No opinion was expressed by Japan.

Source: OECD HCQI Questionnaire, Secondary Use of Health Data and Follow-up Telephone Interviews, 2011/12.

Le suivi des patients à travers le parcours de soins, et d'une base de données à l'autre, nécessite des informations d'identification du patient. Plus de la moitié des pays indiquent que leurs bases de données nationales de santé contiennent des numéros d'identification uniques des patients. Mais contrairement à la France, seulement la moitié apparie et chaîne des données pour suivre la qualité des soins de santé. Au cours des cinq dernières années, il est devenu plus facile de mener des projets de couplage des données en France car le processus d'approbation des projets semble mieux établi et bien compris en comparaison d'autres pays. D'un autre côté, les bases de données et les techniques de couplage des données nécessaires pour les analyser sont de plus en plus complexes.

Table 3.3. Distribution of the regular occurrence of health-related record linkage projects by availability of databases with patient identifiers

	Most national data with a unique patient identifying number (UPI)	Most national data with other patient identifiers	Some national data with a unique patient identifying number (UPI)
Data linkage projects on a regular basis...			
With 7+ national databases	Denmark, Finland, Israel, Korea, Sweden, United Kingdom	United States	
With 5-6 national databases	France, Belgium	Australia	Switzerland
With 3-4 national databases	Canada, Malta, Norway		
With 2 national databases	Singapore		Portugal
None	Japan		Poland, Germany

Source: OECD HCQI Questionnaire, Secondary Use of Health Data, 2011/12.

Sept pays réalisent régulièrement des projets de couplage de données impliquant de nombreuses bases de données nationales (Danemark, Finlande, Israël, la Corée, la Suède, le Royaume-Uni et aux

États-Unis). La Belgique et la France ont une plus grande capacité à mener ces appariements en utilisant un numéro de patient unique d'identification, tandis que d'autres identificateurs sont plus souvent utilisés en Australie et en Suisse. L'Allemagne, le Japon et la Pologne ont des bases de données avec des variables qui pourraient être utilisées pour réaliser des projets de couplage des données, mais aucun de ces pays ne le fait régulièrement.

Annexe 6 : Panorama des principaux dispositifs existants à l'étranger

Cette annexe présente une synthèse des dispositifs existants au Royaume-Uni et dans les principaux pays européens.

Un focus particulier est réservé aux systèmes britanniques :

- « Health & Social Care Information Centre » (HSCIC), mis en place au 1^{er} avril 2013
- "Quality Outcome Framework " (QOF) du NHS

Ces dispositifs incluent, comme en France le PMSI, des données hospitalières de type résumés de sortie mais sont aussi fondées sur la centralisation de données, individuelles ou non, offrant un reflet des principaux signes cliniques des patients et de leurs prises en charge thérapeutique (à la différence du SNIIRAM qui est issu de données de facturation). Elles sont conçues pour permettre la production de tableaux de bord structurés selon les enjeux jugés prioritaires en santé publique mais le chaînage des données individuelles peut s'avérer plus difficile.

Annexe 6.1 : « Health & Social Care Information Centre » du NHS

Le « Health & Social Care Information Centre » (HSCIC) a été créé le 1er avril 2013, avec un statut de « Executive Non Departmental Public Body » (l'équivalent de nos Agences sanitaires) ; il succède ainsi au NHS Information Centre.

Son objectif est d'être l'agence de référence pour le suivi et la mise à disposition des données permettant de rendre compte de la prise en charge des patients et de la qualité des soins.

Ses réalisations récentes sont, pour l'essentiel, les suivantes :

- Mettre à disposition des chercheurs - avec le Medical Research Information Service (MRIS) - des fonctions facilitant la gestion de cohortes patients ; prendre en charge, dans cette logique, la qualité et l'intégrité des données centralisées, avec le service Integrated Database and Administration System (MIDAS)
- Assurer, dans une logique d'Open Data, la diffusion des données médico sanitaires, en liaison avec le NHS et les travaux engagés par ce dernier (cf. www.data.gouv.uk)
 - o Près de 1 000 séries d'indicateurs, couvrant l'ensemble des pathologies chroniques ou de longue durée, sont ainsi déjà mis à disposition sur son site

Son programme de travail couvre, pour la période 2013 -14, les sujets suivants :

- Collecter pour le compte des différents services du Royaume Uni et mettre à la disposition des utilisateurs les informations reflétant la prise en charge sanitaire avec les informations cliniques majeures des patients (cf. fonctions A et B, au § 4.2 La gestion des bases de données médico-administratives : les fonctions)
- Mettre en place un service sécurisé d'appariement des données, capable de fournir des sorties de données chaînées anonymisées pour une large population d'utilisateurs (cf. fonction B au chapitre 4) :

- o Publics
 - o Privés
- Proposer de nouveaux modes de financement de ces travaux, basés sur une logique de facturation auprès de certains acteurs

Il reprend, pour ce faire, l'infrastructure technique de deux agences : NHS Connecting for Health - CfH - et Strategic Health Authorities), qu'il internalise.

Il ressort des contacts pris avec le NHS que le chaînage des données hospitalières entre elles (soins aigus et psychiatrie) semble toutefois progresser plus vite que la collecte des données en provenance des cabinets de. L'information en provenance des pharmacies est peu détaillée et les GPs et le chaînage de l'ensemble se heurte à la qualité inégale des identifiants.

En 2012, son budget se situait ainsi à 37 M£.

Annexe 6.2 : Système QOF (Quality Outcome Framework) du NHS

Description

Le QOF est un programme d'amélioration de la qualité des soins dont les engagements font l'objet d'une contractualisation, sur la base du volontariat, avec les médecins d'un « primary care centre ».

Les « GPs » des centres adhérant à ce contrat s'engagent à respecter des critères qualité fondés sur des indicateurs de moyens et des indicateurs de résultats (données sources : *Read Codes*, examens biologiques, paramètres cliniques), tels que :

- Proportion de patient avec une maladie coronarienne pour lequel un examen de cholestérol a été réalisé dans l'année.
- Pourcentage de patients avec une dépression qui ont répondu à un questionnaire d'évaluation de la sévérité.
- Pourcentage de patient hypertendu pour lesquels la mesure de tension (dans les 9 derniers mois) est inférieure à 150/90.
- Pourcentage de patients pour lesquels les prescriptions de médicaments sont reliées à un diagnostic ou des signes et symptômes.
- ...

CHD 10.2 Reporting and Verification

The percentage of patients on the CHD register who have been prescribed a beta blocker in the last 6 months.

CHD Indicator 11

The percentage of patients with a history of myocardial infarction (diagnosed after 1 April 2003) who are currently treated with an ACE inhibitor or angiotensin II antagonist

CHD 11.1 Rationale

A number of trials have shown reduced mortality following myocardial infarction with the use of ACE inhibitors. The Heart Outcome Prevention Evaluation (HOPE) showed that ACE inhibitors are also of benefit in reducing coronary events and progression of coronary arteriosclerosis in patients without left ventricular systolic dysfunction. There is evidence that angiotensin II antagonists have a similar effect. This indicator is prospective with inclusion of patients diagnosed with a myocardial infarction after 1 April 2003.

Grade A Recommendation SIGN Guideline 41
Grade A Recommendation NICE Guideline A

Further Information: <http://www.sign.ac.uk/guidelines/fulltext/41/index.html>
Further information: <http://www.escardio.org/scinfo/Guidelines/98prevention.pdf>

Exemple d'indicateur QOF

Il existe environ 130 critères, couvrant quatre domaines :

- Clinique : 80 indicateurs concernant 19 thèmes cliniques (maladie coronarienne, hypertension, asthme...).
- Organisationnel : 43 indicateurs sur 5 thèmes : dossiers médicaux et information, information des patients, éducation et formation, gestion du centre, gestion des médicaments.
- Retour/satisfaction patient : 4 indicateurs concernant la durée des consultations et des enquêtes patient.
- Autres services : 8 indicateurs (contraception, protection de l'enfance...).

Chaque année, les GPs adhérant à ce programme doivent satisfaire un objectif pour chacun des critères qualité. Par exemple ci-dessous :

Mental Health (MH)

Indicator	Points	Payment stages
Records		
MH 1. The practice can produce a register of people with severe long-term mental health problems who require and have agreed to regular follow-up	7	
Ongoing management		
MH 2. The percentage of patients with severe long-term mental health problems with a review recorded in the preceding 15 months. This review includes a check on the accuracy of prescribed medication, a review of physical health and a review of co-ordination arrangements with secondary care	23	25-90%
MH 3. The percentage of patients on lithium therapy with a record of lithium levels checked within the previous 6 months	3	25-90%
MH 4. The percentage of patients on lithium therapy with a record of serum creatinine and TSH in the preceding 15 months	3	25-90%
MH 5. The percentage of patients on lithium therapy with a record of lithium levels in the therapeutic range within the previous 6 months	5	25-70%

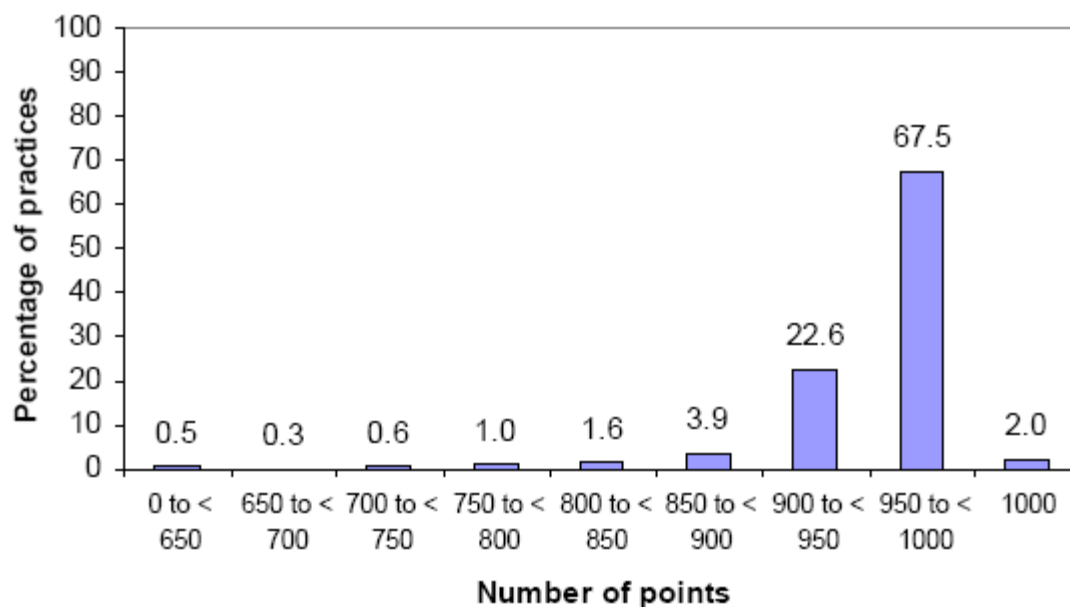
Un certain nombre de points de performance sont attribués aux GPs du « primary care centre » au regard de l'objectif indiqué.

Le programme QOF définit les critères et les objectifs à atteindre au niveau national ; les PCT (autorités locales du NHS) peuvent choisir de demander des objectifs plus élevés sur certains des critères nationaux en fonction de la population du territoire de santé. Cette adaptation locale du programme national s'appelle QOF+.

Les programmes QOF et QOF+ sont basés sur une démarche incitative ; ainsi, chaque année les professionnels de santé perçoivent un financement basé sur l'activité du centre le nombre de points obtenus.

Ce programme a commencé à être proposé aux GPs à partir de 2004 ; désormais, la plupart des « primary care centres » adhèrent à ce dispositif.

A titre d'exemple, le tableau ci-dessous présente les résultats atteints par les 8 000 centres impliqués en Angleterre.



Points obtenus par les centres sur une année

Système d'information associé

Ce dispositif s'est accompagné de la mise en place d'un système d'information global avec :

- Une application nationale pour calculer le niveau d'atteinte des critères par les centres : Quality Management Analysis System (QMAS).
- La définition d'un protocole de transfert d'indicateurs médicalisés entre les « primary care centers » et le QMAS :
 - o Structurés par thématique de santé publique
 - o Produits :
 - Dans une logique d'agrégats, en règle générale,
 - A une maille individuelle et après anonymisation de l'identifiant patient, par exception

Extraits du site <http://www.ic.nhs.uk/datalinkage>

Data linkage services

Unlocking the full potential of data

There is a wide range of data that exists within health and social care. This usually exists for a single purpose and often contains patient level information. The data linkage service adds significant value to individual sets of data by combining and matching them at an individual record level in a secure environment. Maintaining the confidentiality of personal details is paramount and this is achieved by removing identifiable information before any linked data is released.

What is the data linkage service?

Our service provides extracts of anonymised data which are created by:

- Linking data held by the Health and Social Care Information Centre (HSCIC)
- Linking data held by the HSCIC to customer data
- Linking two or more sets of customer data

When stringent Information Governance controls allow we can also provide extracts of linked data sets in an identifiable form.

We have a strong track record in the provision of data linkage services, having provided the Medical Research Information Service (MRIS) since the 1960s. Additionally, we developed a service to link data to Hospital Episode Statistics (HES) in 2011. The Data Linkage Service incorporates these services and builds on this heritage.

Who is the service for?

Any individual or organisation can use our data linkage service. Our customers include commissioners, care providers, academics, universities, clinical auditors, information intermediaries, researchers, pharmaceutical and life science companies.

If you are conducting formal, protocol-based research then we may refer you to the [Clinical Practice Research Datalink \(CPRD\)](#) (external) (the English NHS Observational Data and Interventional Research Service).

- Des évolutions des dossiers médicaux informatisés des centres de « primary care » :
 - o Des rappels des recommandations des programmes QOF et QOF+.
 - o Des fonctions d'aide au codage des informations afin que les GPs utilisent les « read codes » choisis par le NHS pour calculer les critères.
 - o Des tableaux de bord locaux pour évaluer localement l'atteinte des objectifs.
 - o La possibilité d'exclure, dans le respect d'un cadre prédéfini, des patients du calcul de certains indicateurs (patients refusant les soins, patients récemment inscrits dans le centre...)
 - o La fonction de transmission des données vers le QMAS, qui fait l'objet d'une certification.
- Un site Internet permettant aux administrateurs des centres de consulter le nombre de points acquis par leur centre dans le QMAS, et comparer la performance du centre par rapport aux autres centres.

Ima NURSE 36Y - 21/01/1973 (F) 7 Little Grange, Perivale Lane, Greenford, UB6 8TZ - (OFFLINE) - [COPD Quality Indicators v10.0.00]

Consultation Summary Guidelines Add List View Window Help

21/10/2009 Inhaler technique - good quality - COPD Quality Indicators v10.0.00

No data recorded.

Inhaler technique observed Inhaler technique good Inhaler technique moderate Inhaler technique poor

Inhaler technique in last 15m

Flu vaccination

COPD 8 Rule (6 Points)

No data recorded.

No data recorded.

No data recorded.

- COPD patients should have a flu vaccination every year (now 1/9/2007 - 31/3/2008), unless

- History of adverse effect EVER

- Contraindicated, not indicated or refused - annual record required (after 1/1/2007)

Flu vaccination given

Last 3 influenza vaccinations

Flu vaccination adverse effect (persisting exception)

- History of adverse effect EVER

No data recorded.

No data recorded.

H/O allergy flu vaccination Influenza vaccination adverse reaction

Flu vaccination contraindicated, not indicated or refused (expiring exception)

- Contraindicated, not indicated or refused - annual record required (after 1/1/2007)

No data recorded.

No data recorded.

Flu vaccination not indicated (Hx) Flu vaccination contraindicated

- Two buttons have been provided for Flu Vacc Refused; the first enters refusal onto Medical History, the second enters refusal onto the Immunisation screen.

- Note if you are in Wales, you should use the first button and record in Medical History to ensure inclusion in the MIQUEST query

21/10/2009 FLU Stage: 0 Refusal to start or complete course Due 21/10/2010 Dr J Gittins

No data recorded.

No data recorded.

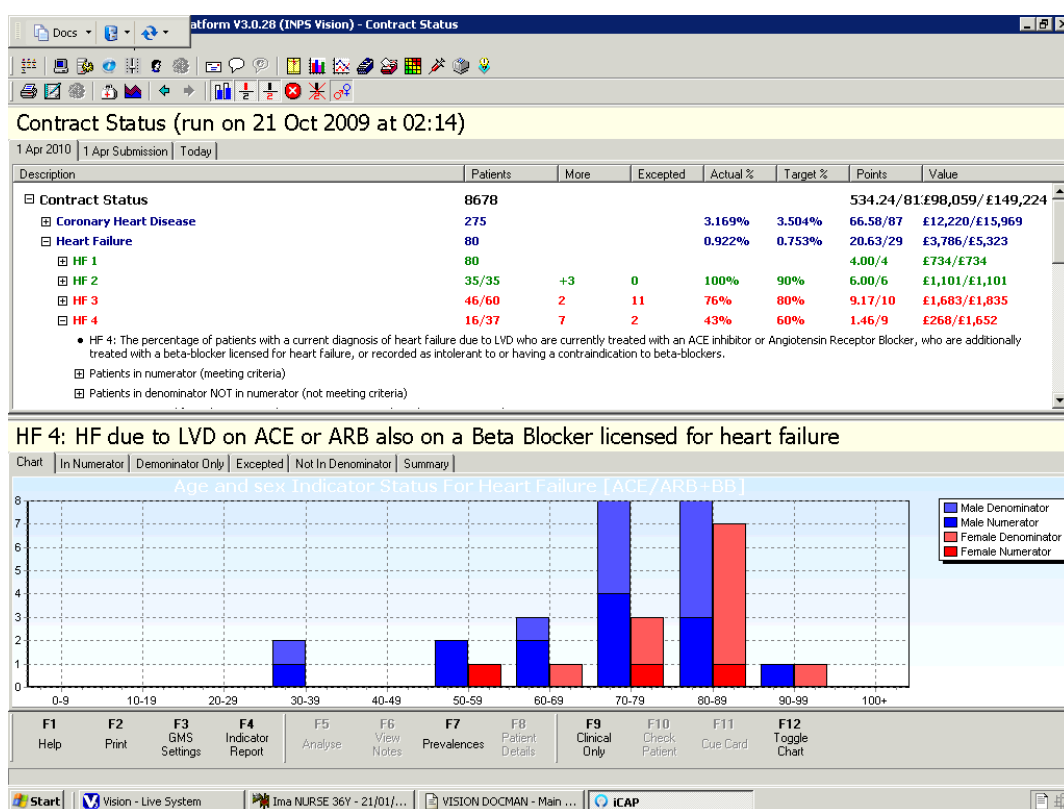
Flu vaccination declined (Hx) Flu vaccination refused (Imms)

Expirations in last 45m

COPD Quality Indicators v10.0.00

Start Vision - Live System Ima NURSE 36Y - 21/01/... VISION DOCMAN - Main ... ICAP P10 - Paint

Exemple de rappel des recommandations et d'accès aux formulaires d'aide au codage.



Exemple de fonction de suivi d'indicateurs.

Certification des dossiers médicaux informatisés

Les systèmes utilisés par les GPs font l'objet d'une certification par le NHS. Cette certification comprend :

- La gestion et la transmission des indicateurs QOF.
- L'enregistrement d'un nouveau patient dans le centre (transfert avec le PCT).
- Pathology Messaging Implementation Project, réception de données de laboratoire.
- La gestion de la prescription
- Le transfert électronique de la prescription (un serveur central assure la relation entre les « primary care centres » et les pharmacies). Les posologies sont complexes à standardiser et sont exprimées en texte libre et interprétées par les pharmaciens.
- Le transfert de dossiers patient entre « primary care practices » lorsqu'un patient choisit un nouveau centre (GP2GP).
- La traçabilité des décisions d'orientation des patients vers les soins spécialisés « referral » avec un accès à l'application nationale de prise de rendez-vous en ligne (Choose and Book).

Conclusion

Du point de vue de l'aide à la décision, l'outil de gestion des QOF donne accès à des synthèses collectives pour le pilotage (rapports de pilotage, statistiques permettant aux praticiens d'évaluer leur niveau d'avancement et les financements en rapport) dont certains font la relation avec les identités des patients. Il ne repose cependant pas sur une base de données analogue au SNIIRAM :

- Pas de données de facturation, compte tenu de l'organisation du NHS ; à l'inverse, transmission d'informations cliniques et thérapeutiques permettant de juger du respect des « bonnes pratiques » notamment
- Pas de données individuelles centralisées, en règle générale
- Production de tableaux de bord structurés selon les enjeux jugés prioritaires en santé publique

Annexe 7 : Gestion des données à caractère personnel - Comparaison des systèmes d'information sanitaire européens et canadien (Congrès ADEL-EMOIS 2012)

Description

Cette étude présente une analyse comparative des modalités de gestion des données à caractère personnel de systèmes d'information sanitaire (Danemark, Royaume-Uni, Canada) et propose des pistes de renforcement d'un système statistique européen.

Cette étude comparative permet de suggérer certaines pistes pour d'autres systèmes d'information sanitaire : renforcer l'indépendance de l'autorité statistique, mettre en place des procédures de contrôle de qualité et favoriser la transparence dans le traitement de l'information statistique. Enfin, l'utilisation de données administratives nécessite une coordination efficace entre tous les producteurs de données.

Méthodologie

Cette analyse se base sur la revue de la littérature et des entretiens avec des experts responsables des institutions statistiques dans les différents contextes d'étude.



UNIVERSITÉ LIBRE DE BRUXELLES,
UNIVERSITÉ D'EUROPE

Gestion des données à caractère personnel: comparaison de systèmes d'information sanitaire européens

Yves Coppieters, Anne Ingenbleek & Alain Levêque

École de santé publique
de l'Université Libre de Bruxelles,
Belgique

Congrès Adelfémois – Dijon, mars 2012

Gestion des données ? Éthique ? Vie privée ? Cadre légal ?

- Toute recherche scientifique sur des données personnelles doit aborder les questions de la gestion des données à caractère personnel, du cadre institutionnel, d'éthique et de protection de la vie privée
 - Peu développé dans les formations médicales de base
 - Confusion chez les professionnels entre les aspects éthiques, législatifs et de protection de la vie privée
 - Cadres institutionnels et législatifs très différents d'un contexte à l'autre
-

Objectifs

- Analyse comparative de cadres institutionnels et des modalités de gestion des données à caractère personnel de différents systèmes d'information sanitaire européens :
 - Danemark
 - Royaume-Uni
 - Belgique
 - +
 - Canada

Evolution de la recherche dans les sociétés « modernes »

Contexte technologique :

- Utilisation de dossiers électroniques, des trajets de soins, de plateformes Web
- Développement de techniques statistiques sophistiquées, de bases de données de population à grande échelle, etc.

Contexte social :

- Depuis une dizaine d'années, le rôle de « l'information » a changé
 - menaces potentielles pour la sécurité nationale > protection de la vie privée des citoyens

Collecte des données :

- Utilisation de statistiques sur base de données administratives
- Lien entre les bases de données
- Accent accru sur les données longitudinales (couplages)
- Nécessité du consentement éclairé
- Augmentation du risque d'identification des individus

4

'Politiques de santé vs respect de la vie privée'

	Surveillance sanitaire	
A court terme	Politiques de santé plus précises ↔ Respect / menace vie privée	
A long terme	Contrôle sur la santé ?	Valeurs démocratiques ?

5

Différentes approches



Eléments qui différencient les approches

- 1. Historique / idéologique
 - Centralisation vs décentralisation
 - 2. Gestion des données : données administratives vs d'enquêtes
 - 3. Numéros d'identification uniques vs multiples
 - 4. Accès et gestion des données à caractère personnel
 - 5. Sécurisation des données et évolutions technologiques
 - 6. ...
-

Système danois

- Basé sur une gestion centralisée (données administratives)
 - Grande variété de données :
 - Usage large des données administratives couvrant toute la population
 - Usage de nombreux registres
 - Numéro d'identification unique tant pour les personnes que pour les sociétés
 - Méthodes statistiques et informatiques modernes
 - Facilitation du couplage des données
 - Bonne protection de la vie privée, de la confidentialité et du secret statistique
 - Bonne confiance de la population dans le système
 - Accès de plus en plus ouvert aux données : connaissances ↑
-

8

Système britannique

- Basé sur une gestion décentralisée (géographique et «thématique»)
 - Peu de recours aux données administratives
 - Beaucoup de sources de données ou d'études développées isolément des unes des autres
 - Objectifs différents,
 - Méthodes de collecte différentes,
 - Définitions et concepts non standardisés.
- ... de nombreuses sources de données
- **Les recensements populations**
 - **Les études épidémiologiques transversales et longitudinales**
 - **Quelques données issues d'administrations**
-

9

Système britannique

- Faible confiance de la population dans le système
 - Faiblesses des systèmes informatiques (2007: perte / vol de données de 25 millions de personne)
 - En cours de réforme de son système administratif et statistique
 - Atouts :
 - indépendance des autorités statistiques conforme à la déclaration de qualité européenne
 - Développement d'un [Plan National](#)
-

10

Au niveau opérationnel

NE

1. Création de **12 Theme working group** qui ont comme missions de définir les besoins de données de domaines spécifiques
2. La création en 2005 d'un **UK Data Forum**
 - constitué de représentants des principaux départements gouvernementaux, différentes organisations ou institutions actives dans le financement ou la production de données.
 - pour réfléchir sur les études futures et les priorités en termes d'indicateurs

11

Au niveau opérationnel

3. Création **UK longitudinal Studies center**

- développer une stratégie nationale propre aux études longitudinales
 - Renforcer les études longitudinales existantes
 - Identifier les lacunes en terme de population
 - Favoriser la complémentarité entre les études
 - Faciliter l'accès aux données et leur utilisation
 - Favoriser le linkage entre bases de données d'études et bases de données officielles

12

Création d'un organisme indépendant au Canada

L'**Institut canadien d'information sur la santé (ICIS)** est un organisme national autonome financé et sous la responsabilité des ministres de la Santé.

Ses missions sont de :

- **Coordonner** l'élaboration et le maintien d'un système complet et intégré d'information sur la santé
- **Définir les besoins en information** sur la santé et les priorités en la matière;
- **Effectuer des analyses** et des études spéciales et appuyer la recherche sur le système de santé ou y participer;
- Appuyer l'élaboration **d'indicateurs** de santé à l'échelle nationale;
- Coordonner et favoriser l'élaboration et la mise à jour **de normes** nationales d'information sur la santé;
- Créer et gérer des **bases de données** et des registres;
- **Fournir un accès** approprié aux données sur la santé;
- **Publier des rapports** et diffuser des renseignements sur la santé.

Sources des données

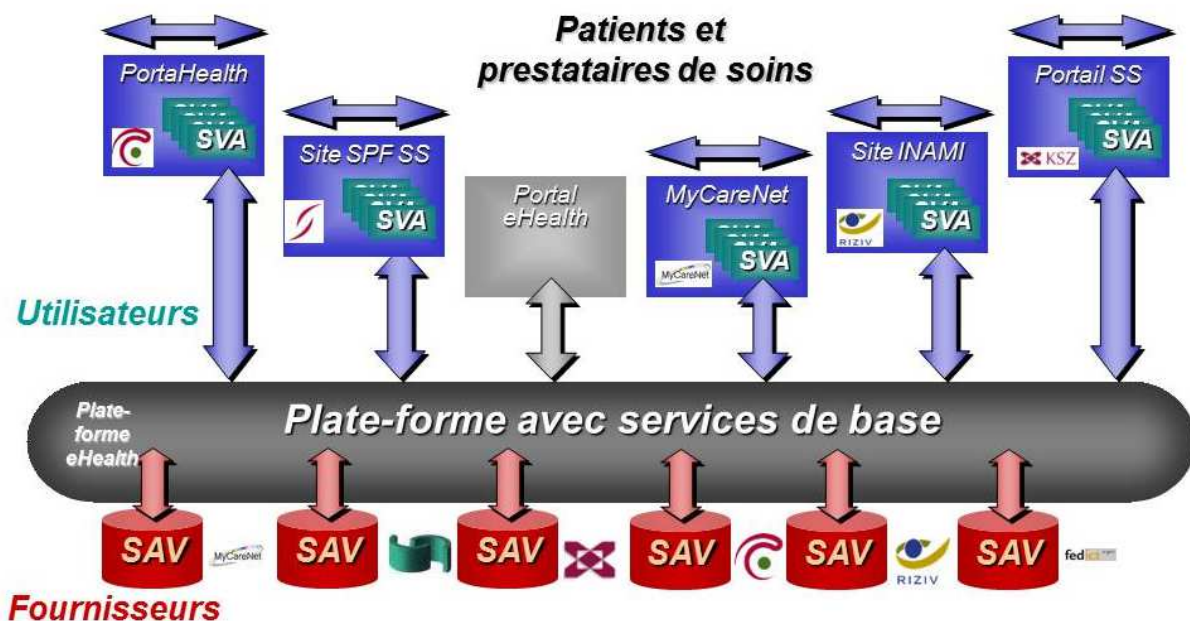
Les informations sanitaires proviennent principalement :

- des ministères de la Santé,
- des hôpitaux,
- des centres de soins de longue durée,
- d'autres organismes de soins de santé.

Les dossiers de l'ICIS comprennent un identificateur unique afin de faciliter l'analyse statistique et la recherche (numéro d'assurance-maladie)

14

La plate-forme eHealth



Système belge

- Développement d'une architecture globale décentralisée de l'information issue du système de santé (historique !)
 - Appui des technologies de l'information et de la communication
 - Augmentation potentielle du volume des enregistrements
 - Accès facilité aux méthodes de cryptage
 - Sécurité accrue (données, transactions, personnes)
 - Développement de nouveaux outils et services
 - dossier médical informatisé
 - plates-formes d'échange (cliniques ou administratives)
 - Implications pour la vision longitudinale
 - Rapprochement de bases de données
-

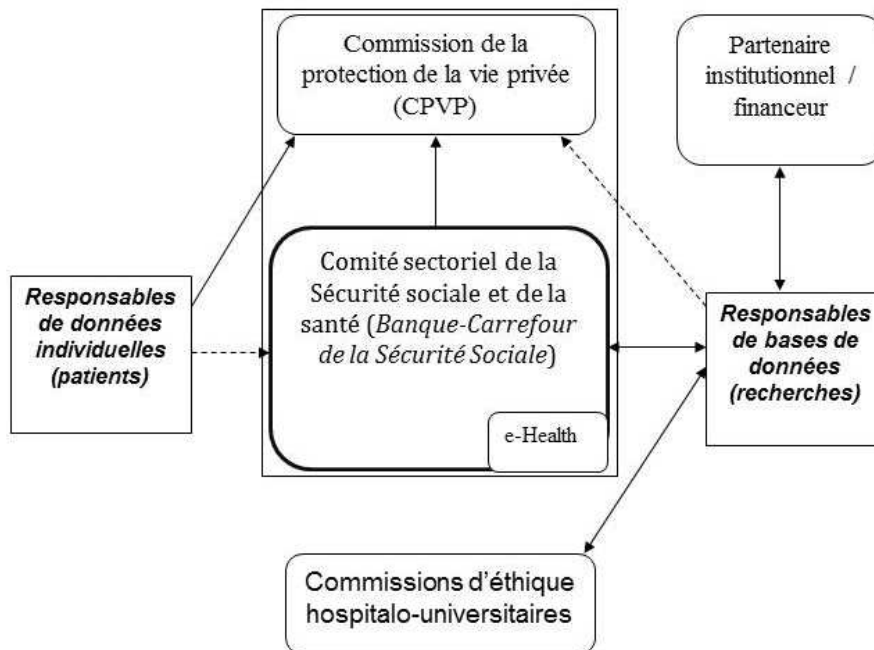
15

Données à caractère personnel

- **Belgique** : pas de Comité d'éthique national
 - **Différents Comités d'éthique hospitalo-universitaires**
 - **3 fonctions (A.R. 12/08/1994)** :
 - une fonction d'**avis** sur tout protocole d'expérimentation sur l'homme
 - une fonction d'**accompagnement** et de **conseil** concernant les aspects éthiques de la pratique des soins hospitaliers
 - une fonction d'**assistance à la décision** concernant les cas individuels, en matière d'éthique
-

17

Procédures de déclaration de traitement de données médicales à caractère personnel



18

Commission de la protection de la vie privée (CPVP)

- **Chargée de rendre des avis** en matière de vie privée dans le cadre de traitements de données ainsi que de traiter des plaintes émanant du citoyen
- **Pas d'obligation légale** de solliciter un comité d'éthique
- La CPVP ne va donc à priori jamais refuser une recherche mais elle peut bloquer une demande si elle ne correspond pas à la procédure de la Loi vie privée.

19

Conclusion

Pistes communes :

- Favoriser la coordination (et la cohérence) entre les fournisseur et utilisateurs de données dans le cadre de l'utilisation de données administratives
- Favoriser les liens entre données administratives et d'enquêtes
- Renforcer l'indépendance et la transparence dans le traitement de l'information statistique
- Mettre en place des procédures de contrôle de qualité
- Favoriser l'accès aux données : suivi longitudinal, couplage, études ponctuelles, ...
- ...

Cela entraîne des ...

- Contraintes opérationnelles et institutionnelles
- Intérêt de systèmes centralisés / décentralisé ?
- Nécessité de faire évoluer les cadres institutionnels et opérationnels pour toutes recherches sur les données à caractère personnel, tant publiques que privées.
- Équilibre entre recherche, politique de santé, aspects éthiques et juridiques, développements technologiques et évolution des besoins des sociétés / systèmes sociaux.

Annexe 8 : Loi Informatique et Libertés, extraits

Article 8

I.-Il est interdit de collecter ou de traiter des données à caractère personnel qui font apparaître, directement ou indirectement, les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses ou l'appartenance syndicale des personnes, ou qui sont relatives à la santé ou à la vie sexuelle de celles-ci.

II.-Dans la mesure où la finalité du traitement l'exige pour certaines catégories de données, ne sont pas soumis à l'interdiction prévue au I :

1° Les traitements pour lesquels la personne concernée a donné son consentement exprès, sauf dans le cas où la loi prévoit que l'interdiction visée au I ne peut être levée par le consentement de la personne concernée ;

2° Les traitements nécessaires à la sauvegarde de la vie humaine, mais auxquels la personne concernée ne peut donner son consentement par suite d'une incapacité juridique ou d'une impossibilité matérielle ;

3° Les traitements mis en œuvre par une association ou tout autre organisme à but non lucratif et à caractère religieux, philosophique, politique ou syndical :

-pour les seules données mentionnées au I correspondant à l'objet de ladite association ou dudit organisme ;

-sous réserve qu'ils ne concernent que les membres de cette association ou de cet organisme et, le cas échéant, les personnes qui entretiennent avec celui-ci des contacts réguliers dans le cadre de son activité ;

-et qu'ils ne portent que sur des données non communiquées à des tiers, à moins que les personnes concernées n'y consentent expressément ;

4° Les traitements portant sur des données à caractère personnel rendues publiques par la personne concernée ;

5° Les traitements nécessaires à la constatation, à l'exercice ou à la défense d'un droit en justice ;

6° Les traitements nécessaires aux fins de la médecine préventive, des diagnostics médicaux, de l'administration de soins ou de traitements, ou de la gestion de services de santé et mis en œuvre par un membre d'une profession de santé, ou par une autre personne à laquelle s'impose en raison de ses fonctions l'obligation de secret professionnel prévue par l'article 226-13 du code pénal ;

7° Les traitements statistiques réalisés par l'Institut national de la statistique et des études économiques ou l'un des services statistiques ministériels dans le respect de la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques, après avis du Conseil national de l'information statistique et dans les conditions prévues à l'article 25 de la présente loi ;

8° Les traitements nécessaires à la recherche dans le domaine de la santé selon les modalités prévues au chapitre IX.

III.-Si les données à caractère personnel visées au I sont appelées à faire l'objet à bref délai d'un procédé d'anonymisation préalablement reconnu conforme aux dispositions de la présente loi par la Commission nationale de l'informatique et des libertés, celle-ci peut autoriser, compte tenu de leur finalité, certaines catégories de traitements selon les modalités prévues à l'article 25. Les dispositions des chapitres IX et X ne sont pas applicables.

IV.-De même, ne sont pas soumis à l'interdiction prévue au I les traitements, automatisés ou non, justifiés par l'intérêt public et autorisés dans les conditions prévues au I de l'article 25 ou au II de l'article 26.

Article 25

I. - Sont mis en œuvre après autorisation de la Commission nationale de l'informatique et des libertés, à l'exclusion de ceux qui sont mentionnés aux articles 26 et 27 :

1° Les traitements, automatisés ou non, mentionnés au 7° du II, au III et au IV de l'article 8 ;

2° Les traitements automatisés portant sur des données génétiques, à l'exception de ceux d'entre eux qui sont mis en œuvre par des médecins ou des biologistes et qui sont nécessaires aux fins de la médecine préventive, des diagnostics médicaux ou de l'administration de soins ou de traitements ;

3° Les traitements, automatisés ou non, portant sur des données relatives aux infractions, condamnations ou mesures de sûreté, sauf ceux qui sont mis en œuvre par des auxiliaires de justice pour les besoins de leurs missions de défense des personnes concernées ;

4° Les traitements automatisés susceptibles, du fait de leur nature, de leur portée ou de leurs finalités, d'exclure des personnes du bénéfice d'un droit, d'une prestation ou d'un contrat en l'absence de toute disposition législative ou réglementaire ;

5° Les traitements automatisés ayant pour objet :

- l'interconnexion de fichiers relevant d'une ou de plusieurs personnes morales gérant un service public et dont les finalités correspondent à des intérêts publics différents ;

- l'interconnexion de fichiers relevant d'autres personnes et dont les finalités principales sont différentes ;

6° Les traitements portant sur des données parmi lesquelles figure le numéro d'inscription des personnes au répertoire national d'identification des personnes physiques et ceux qui requièrent une consultation de ce répertoire sans inclure le numéro d'inscription à celui-ci des personnes ;

7° Les traitements automatisés de données comportant des appréciations sur les difficultés sociales des personnes ;

8° Les traitements automatisés comportant des données biométriques nécessaires au contrôle de l'identité des personnes.

II. - Pour l'application du présent article, les traitements qui répondent à une même finalité, portent sur des catégories de données identiques et ont les mêmes destinataires ou catégories de destinataires peuvent être autorisés par une décision unique de la commission. Dans ce cas, le responsable de chaque traitement adresse à la commission un engagement de conformité de celui-ci à la description figurant dans l'autorisation.

III. - La Commission nationale de l'informatique et des libertés se prononce dans un délai de deux mois à compter de la réception de la demande. Toutefois, ce délai peut être renouvelé une fois sur décision motivée de son président. Lorsque la commission ne s'est pas prononcée dans ces délais, la demande d'autorisation est réputée rejetée

Article 27

I.- Sont autorisés par décret en Conseil d'État, pris après avis motivé et publié de la Commission nationale de l'informatique et des libertés :

1° Les traitements de données à caractère personnel mis en œuvre pour le compte de l'État, d'une personne morale de droit public ou d'une personne morale de droit privé gérant un service public, qui portent sur des données parmi lesquelles figure le numéro d'inscription des personnes au répertoire national d'identification des personnes physiques ;

2° Les traitements de données à caractère personnel mis en œuvre pour le compte de l'État qui portent sur des données biométriques nécessaires à l'authentification ou au contrôle de l'identité des personnes.

II. - Sont autorisés par arrêté ou, en cas de traitement opéré pour le compte d'un établissement public ou d'une personne morale de droit privé gérant un service public, par décision de l'organe délibérant chargé de leur organisation, pris après avis motivé et publié de la Commission nationale de l'informatique et des libertés :

1° Les traitements mis en œuvre par l'État ou les personnes morales mentionnées au I qui requièrent une consultation du répertoire national d'identification des personnes physiques sans inclure le numéro d'inscription à ce répertoire ;

2° Ceux des traitements mentionnés au I :

- qui ne comportent aucune des données mentionnées au I de l'article 8 ou à l'article 9 ;

- qui ne donnent pas lieu à une interconnexion entre des traitements ou fichiers correspondant à des intérêts publics différents ;

- et qui sont mis en œuvre par des services ayant pour mission, soit de déterminer les conditions d'ouverture ou l'étendue d'un droit des administrés, soit d'établir l'assiette, de contrôler ou de recouvrer des impositions ou taxes de toute nature, soit d'établir des statistiques ;

3° Les traitements relatifs au recensement de la population, en métropole et dans les collectivités situées outre-mer ;

4° Les traitements mis en œuvre par l'État ou les personnes morales mentionnées au I aux fins de mettre à la disposition des usagers de l'administration un ou plusieurs téléservices de l'administration électronique, si ces traitements portent sur des données parmi lesquelles figurent le numéro d'inscription des personnes au répertoire national d'identification ou tout autre identifiant des personnes physiques.

III. - Les dispositions du IV de l'article 26 sont applicables aux traitements relevant du présent article.

Chapitre IX : Traitements de données à caractère personnel ayant pour fin la recherche dans le domaine de la santé

Article 53

Les traitements de données à caractère personnel ayant pour fin la recherche dans le domaine de la santé sont soumis aux dispositions de la présente loi, à l'exception des articles 23 à 26, 32 et 38.

Les traitements de données ayant pour fin le suivi thérapeutique ou médical individuel des patients ne sont pas soumis aux dispositions du présent chapitre. Il en va de même des traitements permettant d'effectuer des études à partir des données ainsi recueillies si ces études sont réalisées par les personnels assurant ce suivi et destinées à leur usage exclusif.

Article 54

Pour chaque demande de mise en œuvre d'un traitement de données à caractère personnel, un comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé, institué auprès du ministre chargé de la recherche et composé de personnes compétentes en matière de recherche dans le domaine de la santé, d'épidémiologie, de génétique et de biostatistique, émet un avis sur la méthodologie de la recherche au regard des dispositions de la présente loi, la nécessité du recours à des données à caractère personnel et la pertinence de celles-ci par rapport à l'objectif de la recherche, préalablement à la saisine de la Commission nationale de l'informatique et des libertés.

Le comité consultatif dispose d'un mois pour transmettre son avis au demandeur. A défaut, l'avis est réputé favorable. En cas d'urgence, ce délai peut être ramené à quinze jours.

Le président du comité consultatif peut mettre en œuvre une procédure simplifiée.

La mise en œuvre du traitement de données est ensuite soumise à l'autorisation de la Commission nationale de l'informatique et des libertés, qui se prononce dans les conditions prévues à l'article 25.

Pour les catégories les plus usuelles de traitements automatisés ayant pour finalité la recherche dans le domaine de la santé et portant sur des données ne permettant pas une identification directe des personnes concernées, la commission peut homologuer et publier des méthodologies de référence, établies en concertation avec le comité consultatif ainsi qu'avec les organismes publics et privés représentatifs, et destinées à simplifier la procédure prévue aux quatre premiers alinéas du présent article.

Ces méthodologies précisent, eu égard aux caractéristiques mentionnées à l'article 30, les normes auxquelles doivent correspondre les traitements pouvant faire l'objet d'une demande d'avis et d'une demande d'autorisation simplifiées.

Pour les traitements répondant à ces normes, seul un engagement de conformité à l'une d'entre elles est envoyé à la commission. Le président de la commission peut autoriser ces traitements à l'issue d'une procédure simplifiée d'examen.

Pour les autres catégories de traitements, le comité consultatif fixe, en concertation avec la Commission nationale de l'informatique et des libertés, les conditions dans lesquelles son avis n'est pas requis.

Article 55

Nonobstant les règles relatives au secret professionnel, les membres des professions de santé peuvent transmettre les données à caractère personnel qu'ils détiennent dans le cadre d'un traitement de données autorisé en application de l'article 53.

Lorsque ces données permettent l'identification des personnes, elles doivent être codées avant leur transmission. Toutefois, il peut être dérogé à cette obligation lorsque le traitement de données est associé à des études de pharmacovigilance ou à des protocoles de recherche réalisés dans le cadre d'études coopératives nationales ou internationales ; il peut également y être dérogé si une particularité de la recherche l'exige. La demande d'autorisation comporte la justification scientifique et technique de la dérogation et l'indication de la période nécessaire à la recherche. A l'issue de cette période, les données sont conservées et traitées dans les conditions fixées à l'article 36.

La présentation des résultats du traitement de données ne peut en aucun cas permettre l'identification directe ou indirecte des personnes concernées.

Les données sont reçues par le responsable de la recherche désigné à cet effet par la personne physique ou morale autorisée à mettre en œuvre le traitement. Ce responsable veille à la sécurité des informations et de leur traitement, ainsi qu'au respect de la finalité de celui-ci.

Les personnes appelées à mettre en œuvre le traitement de données ainsi que celles qui ont accès aux données sur lesquelles il porte sont astreintes au secret professionnel sous les peines prévues à l'article 226-13 du code pénal.

Article 56

Toute personne a le droit de s'opposer à ce que des données à caractère personnel la concernant fassent l'objet de la levée du secret professionnel rendue nécessaire par un traitement de la nature de ceux qui sont visés à l'article 53.

Dans le cas où la recherche nécessite le recueil de prélèvements biologiques identifiants, le consentement éclairé et exprès des personnes concernées doit être obtenu préalablement à la mise en œuvre du traitement de données.

Les informations concernant les personnes décédées, y compris celles qui figurent sur les certificats des causes de décès, peuvent faire l'objet d'un traitement de données, sauf si l'intéressé a, de son vivant, exprimé son refus par écrit.

Article 57

Les personnes auprès desquelles sont recueillies des données à caractère personnel ou à propos desquelles de telles données sont transmises sont, avant le début du traitement de ces données, individuellement informées :

1° De la nature des informations transmises ;

2° De la finalité du traitement de données ;

3° Des personnes physiques ou morales destinataires des données ;

4° Du droit d'accès et de rectification institué aux articles 39 et 40 ;

5° Du droit d'opposition institué aux premier et troisième alinéas de l'article 56 ou, dans le cas prévu au deuxième alinéa de cet article, de l'obligation de recueillir leur consentement.

Toutefois, ces informations peuvent ne pas être délivrées si, pour des raisons légitimes que le médecin traitant apprécie en conscience, le malade est laissé dans l'ignorance d'un diagnostic ou d'un pronostic grave.

Dans le cas où les données ont été initialement recueillies pour un autre objet que le traitement, il peut être dérogé à l'obligation d'information individuelle lorsque celle-ci se heurte à la difficulté de retrouver les personnes concernées. Les dérogations à l'obligation d'informer les personnes de l'utilisation de données les concernant à des fins de recherche sont mentionnées dans le dossier de demande d'autorisation transmis à la Commission nationale de l'informatique et des libertés, qui statue sur ce point.

Article 58

Sont destinataires de l'information et exercent les droits prévus aux articles 56 et 57 les titulaires de l'autorité parentale, pour les mineurs, ou le représentant légal pour les personnes faisant l'objet d'une mesure de tutelle.

Article 59

Une information relative aux dispositions du présent chapitre doit être assurée dans tout établissement ou centre où s'exercent des activités de prévention, de diagnostic et de soins donnant lieu à la transmission de données à caractère personnel en vue d'un traitement visé à l'article 53.

Article 60

La mise en œuvre d'un traitement de données en violation des conditions prévues par le présent chapitre entraîne le retrait temporaire ou définitif, par la Commission nationale de l'informatique et des libertés, de l'autorisation délivrée en application des dispositions de l'article 54.

Il en est de même en cas de refus de se soumettre aux vérifications prévues par le f du 2° de l'article 11.

Article 61

La transmission vers un État n'appartenant pas à la Communauté européenne de données à caractère personnel non codées faisant l'objet d'un traitement ayant pour fin la recherche dans le domaine de la santé n'est autorisée, dans les conditions prévues à l'article 54, que sous réserve du respect des règles énoncées au chapitre XII.

Chapitre X : Traitements de données de santé à caractère personnel à des fins d'évaluation ou d'analyse des pratiques ou des activités de soins et de prévention

Article 62

Les traitements de données de santé à caractère personnel qui ont pour fin l'évaluation des pratiques de soins et de prévention sont autorisés dans les conditions prévues au présent chapitre.

Les dispositions du présent chapitre ne s'appliquent ni aux traitements de données à caractère personnel effectuées à des fins de remboursement ou de contrôle par les organismes chargés de la gestion d'un régime de base d'assurance maladie, ni aux traitements effectués au sein des établissements de santé par les médecins responsables de l'information médicale dans les conditions prévues au deuxième alinéa de l'article L. 6113-7 du code de la santé publique.

Article 63

Les données issues des systèmes d'information visés à l'article L. 6113-7 du code de la santé publique, celles issues des dossiers médicaux détenus dans le cadre de l'exercice libéral des professions de santé, ainsi que celles issues des systèmes d'information des caisses d'assurance maladie, ne peuvent être communiquées à des fins statistiques d'évaluation ou d'analyse des pratiques et des activités de soins et de prévention que sous la forme de statistiques agrégées ou de données par patient constituées de telle sorte que les personnes concernées ne puissent être identifiées.

Il ne peut être dérogé aux dispositions de l'alinéa précédent que sur autorisation de la Commission nationale de l'informatique et des libertés dans les conditions prévues aux articles 64 à 66. Dans ce cas, les données utilisées ne comportent ni le nom, ni le prénom des personnes, ni leur numéro d'inscription au Répertoire national d'identification des personnes physiques.

Article 64

Pour chaque demande, la commission vérifie les garanties présentées par le demandeur pour l'application des présentes dispositions et, le cas échéant, la conformité de sa demande à ses missions ou à son objet social. Elle s'assure de la nécessité de recourir à des données à caractère personnel et de la pertinence du traitement au regard de sa finalité déclarée d'évaluation ou d'analyse des pratiques ou des activités de soins et de prévention. Elle vérifie que les données à caractère personnel dont le traitement est envisagé ne comportent ni le nom, ni le prénom des personnes concernées, ni leur numéro d'inscription au Répertoire national d'identification des personnes physiques. En outre, si le demandeur n'apporte pas d'éléments suffisants pour attester la nécessité de disposer de certaines informations parmi l'ensemble des données à caractère personnel

dont le traitement est envisagé, la commission peut interdire la communication de ces informations par l'organisme qui les détient et n'autoriser le traitement que des données ainsi réduites.

La commission détermine la durée de conservation des données nécessaires au traitement et apprécie les dispositions prises pour assurer leur sécurité et la garantie des secrets protégés par la loi.

Article 65

La commission dispose, à compter de sa saisine par le demandeur, d'un délai de deux mois, renouvelable une seule fois, pour se prononcer. A défaut de décision dans ce délai, ce silence vaut décision de rejet.

Les traitements répondant à une même finalité portant sur des catégories de données identiques et ayant des destinataires ou des catégories de destinataires identiques peuvent faire l'objet d'une décision unique de la commission.

Article 66

Les traitements autorisés conformément aux articles 64 et 65 ne peuvent servir à des fins de recherche ou d'identification des personnes. Les personnes appelées à mettre en œuvre ces traitements, ainsi que celles qui ont accès aux données faisant l'objet de ces traitements ou aux résultats de ceux-ci lorsqu'ils permettent indirectement d'identifier les personnes concernées, sont astreintes au secret professionnel sous les peines prévues à l'article 226-13 du code pénal.

Les résultats de ces traitements ne peuvent faire l'objet d'une communication, d'une publication ou d'une diffusion que si l'identification des personnes sur l'état desquelles ces données ont été recueillies est impossible.

Annexe 9 : Extraits du Code de la sécurité sociale

Article L. 161-28-1 créant le SNIIRAM

Il est créé un système national d'information interrégimes de l'assurance maladie qui contribue :

1° A la connaissance des dépenses de l'ensemble des régimes d'assurance maladie par circonscription géographique, par nature de dépenses, par catégorie de professionnels responsables de ces dépenses et par professionnel ou établissement ;

2° A la transmission en retour aux prestataires de soins d'informations pertinentes relatives à leur activité et leurs recettes, et s'il y a lieu à leurs prescriptions ;

3° A la définition, à la mise en œuvre et à l'évaluation de politiques de santé publique.

Le système national d'information interrégimes est mis en place par les organismes gérant un régime de base d'assurance maladie. Ces derniers transmettent au système national d'information interrégimes de l'assurance maladie les données nécessaires.

Les modalités de gestion et de renseignement du système national d'information interrégimes de l'assurance maladie, définies conjointement par protocole passé entre au moins la Caisse nationale de l'assurance maladie des travailleurs salariés, la Caisse centrale de mutualité sociale agricole et la Caisse nationale du régime social des indépendants, sont approuvées par un arrêté du ministre chargé de la sécurité sociale.

Cet arrêté est pris après avis motivé de la Commission nationale de l'informatique et des libertés.

Les données reçues et traitées par le système national d'information interrégimes de l'assurance maladie préservent l'anonymat des personnes ayant bénéficié des prestations de soins.

Article L. 161-36-5 créant l'IDS

Un groupement d'intérêt public dénommé " Institut des données de santé ", régi par le chapitre II de la loi n° 2011-525 du 17 mai 2011 de simplification et d'amélioration de la qualité du droit, est constitué notamment entre l'État, les caisses nationales d'assurance maladie, l'Union nationale des organismes d'assurance maladie complémentaire et l'Union nationale des professionnels de santé, mentionnées au chapitre II bis du titre VIII du livre Ier.

L'Institut des données de santé a pour mission d'assurer la cohérence et de veiller à la qualité des systèmes d'information utilisés pour la gestion du risque maladie et de veiller à la mise à disposition de ses membres, de la Haute Autorité de santé, des unions régionales des professionnels de santé ainsi que d'organismes désignés par décret en Conseil d'État, à des fins de gestion du risque maladie ou pour des préoccupations de santé publique, des données issues des systèmes d'information de ses membres, dans des conditions garantissant l'anonymat fixées par décret en Conseil d'État pris après avis de la Commission nationale de l'informatique et des libertés.

Chaque année, l'Institut des données de santé transmet son rapport d'activité au Parlement.

Article R. 115-2 relatif à l'utilisation du NIR par les organismes de Sécurité Sociale

L'autorisation donnée à l'article R. 115-1 vaut exclusivement pour les traitements, mis en œuvre dans le respect des dispositions de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés :

1° Que les organismes visés aux 1° et 2° dudit article effectuent dans l'exercice de leurs missions de sécurité sociale conformément aux lois et règlements en vigueur ou, le cas échéant, lors de la phase administrative initiale de traitement de la déclaration préalable à l'embauche prévue par l'article L. 320 du code du travail, à seule fin de vérifier l'identité du salarié faisant l'objet de cette déclaration ;

2° Que les organismes visés au 3° du même article effectuent dans l'exercice de leurs activités d'assurance maladie, maternité et invalidité complémentaire conformément à la loi n° 89-1009 du 31 décembre 1989 renforçant les garanties offertes aux personnes assurées contre certains risques, ou dans leurs activités d'assurance vieillesse complémentaire ;

3° Que les professionnels, institutions ou établissements visés au 4° du même article effectuent pour leurs échanges avec les organismes mentionnés aux 1° et 3° dudit article ;

4° Que les comptables publics visés au 4° du même article effectuent pour le recouvrement de créances auprès des assurés sociaux soignés par leurs établissements ;

5° Que les collectivités publiques qui servent des prestations d'aide sociale mettent en œuvre dans le cadre des dispositions du chapitre II du titre VIII du livre Ier du présent code, relatives aux bénéficiaires de l'aide médicale ;

6° Que les organismes mentionnés aux 1°, 3° et 6° du même article effectuent en application du I de l'article L. 133-5 pour les déclarations sociales auxquelles sont assujetties les entreprises.

Annexe 10 : Extraits du Code de la santé publique

Article L. 1435-6 (Loi n°2009-879 du 21 juillet 2009 - art. 118) : accès des ARS aux SI de santé

L'agence régionale de santé a accès aux données nécessaires à l'exercice de ses missions contenues dans les systèmes d'information des établissements de santé et des établissements et services médico-sociaux ainsi que des organismes d'assurance maladie et de la Caisse nationale de solidarité pour l'autonomie, notamment à ceux mentionnés aux articles L. 161-28-1 du code de la sécurité sociale et L. 247-2 du code de l'action sociale et des familles. Cet accès est assuré dans des conditions garantissant l'anonymat des personnes bénéficiant de prestations de soins ou de prises en charge et d'accompagnements médico-sociaux dans le respect des dispositions de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. L'agence régionale de santé est tenue informée par les organismes situés dans son ressort de tout projet concernant l'organisation et le fonctionnement de leurs systèmes d'information. Le directeur général détermine, en fonction de la situation sanitaire, pour chaque établissement, les données utiles que celui-ci doit transmettre de façon régulière, et notamment les disponibilités en lits et places. Le directeur général décide également de la fréquence de mise à jour et de transmission des données issues des établissements de soins et médico-sociaux.

Les agents de l'agence régionale de santé ayant la qualité de médecin n'ont accès aux données de santé à caractère personnel que si elles sont strictement nécessaires à l'accomplissement de leurs missions, dans le respect de l'article 226-13 du code pénal.

Avant le 1er janvier 2011, la Commission nationale de l'informatique et des libertés présente au Parlement un rapport évaluant les conditions d'accès aux données de santé par les agences régionales de santé. [...]

Article L. 5121-28 créant le GIP « études en santé publique » (loi du 29/12/2011)

Lorsque la réalisation d'études de vigilance et d'épidémiologie impliquant notamment les produits mentionnés à l'article L. 5311-1 rend nécessaire un accès au système national d'information inter-régimes de l'assurance maladie mentionné à l'article L. 161-28-1 du code de la sécurité sociale ou une extraction de ses données, l'accès ou l'extraction peuvent être autorisés par un groupement d'intérêt public constitué à cette fin entre l'État, la Haute Autorité de santé, l'Agence nationale de sécurité du médicament et des produits de santé, l'Institut de veille sanitaire et la Caisse nationale d'assurance maladie des travailleurs salariés. Ce groupement d'intérêt public est régi par la loi n° 2011-525 du 17 mai 2011 de simplification et d'amélioration de la qualité du droit.

L'autorisation est accordée en fonction des finalités poursuivies par ces études et de la contribution qu'elles sont susceptibles d'apporter, par leur qualité scientifique et dans le respect de la protection des données personnelles, notamment du secret médical, à la santé publique ou à l'efficacité des dépenses d'assurance maladie.

Le groupement d'intérêt public mentionné au premier alinéa peut conduire lui-même des études de vigilance et d'épidémiologie impliquant notamment les produits mentionnés à l'article L. 5311-1. Il peut aussi lancer des appels d'offres pour la réalisation d'études, sous réserve que celles-ci ne fassent pas l'objet d'un financement par une ou plusieurs entreprises produisant ou commercialisant des produits mentionnés au même article L. 5311-1 ou assurant des prestations associées à ces produits.

Un rapport d'activité décrivant le résultat des études menées et formulant des recommandations est remis chaque année au Parlement.

Un décret en Conseil d'État, pris après avis de la Commission nationale de l'informatique et des libertés, fixe les modalités d'application du présent article.

Annexe 11 : « Arrêté SNIIRAM » du 19 juillet 2013

Arrêté du 19 juillet 2013 relatif à la mise en œuvre du Système national d'information interrégimes de l'assurance maladie

NOR: AFSS1318985A

Version consolidée au 15 septembre 2013

La ministre des affaires sociales et de la santé,

Vu le code de la santé publique, notamment ses articles L. 4113-7, L. 6113-8 et L. 6113-9 ;

Vu le code de la sécurité sociale, notamment ses articles L. 115-4, L. 161-28, L. 161-28-1 et L. 161-29 ;

Vu la loi n° 51-711 du 7 juin 1951 modifiée sur l'obligation, la coordination et le secret en matière statistique ;

Vu la loi n° 78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés ;

Vu la loi n° 93-8 du 4 janvier 1993 relative aux relations entre les professions de santé et l'assurance maladie ;

Vu la loi n° 2004-810 du 13 août 2004 relative à l'assurance maladie ;

Vu l'avis de la Commission nationale de l'informatique et des libertés en date du 27 juin 2013,

Arrête :

Article 1

Est approuvé le protocole du 8 juin 2012 et ses annexes définissant les modalités de gestion et de renseignement du Système national d'information interrégimes de l'assurance maladie (SNIIRAM), signé par la Caisse nationale de l'assurance maladie des travailleurs salariés, la Caisse centrale de mutualité sociale agricole et la Caisse nationale du régime social des indépendants.

Article 2

Les traitements mis en œuvre dans le cadre du Système national d'information interrégimes de l'assurance maladie ont pour finalités :

1° D'améliorer la qualité des soins, notamment par :

— la comparaison des pratiques aux référentiels, accords de bons usages ou contrats de bonne pratique, au sens des articles L. 162-12-15, L. 162-12-17, L. 162-12-18, et L. 162-12-20 du code de la sécurité sociale, et moyennes professionnels ;

— l'évaluation des comportements de consommation de soins ;

— l'analyse des caractéristiques et des déterminants de la qualité des soins ;

2° De contribuer à une meilleure gestion de l'assurance maladie, notamment par :

— la connaissance des dépenses de l'ensemble des régimes d'assurance maladie par circonscription géographique, par nature de dépense, par catégorie de professionnels responsables de ces dépenses et par professionnel ou établissement ;

— l'évaluation des transferts entre enveloppes correspondant aux objectifs sectoriels de dépenses fixés, en fonction de l'objectif national de dépenses d'assurance maladie, dans le cadre de la loi de financement de la sécurité sociale ;

— l'analyse quantitative des déterminants de l'offre de soins et la mesure de leurs impacts sur l'évolution des dépenses d'assurance maladie ;

3° De contribuer à une meilleure gestion des politiques de santé, notamment par :

- l'identification des parcours de soins des patients ;
- le suivi et l'évaluation de l'état de santé des patients et leurs conséquences sur la consommation de soins ;
- l'analyse de la couverture sociale des patients ;
- la surveillance de la consommation de soins en fonction de différents indicateurs de santé publique ou de risque ;

4° De transmettre aux prestataires de soins les informations pertinentes relatives à leur activité, à leurs recettes et, s'il y a lieu, à leurs prescriptions.

Article 3

Les informations nécessaires à l'établissement du Système national d'information interrégimes de l'assurance maladie (SNIIRAM) sont transmises par les organismes gérant un régime de base d'assurance maladie et, pour ce qui concerne les informations relatives à l'activité hospitalière, par l'Agence technique de l'information sur l'hospitalisation (ATIH).

Ces informations, détaillées à l'annexe 1 du protocole, concernent :

- l'identification des organismes de prise en charge ;
- les caractéristiques des décomptes de remboursement ;
- les numéros d'anonymat de l'assuré et du bénéficiaire, le sexe, l'année et le mois de naissance, le cas échéant, la date de décès, le département et la commune de résidence ;
- les informations relatives aux prestations servies ;
- la nature détaillée des actes, biens et services présentés au remboursement ;
- les dates de soins et de remboursement ;
- le mode de prise en charge ;
- les informations relatives au parcours de soins ;
- les informations médico-administratives (notamment le numéro d'affection de longue durée au sens de l'article D. 322-1 du code de la sécurité sociale, le numéro de maladie professionnelle, les codes de pathologie suivant la codification internationale des maladies en vigueur, les dates de grossesse, le numéro de dent) ;
- les montant, cotation et coefficient de la prestation ;
- le numéro d'identification du professionnel et, le cas échéant, de l'établissement de rattachement, le sexe, la date de naissance, la spécialité médicale, la nature d'exercice, le statut conventionnel, la caisse de rattachement, département, commune et localisation infracommunale d'implantation ;
- les informations relatives à l'activité des établissements de santé : résumés de sorties anonymes établis dans le cadre du programme de médicalisation des systèmes d'information au sens de l'article L. 6113-7 du code de la santé publique et les informations de séjour pour les établissements financés par dotation globale. La durée de conservation de ces données est de dix ans au-delà de l'année en cours. Passé ce délai, ces données sont archivées pour une durée de dix ans ;
- les données comptables.

Les données individuelles concernant les bénéficiaires sont conservées pendant trois ans au-delà de l'année en cours, excepté lors de constitution d'échantillons. Passé ce délai, ces données sont archivées pour une durée de dix ans.

Un échantillon généraliste de ces données représentatif des personnes protégées des régimes est constitué afin d'assurer le suivi de la consommation de soins et des taux de recours aux soins.

La durée de conservation est de vingt ans au-delà de l'année en cours.

D'autres échantillons de ces données peuvent être réalisés, conformément aux dispositions du chapitre X de la loi du 6 janvier 1978 susvisée. Ils sont soumis à l'approbation préalable du bureau de l'Institut des données de santé, excepté lorsqu'ils sont réalisés pour le compte d'organismes d'assurance maladie obligatoires ou lorsqu'ils ont déjà fait l'objet d'un avis du Conseil national de l'information statistique.

Article 4

I. — Pour tout traitement de données individuelles relatives aux bénéficiaires de l'assurance maladie, seuls sont autorisés à effectuer des recherches mettant en œuvre simultanément plus d'une des quatre variables sensibles (code commune, date des soins, mois et année de naissance, date de décès) avec d'autres données :

- les médecins-conseils de l'assurance maladie et les personnels habilités placés sous leur responsabilité, nommément désignés par les médecins responsables selon l'organisation des régimes d'assurance maladie ;

- à titre expérimental et pour une durée de trois ans, les médecins de l'Institut de veille sanitaire et les personnels habilités placés sous leur responsabilité, nommément désignés par le directeur de l'Institut de veille sanitaire, autorité médicale de l'institut, dans la mesure où ces variables sont strictement nécessaires à l'accomplissement des missions de surveillance et d'alerte sanitaires.

II. — Dans le cadre de la charte d'engagements pour la mise à disposition et les principes d'utilisation des données issues du SNIIRAM dans les ARS entre les membres de l'UNCAM, le ministère des affaires sociales et de la santé et l'Union nationale des professionnels de santé, pour tout traitement de données individuelles relatives :

- aux bénéficiaires de l'assurance maladie, seuls les médecins membres du personnel des agences régionales de santé (ARS), nommément désignés par les directeurs généraux des ARS, sont autorisés à effectuer des recherches sur le champ de leur compétence régionale ;

- aux professionnels de santé, seuls les médecins membres du personnel des agences régionales de santé (ARS), nommément désignés par les directeurs généraux des ARS, sont autorisés, sur le champ de leur compétence régionale, à effectuer des recherches localisées (code commune et données infracommunales).

III. — Les destinataires des informations contenues dans le Système national d'information interrégimes de l'assurance maladie sont, à raison de leurs fonctions et selon les règles d'habilitation détaillées à l'annexe 2 du protocole :

1° Pour l'ensemble des informations énumérées à l'article 3, les médecins-conseils et les personnels habilités placés sous leur responsabilité ainsi que les agents administratifs habilités des caisses des différents régimes de base d'assurance maladie ou de la Caisse nationale de solidarité pour l'autonomie nommément désignés, suivant leur rattachement administratif, par les directeurs ou les agents comptables des caisses ;

2° Pour l'ensemble des informations énumérées à l'article 3, dans la mesure strictement nécessaire à l'accomplissement des missions de l'organisme et à l'exclusion de toute donnée d'identification des professionnels de santé :

- les agents habilités de l'Institut de veille sanitaire (InVS) nommément désignés par leur directeur, sont autorisés à effectuer des recherches localisées (code commune et données infra-communales) ;

- les agents habilités de la Haute Autorité de santé (HAS) nommément désignés par leur président ;

- les agents habilités de l'Agence nationale de sécurité du médicament et des produits de santé nommément désignés par leur directeur ;

- sur le champ de leur compétence régionale, les médecins des Agences régionales de santé nommément désignés par les directeurs généraux ;

— l'expérimentation d'utilisation de données identifiantes des professionnels de santé encadrée par la charte d'engagements pour la mise à disposition et les principes d'utilisation des données issues du SNIIRAM dans les ARS, doit faire l'objet, conformément aux dispositions du chapitre X de la loi du 6 janvier 1978 susvisée, d'une autorisation de la CNIL ;

3° Pour l'ensemble des informations énumérées à l'article 3, exclusivement sous forme de statistiques agrégées ou sous forme individualisée sur l'échantillon généraliste, les données relatives aux bénéficiaires de l'assurance maladie comprenant l'ensemble des données relatives aux professionnels de santé ou aux établissements de santé, à l'exclusion de toute donnée d'identification des professionnels de santé :

— les agents du ministère chargé de l'assurance maladie, de la santé, de l'action sociale des personnes handicapées et de la dépendance ou de ses services déconcentrés nommément désignés par les directeurs d'administration centrale ou des services déconcentrés de ce ministère ;

— les agents des agences sanitaires sous tutelle du ministère chargé de l'assurance maladie et de la santé nommément désignés par les directeurs de ces agences, notamment l'Agence technique de l'information sur l'hospitalisation (ATIH), l'Agence de biomédecine et les agences régionales de santé (ARS) ;

— les agents habilités et nommément désignés par le directeur général des politiques agricole, agroalimentaire et des territoires au ministère chargé de l'agriculture, pour les données relatives aux ressortissants des régimes de protection sociale agricole ;

— les agents des ministères chargés de l'économie, des finances et du budget, nommément désignés par le directeur général du Trésor ou le directeur du budget ;

— les membres du Haut Conseil pour l'avenir de l'assurance maladie (HCAAM) nommément désignés par leur président ;

— les membres de l'Institut des données de santé (IDS) nommément désignés par leur président ;

— les membres de l'Union nationale des professions de santé (UNPS) nommément désignés par leur président ;

— les chercheurs habilités et nommément désignés par le président de l'Institut de recherche et de documentation en économie de la santé (IRDES) ;

— les agents habilités et nommément désignés par le président du conseil d'administration du centre technique d'appui et de formation des centres d'examen de santé (CETAF) ;

— les chercheurs des services chargés de l'assurance maladie, de la santé, de l'action sociale, des personnes handicapées ou de la dépendance, habilités et nommément désignés par le directeur général de l'Institut national de la santé et de la recherche médicale (INSERM) ;

— les chercheurs des services chargés de l'assurance maladie, de la santé, de l'action sociale, des personnes handicapées ou de la dépendance habilités et nommément désignés par le président du conseil d'administration du Centre national de la recherche scientifique (CNRS) ;

— les agents de l'Institut national du cancer (INCA) nommément désignés par leur directeur ;

— les agents du fonds CMU nommément désignés par leur directeur ;

— les membres de l'Observatoire français des drogues et de la toxicomanie (OFDT) nommément désignés par leur président ;

— les membres de l'UNOCAM signataires de la charte d'engagements entre l'assurance maladie obligatoire et les assureurs maladie complémentaires pour le partage et la mise en commun des données de remboursement agrégées ou individuelles anonymisées issues de leurs systèmes d'information, présentée en annexe 6, nommément désignés par le président de l'Institut des données de santé.

Le traitement des informations énumérées à l'article 3 demandé par tout autre organisme de recherche, des universités, écoles ou autres structures d'enseignement liés à la recherche que ceux mentionnés au paragraphe précédent est soumis à l'approbation du bureau de l'Institut des données de santé. Aucun organisme de recherche, université, école ou autre structure d'enseignement lié à la

recherche poursuivant un but lucratif ne peut accéder aux informations de l'article 3. La CNIL, conformément aux dispositions du chapitre X de la loi du 6 janvier 1978 susvisée, autorise ces traitements.

4° Pour les données relatives à leur activité, leurs recettes ou leurs prescriptions, l'ensemble des prestataires de soins ; les professions médicales et les pharmaciens ayant seuls accès aux informations médico-administratives (numéro d'affection de longue durée au sens de l'article D. 322-1 du code de la sécurité sociale, numéro de maladie professionnelle, codes de pathologie suivant la codification internationale des maladies en vigueur) ;

5° Pour l'ensemble des informations énumérées à l'article 3, à l'exclusion de l'échantillon généraliste, sous forme de statistiques agrégées, les données relatives aux bénéficiaires de l'assurance maladie comprenant l'ensemble des données relatives aux professionnels de santé de leur région, à l'exclusion de toute donnée d'identification :

— les membres des unions régionales des professionnels de santé (URPS) nommément désignés par leur président ;

6° Pour l'ensemble des informations énumérées à l'article 3, à l'exclusion de l'échantillon généraliste, sous forme de statistiques agrégées, les données relatives aux bénéficiaires de l'assurance maladie comprenant l'ensemble des données relatives aux professionnels de santé ou aux établissements de santé à l'exclusion de toute donnée d'identification des professionnels de santé :

— les structures adhérant aux membres de l'Institut des données de santé ou les constituant, à l'exclusion des membres de l'UNOCAM, nommément désignées par le président de l'IDS.

7° Pour l'ensemble des informations énumérées à l'article 3, à l'exclusion de l'échantillon généraliste, sous forme de statistiques agrégées, les données relatives aux bénéficiaires de l'assurance maladie comprenant l'ensemble des données relatives aux professionnels de santé ou aux établissements de santé à l'exclusion de toute donnée d'identification :

— les organismes complémentaires contributeurs de données membres des fédérations constitutives de l'UNOCAM signataires de la charte d'engagement entre l'assurance maladie obligatoire et les assureurs maladie complémentaires pour le partage et la mise en commun des données de remboursement agrégées ou individuelles anonymisées issues de leurs systèmes d'information, présentée en annexe 6.

Article 5

Pour la mise en œuvre du système, il est créé une base de données nationale dont la gestion technique est confiée à la Caisse nationale de l'assurance maladie des travailleurs salariés.

Afin de garantir l'anonymat des personnes ayant bénéficié des prestations de soins, les données transmises ne comportent pas l'identité de ces personnes. Un numéro d'anonymat est établi par codage informatique irréversible à partir du numéro d'inscription au répertoire national d'identification des personnes physiques.

Ce procédé d'anonymisation s'opère à un double niveau, une première fois avant transmission des informations par les régimes à la base nationale et, une deuxième fois, préalablement à leur enregistrement dans la base de données nationale. Ce même procédé est appliqué aux numéros d'identification des titulaires de pensions d'invalidité et de rentes d'accidents du travail ou de maladies professionnelles ainsi qu'aux numéros d'entrée des patients. Toutes les données sensibles sont chiffrées lors de leur sauvegarde.

Lors des transferts vers la base nationale, les informations sont transmises selon un mode sécurisé

recourant à des procédures de chiffrement et d'authentification.

Un annuaire sécurisé, créé et mis à jour par une infrastructure de gestion de clés, des personnes habilitées à accéder au Système national d'information interrégimes de l'assurance maladie, dont la gestion technique est confiée à la Caisse nationale de l'assurance maladie des travailleurs salariés, est établi pour contrôler toutes les autorisations d'accès.

Des dispositifs de journalisation des interrogations sont mis en œuvre pour contrôler les accès au système. Des procédures d'exploitation systématique de ces journaux sont élaborées et mises en œuvre de manière à permettre l'audit des accès au système.

Article 6

Le droit d'accès et de rectification des données à caractère personnel prévu par les articles 39 et 40 s'exerce, pour les professionnels de santé, auprès de la caisse dont ils relèvent au titre de leur activité principale.

Article 7

Conformément aux dispositions du dernier alinéa de l'article 38 de la loi du 6 janvier 1978 susvisée, le droit d'opposition ne s'applique pas au présent traitement.

Article 8

Le protocole et ses annexes seront publiés au Bulletin officiel solidarité-santé du ministère des affaires sociales et de la santé.

Article 9

A modifié les dispositions suivantes :

- Abroge Arrêté du 11 juillet 2012 (Ab)
- Abroge Arrêté du 11 juillet 2012 - art. 1 (Ab)
- Abroge Arrêté du 11 juillet 2012 - art. 10 (Ab)
- Abroge Arrêté du 11 juillet 2012 - art. 2 (Ab)
- Abroge Arrêté du 11 juillet 2012 - art. 3 (Ab)
- Abroge Arrêté du 11 juillet 2012 - art. 4 (Ab)
- Abroge Arrêté du 11 juillet 2012 - art. 5 (Ab)
- Abroge Arrêté du 11 juillet 2012 - art. 6 (Ab)
- Abroge Arrêté du 11 juillet 2012 - art. 7 (Ab)
- Abroge Arrêté du 11 juillet 2012 - art. 8 (Ab)

Article 10

Le directeur de la sécurité sociale est chargé de l'exécution du présent arrêté, qui sera publié au Journal officiel de la République française.

Fait le 19 juillet 2013.

Pour la ministre et par délégation :
Le directeur de la sécurité sociale,
T. Fatome

