



Commission open data en santé

RAPPORT

**Remis à Madame Marisol TOURAINE,
Ministre des Affaires sociales et de la Santé,**

Le 9 juillet 2014

Sommaire

Introduction	p.5
Partie 1 : Etat des lieux	p.9
1.1 Définitions	
1.2 L'état du droit : le cadre juridique de la protection et de l'utilisation des données personnelles de santé	
1.3 La gouvernance	
1.4 Cartographie des accès au SNIIRAM et au PMSI	
1.5 Cartographie des données de santé	
1.6 Enseignements du panorama international	
Partie 2 : Enjeux liés à l'ouverture des données de santé	p.37
2.1 Renforcer la démocratie sanitaire	
2.2 Renforcer l'autonomisation du patient	
2.3 Accroître l'efficacité de l'action publique	
2.4 Permettre aux professionnels de santé de renforcer la connaissance de leur exercice et l'amélioration de sa qualité	
2.5 Développer la recherche	
2.6 Encourager l'innovation et le développement industriel	
Partie 3 : Attentes et éléments de doctrine	p.43
3.1 L'accès aux données anonymes ou rendues anonymes	
3.1.1 Doctrine de l'open data en santé	
3.1.2 Attentes sur les données anonymes à ouvrir	
3.1.3 Recommandations pour pérenniser le mouvement de publication en open data des données en santé	
3.1.4 Actions en cours et à poursuivre pour permettre l'ouverture en open data de données supplémentaires à partir du SNIIRAM-PMSI	
3.2 L'enrichissement à poursuivre du SNIIRAM et du PMSI	
3.3 Les données des patients porteuses de risques de réidentification	
3.4 Les données relatives aux professionnels et établissements de santé	
3.4.1 Les professionnels de santé libéraux	
3.4.2 Les établissements de santé	
3.4.3 Les professionnels de santé exerçant au sein des établissements de santé	
3.5 Un niveau de gouvernance qui assure le fonctionnement du dispositif	
3.5.1 La décision stratégique et la réglementation	
3.5.2 L'expression des parties prenantes	
3.5.3 La décision de donner ou non accès aux données	
3.5.4 Un niveau opérationnel de mise à disposition des données	
3.5.5 Le renforcement de l'effectivité des sanctions	
Conclusion	p.61
Annexes¹	

¹ Les annexes, comme le rapport lui-même, sont disponibles sur internet à l'adresse <http://www.drees.sante.gouv.fr/rapport-de-la-commission-open-data-en-sante,11323.html>

Introduction

Contexte et enjeux

La commission dite « open data en santé », qui s'est réunie de novembre 2013 à mai 2014, avait pour mission de débattre, dans un cadre pluraliste associant les parties prenantes, des enjeux et des propositions en matière d'accès aux données de santé.

Ses travaux se sont déroulés dans un double contexte :

- une stratégie gouvernementale volontariste en faveur de *l'open data*, c'est-à-dire l'ouverture, le partage et la réutilisation des données publiques, portée par Etalab, service du Premier Ministre ;
- une réflexion approfondie menée par le Ministère de la santé, en vue d'ouvrir davantage l'accès aux données de santé, et notamment aux données du SNIIRAM² et du PMSI³, dans le respect de l'anonymat des patients ; cette réflexion a fait l'objet du rapport de Pierre-Louis Bras, remis à Marisol Touraine, Ministre des affaires sociales et de la santé, le 3 octobre 2013.

Les enjeux de l'accès aux données de santé sont nombreux. Ils touchent à la démocratie sanitaire, à l'autonomisation des patients, à l'efficacité de l'action publique, aux progrès de la recherche et à la capacité de notre pays à favoriser l'innovation et à rester compétitif en matière d'accueil d'entreprises du champ de la santé.

Si la situation actuelle n'est pas toujours satisfaisante, de nombreuses avancées ont néanmoins déjà été réalisées.

D'une part, un grand nombre de données sont déjà accessibles au public : la cartographie élaborée pour la commission et présentée dans la première partie en témoigne.

D'autre part, les accès au SNIIRAM-PMSI et les extractions qui en sont faites ont connu de grands progrès dans les années récentes, notamment grâce aux actions déployées par la CNAMTS et l'ATIH⁴, ainsi que grâce à l'impulsion et aux décisions prises par l'Institut des données de santé.

Le panorama international réalisé pour la commission, et présenté dans ce rapport, montre que d'autres pays ont engagé des stratégies volontaristes en matière d'ouverture des données de santé. La France, si elle n'a pas à rougir de sa situation, pourrait ainsi se doter d'objectifs ambitieux de moyen-terme la portant au niveau des meilleurs.

² Système national d'information interrégimes de l'assurance maladie.

³ Programme de médicalisation des systèmes d'information.

⁴ Agence technique de l'information sur l'hospitalisation.

Objectifs et méthode de travail de la commission

La Ministre des affaires sociales et de la santé a confié à Philippe Burnel, délégué à la stratégie des systèmes d'information de santé, et Franck von Lenep, directeur de la recherche, des études, de l'évaluation et des statistiques, la mission d'animer une commission dite *open data*, en lien avec la mission Etalab.

La lettre de mission des animateurs de la commission (voir annexe 2) précisait que les débats de cette dernière devaient tenir compte de la distinction entre données publiques et données indirectement nominatives. S'agissant des données anonymes, l'objectif était d'établir un état des lieux et l'identification de jeux de données publiques susceptibles d'être mis en ligne. S'agissant des données comportant un risque de réidentification, il était demandé de confronter les points de vue sur les principes présidant à l'accès à ces données et sur les évolutions possibles, dans le respect de la vie privée.

On notera d'emblée que, si le choix a été fait de baptiser cette commission « open data », le périmètre de sa réflexion n'était pas cantonné aux données en open data, mais portait sur l'ensemble des données de santé anonymes et désidentifiées, y compris donc celles qui présentent un risque de réidentification des patients, ainsi que sur les données nominatives des professionnels et établissements de santé.

Les données nécessaires aux personnels de soins dans l'exercice de leurs missions auprès des patients, et en particulier les données des dossiers médicaux, ne relevaient pas du périmètre de la commission.

La commission était composée de 43 membres (cf. liste en annexe 1) :

- Représentants des patients et usagers,
- Représentants des professionnels de santé,
- Représentants des établissements de santé,
- Représentants des industriels,
- Représentants des organismes complémentaires,
- Chercheurs,
- Producteurs de données,
- Institut des données de santé,
- Services de la CNIL,
- Juristes.

La commission s'est réunie huit fois entre novembre 2013 et mai 2014. Elle a notamment débattu des préconisations du rapport Bras, examiné les travaux d'élaboration de bases anonymes à partir du PMSI, pris connaissance des travaux menés par le Secrétariat Général pour la Modernisation de l'Action Publique (SGMAP) en matière d'open data (enseignements internationaux, cartographie de l'existant en France, préconisations de données à ouvrir).

Le présent rapport retrace les travaux et discussions de la commission :

- Un panorama de l'existant (partie 1) : définitions des concepts, état du droit, présentation de la gouvernance, présentation de l'accès aux données du SNIIRAM et du PMSI, cartographie des données de santé et enseignements tirés des expériences étrangères ;
- Les enjeux pour l'avenir (partie 2) ;
- Les actions à mener (partie 3) : données à ouvrir en open data, orientations en matière de données réidentifiantes, données relatives aux professionnels et aux établissements.

Ce rapport a été adopté consensuellement par l'ensemble des membres de la commission⁵, qui partagent des attentes communes et fortes.

Ce sont ces attentes que ce document retrace ; les orientations et préconisations présentées ici visent à répondre à ces attentes ; c'est à l'aune de leurs résultats que leur mise en œuvre devra être évaluée.

⁵ Les services de la Cnil ont participé aux travaux sans approuver formellement le rapport.

Partie 1

Etat des lieux

Il a paru utile, avant d'apporter des informations détaillées sur la situation actuelle en matière de diffusion et d'accès aux données de santé, de donner quelques définitions et références sur les principaux concepts nécessaires à la bonne compréhension des débats.

1.1 Définitions

1.1.1 Open Data

L'Open Data désigne l'ouverture et le partage de données par leur mise en ligne dans des formats ouverts, en autorisant la réutilisation libre et gratuite par toute personne.

Les données ouvertes peuvent être utilisées par une organisation publique ou privée, pour des fins différentes de celles ayant motivé la création des données.

Les modalités de mise à disposition doivent respecter le principe de la plus grande liberté de réutilisation des données publiques, en évitant le plus possible les contraintes d'ordre technique, financier, juridique ou autre.

La donnée occupe désormais une place centrale dans l'économie et la société. L'open data s'inscrit ainsi dans un mouvement plus vaste de fluidification de la donnée. Ce mouvement appelle une gouvernance de la donnée et une architecture des systèmes d'information concernant à la fois les données ayant vocation à être ouvertes à tous et celles n'ayant pas vocation à le devenir.

1.1.2 L'ouverture des données publiques

Le droit d'accès et de réutilisation des données publiques concerne les textes, mémorandums, documents, tableaux ou statistiques produits par l'administration dans le cadre d'une mission de service public.

Appliqué à la sphère publique, l'open data, ou ouverture et partage des données publiques, sert trois objectifs :

- C'est la manière, pour un Etat moderne, de s'organiser afin de rendre des comptes, d'ouvrir le dialogue, et de transformer la relation entre l'Etat et les citoyens ;
- C'est aussi - souvent - le moyen de renforcer l'efficacité de l'action publique, en simplifiant le fonctionnement des administrations ;
- C'est enfin un levier pour stimuler l'innovation et la recherche, et encourager la créativité, par exemple à travers de nouveaux services utiles aux citoyens.

Cette politique est promue par le gouvernement et est un levier important de l'action de modernisation de l'action publique. Les enjeux et modalités de mise en œuvre de cette politique ont été établis et rappelés à de nombreuses reprises, au travers notamment de la feuille de route du gouvernement français sur l'ouverture et le partage des données publiques publiée en février 2013⁶, et des Comités Interministériels pour la Modernisation de l'Action Publique (CIMAP), dont de nombreuses décisions concernent l'open data⁷. Celles-ci ont notamment conduit à réaffirmer le principe de gratuité de la réutilisation des données publiques (*Décisions 24, 25 et 26 du CIMAP du 18 décembre 2013*⁸), et à l'étendre, en concertation, aux collectivités, aux délégataires de service public, à la sphère sociale et aux autorités administratives indépendantes (Voir Annexe 3). Un *Vademecum* pour l'ouverture des données publiques, fournissant à tous les agents publics un guide pratique des modalités d'ouverture des données, a également été adressé à toutes les administrations par une circulaire du Premier ministre.

Ainsi, la politique d'ouverture des données publiques dépasse l'obligation légale de réponse de l'administration aux demandes du public. Elle organise la publication volontaire, gratuite, dans des formats faciles à utiliser par ordinateur, de ces informations, et en autorise par avance toutes les réutilisations.

Toutes les informations ou données contenues dans un document administratif communicable (selon les critères fixés par la loi CADA, voir infra 1.2.2.3) sont réutilisables, à l'exception :

- des « documents produits ou reçus par les administrations mentionnées à l'article 1er de la loi CADA dans l'exercice d'une mission de service public à caractère industriel ou commercial »,
- des documents « sur lesquels des tiers détiennent des droits de propriété intellectuelle » (article 10 de la loi CADA).

1.1.3 Les données à caractère personnel et la protection de la vie privée

On entend par 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, Loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés). Ainsi, une donnée est considérée « à caractère personnel » dès lors qu'elle concerne des personnes physiques identifiées directement ou indirectement.

Il peut s'agir du nom d'une personne, d'une photo, du numéro de sécurité sociale, de téléphone, de compte bancaire, d'une adresse e-mail, d'une empreinte digitale, de l'ADN...

La loi CADA, qui organise le devoir de redevabilité de l'Etat quant aux données qu'il détient, limite drastiquement la diffusion des données à caractère personnel par les administrations.

⁶ <http://www.etalab.gov.fr/article-la-feuille-de-route-du-gouvernement-en-matiere-d-ouverture-et-de-partage-des-donnees-publiques-115767801.html>

⁷ <http://www.modernisation.gouv.fr/le-sgmap/le-cimap>

⁸ http://www.modernisation.gouv.fr/sites/default/files/fichiers-attaches/relevedecisions_cimap4.pdf

La loi n°78-17 relative à l'informatique, aux fichiers et aux libertés du 6 janvier 1978 (plus connue sous le nom de loi Informatique et Libertés, loi de 1978 ou LIL) régit les traitements de données à caractère personnel quel que soit le procédé utilisé de leur mise en œuvre (papier ou informatique). Elle autorise la réutilisation sous certaines conditions, notamment de finalité et de proportionnalité des données recueillies pour la mise en œuvre d'un traitement (et après accord de la Cnil). Elle reconnaît aux personnes dont les données font l'objet d'un traitement un droit d'information préalable, un droit d'opposition, un droit d'accès à ces données ainsi qu'un droit de rectification de celles-ci.

1.1.4 Les données de santé

Il existe un grand nombre d'informations en rapport plus ou moins direct avec la santé. On peut distinguer d'abord les données individuelles (relatives initialement à des personnes, que celles-ci soient identifiables ou non) et les autres, par nature anonymes (qualité des eaux, taux de pollution, objectif national de dépenses d'assurance maladie, numérisation...).

Les données de santé sont mentionnées, dans la loi « Informatique et Libertés », comme des données sensibles, appelant une vigilance particulière.

On peut (à titre d'éclairage, sans prétention ici à l'exhaustivité) classer ces données par leur objet, leur forme, ou leurs sources :

- ✓ Ce que les données décrivent (leur objet) :
 - Informations statistiques sur l'état de santé d'une population, la prévalence et l'incidence des maladies, les variations selon l'âge, le sexe, les catégories professionnelles et sociales, le revenu du ménage, le niveau d'études, certains caractères génétiques, la qualité de l'air ou de l'eau, d'autres facteurs liés à l'environnement, ou selon l'indice de masse corporelle, l'alimentation, la consommation d'alcool ou de tabac, les pratiques sexuelles etc. ;
 - Informations sur l'offre de soins : établissements et professionnels de santé, classés selon la nature de l'offre (spécialités, équipements, capacités...), l'activité, la distance, la disponibilité, la réputation et d'autres indicateurs de la qualité, les tarifs... ;
 - Informations sur la dépense de santé (consommation de soins et biens médicaux, indemnités journalières, prévention environnementale et alimentaire, formation des professions de santé, recherche dans le domaine de la santé, dépenses de santé liées au handicap et à la dépendance...) ou informations sur le financement de la dépense de santé (assurance maladie obligatoire, assurances complémentaires, ménages...);
 - Informations sur les professionnels de santé (démographie, activités, revenus...);
 - Informations sur les facteurs de l'état de santé (environnement, données sociales...).

Une typologie des données de santé a été élaborée pour décrire les initiatives d'open data en santé à l'étranger et pour cartographier les données existantes en France. Cette typologie est présentée plus loin.

✓ Les différentes formes de données :

- *données brutes*, également appelées « granulaires », correspondant au niveau le plus fin qu'il est possible d'obtenir en fonction de l'origine de la donnée et du système de collecte ;
- *données extraites*, telles qu'un échantillon de données brutes ou des données rendues moins précises sur une ou plusieurs dimensions ;
- *données agrégées et statistiques* plus ou moins élaborées : totaux, indicateurs, évolutions, comparaisons...

✓ Leurs sources :

- *Données générées par les professionnels de santé* : prescriptions médicales, ventes de médicaments, ...
- *Données générées par les établissements de santé* : informations administratives des établissements de santé, données d'activité, données relatives à la qualité...
- *Données générées par les autorités dans la réalisation de leur mission de service public* : indicateurs financiers agrégés, données d'évaluation et de suivi des politiques publiques, données sur les remboursements inter-régimes... ;
- *Données générées par les industriels et entreprises du secteur de la santé* : laboratoires pharmaceutiques, assureurs, entreprises produisant du matériel médical... ;
- *Données générées par les chercheurs* : enquêtes, recherches médicales ou interdisciplinaires (ex : sociologie de la santé) ;
- *Données générées par les patients* : témoignages, informations remontées sur des événements indésirables, informations générées à l'aide d'applications et de capteurs électroniques analysant l'activité physique... ; ce champ est en pleine expansion.

Qu'appelle-t-on donnée personnelle de santé ?

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 ou une prédisposition à la maladie. 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 – voir infra le 1.2.3 pour davantage de précisions).

Ainsi le fait de consulter un psychiatre ou d'acheter des médicaments antidiabétiques et de faire réaliser certains examens biologiques est révélateur de l'état de santé d'une personne. Inversement, on peut s'interroger sur le caractère confidentiel de certaines informations de santé pourtant personnelles mais apparentes, telles que la taille ou le poids par exemple.

Qu'est-ce qu'un jeu de données réidentifiantes ?

La réflexion de la commission portait sur le champ des données « désidentifiées »⁹, c'est-à-dire ne comportant pas de données sur les patients directement identifiantes (nom, adresse, NIR).

⁹ C'est la raison pour laquelle l'accès des patients aux données médicales les concernant, garanti par la loi, n'est pas traité dans ce rapport.

Les jeux de données réidentifiantes sont des ensembles de données désidentifiées (ne comportant donc pas de nom, d'adresse, ni de NIR) mais permettant d'identifier la personne dès lors que l'on dispose par ailleurs d'informations sur cette personne qui figurent dans le jeu de données et qui, prises ensemble, lui sont propres (par exemple un parcours de soins, des dates d'hospitalisation, un code postal, un âge...). Si une seule personne, dans une base exhaustive, présente un ensemble de telles caractéristiques :

- elle peut être identifiée à coup sûr,
- mais seulement par un tiers disposant de l'information sur la personne et ayant accès aux données.

Ce point est abordé plus en détail dans la suite de ce document.

1.2 L'état du droit : le cadre juridique de la protection et de l'utilisation des données personnelles de santé

Cette partie présente d'abord les principes constitutionnels et les règles de droit international applicables en matière de traitement des données personnelles de santé, qui s'imposent à tout texte législatif ou réglementaire (1). Elle expose ensuite le cadre législatif actuel (2) avant d'analyser les évolutions contenues dans la proposition de règlement de l'Union européenne relatif à la protection des données personnelles (3). Deux annexes sont consacrées à l'hébergement des données de santé et aux dossiers personnalisés de santé.

1.2.1 Les principes constitutionnels et les règles de droit international applicables

1.2.1.1 Les principes constitutionnels

Les normes du « bloc de constitutionnalité » (Constitution, Déclaration des droits de l'homme et du citoyen et Préambule de la Constitution de 1946) ne traitent pas expressément du droit à la protection des données personnelles. Toutefois, la jurisprudence du Conseil constitutionnel a progressivement consacré ce droit, d'abord en jugeant que la liberté individuelle proclamée par l'article 2 de la Déclaration des droits de l'homme et du citoyen impliquait le droit au respect de la vie privée¹⁰, puis, de manière plus précise, en en déduisant que les traitements de données à caractère personnel devaient être justifiés par un motif d'intérêt général et mis en œuvre de manière adéquate et proportionnée à cet objectif¹¹.

La jurisprudence du Conseil constitutionnel ne fait pas apparaître aujourd'hui de déclinaison spécifique concernant les données de santé¹². Toutefois, on peut supposer que le Conseil constitutionnel exerce un contrôle particulièrement vigilant sur l'atteinte au droit à la vie privée dont le traitement des données de santé peut être porteur. En outre, il ressort des commentaires aux Cahiers du Conseil constitutionnel sur la décision du 22 mars 2012 que « *dès que n'est plus en cause un fichier de police et de justice, le contrôle du Conseil se renforce* », l'objectif de recherche des auteurs d'infraction n'étant en effet plus invocable.

Il résulte de la jurisprudence du Conseil constitutionnel¹³ et du Conseil d'Etat¹⁴ que les garanties prises en compte pour s'assurer de l'adéquation et de la proportionnalité du traitement sont notamment les suivantes :

¹⁰ Décision n° 99-416 DC du 23 juillet 1999, *Loi portant création d'une couverture maladie universelle*, cons. 45, au sujet des dispositions législatives relatives à la carte Vitale.

¹¹ Décision n° 2012-652 DC du 22 mars 2012, *Loi relative à la protection de l'identité*, cons. 7 et 8.

¹² Dans la décision n° 99-416 DC comme dans la décision n° 2010-25 QPC du 16 septembre 2010 relative au Fichier national automatisé des empreintes génétiques, le fait que les données concernées soient des données de santé ne semble pas de jouer de rôle particulier dans le raisonnement du Conseil constitutionnel. En revanche, dans la décision n° 2012-652 DC du 22 mars 2012, le Conseil constitutionnel a relevé que les données biométriques étaient « *particulièrement sensibles* » (cons. 10).

¹³ Cf. notamment les décisions n° 99-416 DC du 23 juillet 1999 (cons. 46 et 47), n° 2003-467 DC du 13 mars 2003 (cons. 17 à 46) et n° 2010-25 QPC du 16 septembre 2010 (cons. 12 à 20).

¹⁴ Cf. la décision CE Ass. 26/10/2011, *Association pour la promotion de l'image et autres*, n° 317827 et l'avis du 6 juin 2013 relatif à la création d'un registre national des crédits aux particuliers (RNCP), à paraître au rapport d'activité 2014.

- Garanties prévues de manière générale par la loi du 6 janvier 1978 (le Conseil constitutionnel tient compte du fait que la loi qui lui est soumise n'y déroge pas) : droits d'information, d'accès et de rectification, contrôle de la Cnil ;
- Précision dans la loi des finalités du fichier ;
- Délimitation précise des personnes pouvant accéder aux fichiers, le cas échéant par une procédure d'habilitation, et des données auxquelles elles ont accès (la loi n'a pas à fixer ces règles de manière précise, mais le Conseil constitutionnel prend en compte le fait qu'elle renvoie sur ce point à un décret en Conseil d'Etat pris après avis de la Cnil) ;
- Application du secret professionnel aux personnes ayant accès aux données ;
- Sanction pénale du non-respect des règles d'utilisation et de transmission des données contenues dans le traitement.

A l'inverse, les éléments qui ont conduit le Conseil constitutionnel à censurer des dispositions législatives relatives à des traitements de données sont les suivants :

- Flou sur les finalités et les conditions d'utilisation du traitement de données¹⁵ ;
- Flou sur la nature des données collectées et sur les garanties assurant leur intégrité et leur confidentialité¹⁶ ;
- Atteinte disproportionnée à la vie privée compte tenu du nombre de personnes fichées, du caractère biométrique des données et de la pluralité des finalités possibles du traitement¹⁷.

1.2.1.2 Les règles de droit international

La convention n° 108 du Conseil de l'Europe du 28 janvier 1981 pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel et la directive 95/46/CE du 24 octobre 1995 de l'Union européenne relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données énoncent des règles similaires, notamment pour les principes relatifs à la « *qualité des données* » : loyauté de la collecte des données, collecte pour des finalités déterminées, proportionnalité de la collecte au regard de ces finalités, exactitude des données, limitation dans le temps de la conservation des données.

La directive 95/46/CE énonce en outre des règles particulières relatives aux données de santé, qui font partie des données « *sensibles* » énumérées par l'article 8. Le traitement des données personnelles de santé est prohibé, sauf dans les cas énumérés par les articles 8.2 et 8.3, notamment lorsque la personne a donné son consentement explicite ou lorsque le traitement est nécessaire « *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 que le traitement de ces données est effectué par un praticien de la santé soumis par le droit national ou par des réglementations arrêtées par les autorités nationales compétentes au secret professionnel, ou par une autre personne également soumise à une obligation de secret équivalente* ». L'article 8.4 permet aux Etats de prévoir d'autres dérogations justifiées par un « *motif d'intérêt public important* ». Le considérant 34 de la directive indique à ce sujet que « *les États membres doivent également être autorisés à déroger à l'interdiction*

¹⁵ Cf. la décision n° 2004-499 DC du 29 juillet 2004, cons. 11 et 12.

¹⁶ Décision n° 2012-652 DC du 22 mars 2012, cons. 14.

¹⁷ Décision n° 2012-652 DC du 22 mars 2012, cons. 11.

de traiter des catégories de données sensibles lorsqu'un motif d'intérêt public important le justifie dans des domaines tels que la santé publique et la protection sociale — particulièrement afin d'assurer la qualité et la rentabilité en ce qui concerne les procédures utilisées pour régler les demandes de prestations et de services dans le régime d'assurance maladie — et tels que la recherche scientifique et les statistiques publiques; qu'il leur incombe, toutefois, de prévoir les garanties appropriées et spécifiques aux fins de protéger les droits fondamentaux et la vie privée des personnes ».

Il faut enfin souligner que la Charte des droits fondamentaux de l'Union européenne, qui s'impose depuis le traité de Lisbonne aux institutions européennes et aux Etats membres lorsqu'ils appliquent le droit de l'Union, consacre à son article 8 le droit à la protection des données personnelles, et que sur le fondement du droit à la vie privée reconnu par l'article 8 de la convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales, la Cour européenne des droits de l'homme a développé une jurisprudence sur ce même droit à la protection. Les principes énoncés par l'article 8 de la Charte et par la jurisprudence de la CEDH sont similaires à ceux prévus par la convention n° 108 du Conseil de l'Europe et la directive 95/46/CE, exposés ci-dessus.

1.2.2 Le cadre législatif en vigueur et ses implications pour l'accès et l'utilisation des données de santé

1.2.2.1. Le cadre général de la loi du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés

Le cadre général

La loi du 6 janvier 1978, modifiée notamment par la loi du 6 août 2004 qui a assuré la transposition de la directive 95/46/CE, reprend les dispositions de celle-ci :

- principes relatifs à la qualité des données, énoncés à l'article 6 ;
- obligation de recueillir le consentement de la personne ou de se situer dans l'un des cas alternatifs énumérés par l'article 7 (respect d'une obligation légale, sauvegarde de la vie de la personne concernée, etc.) ;
- mention des données de santé parmi les données « *sensibles* » énumérées par l'article 8, dont le traitement est en principe prohibé sauf dans les cas qu'il définit¹⁸ ;
- obligation de fournir aux personnes concernées les informations énumérées par l'article 32 (identité du responsable de traitement, finalité poursuivie, droits des personnes...);
- obligation générale de sécurité incombant au responsable du traitement définie à l'article 34.

La loi comporte en outre deux chapitres spécifiques propres aux données de santé :

¹⁸ Les exceptions pertinentes pour les données de santé sont : les traitements pour lesquels la personne a donné son consentement exprès ; les traitements nécessaires à la sauvegarde de la vie humaine ; 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 ; les traitements nécessaires à la recherche dans le domaine de la santé, selon les modalités prévues par le chapitre IX ; les traitements justifiés par l'intérêt public et autorisés par la Cnil.

- Le chapitre IX, relatif aux traitements de données nominatives ayant pour fin la recherche dans le domaine de la santé, est issu d'une loi du 1^{er} juillet 1994. La mise en œuvre d'un tel traitement est subordonnée à une autorisation de la Cnil, après consultation préalable du « *comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé* » (CCTIRS), composé d'experts en matière de recherche en santé, qui donne un avis « *sur la méthodologie de la recherche au regard des dispositions de la présente loi, la nécessité du recours à des données nominatives et la pertinence de celles-ci par rapport à l'objectif de la recherche* ».

- Le chapitre X, relatif au traitement des données personnelles de santé à des fins d'évaluation ou d'analyse des activités de soins et de prévention, a été créé à l'occasion de la loi du 27 juillet 1999 sur la CMU. Créé en vue de la mise en place du « *programme de médicalisation des systèmes d'information* » (PMSI) dans les hôpitaux, ce chapitre n'autorise le traitement de données de santé à cette fin que si celles-ci sont agrégées ou anonymisées. Seule la Cnil peut autoriser une dérogation à cette règle, auquel cas les données utilisées ne doivent comporter ni le nom, ni le prénom, ni le NIR des personnes concernées.

La notion de donnée de santé

Toutes les dispositions propres aux données de santé soulèvent la question de leur champ d'application et donc de la définition de cette notion. Tant la directive 95/46/CE que la loi du 6 janvier 1978 sont muettes à ce sujet. On ne dispose à ce jour que d'éclairages partiels donnés par les travaux préparatoires de la loi, par la jurisprudence ou par la doctrine des autorités administratives indépendantes chargées de la protection des données :

- Les travaux préparatoires de la loi du 6 août 2004 indiquent que les données génétiques sont considérées comme des données de santé.

- Le Conseil d'Etat a jugé¹⁹, dans une affaire relative au fichier « Base élèves » de l'éducation nationale, que la mention exacte de la catégorie de classe d'intégration scolaire (CLIS) d'un élève handicapé, qui « *permet (...) d'identifier immédiatement la nature de l'affection ou du handicap propre à l'élève concerné* », est une donnée de santé au sens de la loi du 6 janvier 1978. En revanche, la seule mention du fait que l'élève est scolarisé dans une structure de soins ne donne pas assez d'informations sur la nature ou la gravité de l'affection dont il souffre pour être considérée comme une donnée de santé.

- La Cour de justice de l'Union européenne²⁰ a considéré que « *l'indication du fait qu'une personne s'est blessée au pied et est en congé de maladie partiel constitue une donnée à caractère personnel relative à la santé au sens de l'article 8, paragraphe 1, de la directive 95/46* ». A cette occasion, elle a jugé « *qu'eu égard à l'objet de cette directive, il convient de donner à l'expression «données relatives à la santé» employée à son article 8, paragraphe 1, une interprétation large de sorte qu'elle comprenne des informations concernant tous les aspects, tant physiques que psychiques, de la santé d'une personne* ».

¹⁹ CE 19/7/2010, *M. Fristot et Mme Charpy*, n° 317182 et n° 334014.

²⁰ CJCE 6/11/2003, *Bodil Lindqvist*, C-101/01.

- Enfin, le G29, instance qui regroupe les autorités de protection des données homologues de la Cnil dans chaque Etat européen, a défendu dans un document de travail une approche large de la donnée de santé : une donnée en relation étroite avec l'état de santé de la personne, telle qu'une information sur la consommation d'alcool, de drogues ou de médicaments doit être considérée comme une donnée de santé au sens de la directive²¹.

1.2.2.2 Le cas particulier du SNIIRAM et du PMSI

Le cadre juridique du SNIIRAM et du PMSI

La base légale du Système national d'information interrégimes de l'assurance-maladie (SNIIRAM) a été créée par la LFSS pour 1999. L'article L. 161-28-1 du code de la sécurité sociale définit aujourd'hui les finalités du SNIIRAM (connaissance des dépenses d'assurance-maladie, transmission d'informations aux prestataires de soins sur leur activité, mise en œuvre et évaluations de politiques de santé publique) et renvoie à un protocole passé entre la CNAMTS, le RSI et la MSA, approuvé par arrêté ministériel pris après avis de la CNIL, le soin de définir « *les modalités de gestion et de renseignement* » du SNIIRAM. Le texte précise enfin que les traitements de données du SNIIRAM préservent l'anonymat des personnes ayant bénéficié de prestations de soins.

L'arrêté auquel renvoie l'article L. 161-28-1²² joue de fait un rôle majeur. Il ne se borne pas à approuver le protocole de gestion mais précise les finalités du SNIIRAM, définit les catégories de données qui y sont versées et organise les procédures d'accès aux données. L'arrêté impose la génération d'un numéro d'anonymat, en prévoyant un double chiffrement du NIR.

Quant au PMSI, son existence est prévue par l'article L. 6113-8 du code de la santé publique. L'article L. 6113-8 définit ainsi les catégories de données transmises par les établissements de santé (moyens de fonctionnement, activité, données sanitaires, démographiques et sociales) et les finalités du PMSI (élaboration du projet régional de santé, détermination des ressources des établissements, évaluation de qualité des soins, vigilance sanitaire, contrôle de la facturation). Les dispositions du chapitre X de la loi du 6 janvier 1978 sont applicables et les textes du code de la santé publique propres au PMSI les reprennent à plusieurs reprises, notamment sur la condition d'anonymat des données transmises ou, à défaut, d'effacement du nom, du prénom et du NIR.

Il faut noter que les conditions d'accès aux données du PMSI sont plus simples et moins strictes que celles du SNIIRAM. En effet, les dispositions combinées du chapitre X de la loi du 6 janvier 1978 et du code de la santé publique définissent deux régimes d'accès :

- En vertu de l'article 63 de la loi du 6 janvier 1978, l'accès à des fins statistiques d'évaluation ou d'analyse des pratiques et des activités de soins et de prévention est libre, dès lors que les données transmises sont agrégées ou ne permettent pas d'identifier les patients concernés.

- L'accès aux données identifiantes (qui ne peuvent en aucun cas comprendre le nom, le prénom et le NIR) est subordonné en vertu des articles 63 et 64 à une autorisation de la CNIL, qui

²¹ G29, "Working Document on the processing of personal data relating to health in electronic health records (EHR)", février 2007.

²² Le dernier en date est un arrêté du 19 juillet 2013, modifié par arrêté du 14 février 2014.

s'assure notamment des garanties présentées par le demandeur et de la nécessité de recourir à des données à caractère personnel au regard de la finalité d'évaluation ou d'analyse des pratiques.

Les questions posées par l'utilisation du NIR

Le numéro d'inscription au répertoire national d'identification des personnes physiques, ou NIR (couramment appelé numéro INSEE), fait l'objet d'un encadrement particulier par la loi du 6 janvier 1978. Les traitements portant sur des données parmi lesquelles figure le NIR doivent être autorisés par la Cnil (article 25). Lorsque de tels traitements sont mis en œuvre pour le compte de l'Etat, d'une personne morale de droit public ou d'une personne morale de droit privé gérant un service public, ils doivent être en principe²³ autorisés par décret en Conseil d'Etat pris après avis de la CNIL (article 27).

Ces règles relatives à l'utilisation du NIR trouvent à s'appliquer dès lors qu'il est envisagé d'exploiter les données gérées par l'assurance-maladie parmi lesquelles figure le NIR. Les organismes d'assurance-maladie disposent en vertu des articles R. 115-1 et R. 115-2 du code de la sécurité sociale d'une autorisation générale d'utilisation du NIR dans le cadre de leur mission de sécurité sociale. Cependant, une nouvelle autorisation est nécessaire dès lors qu'une utilisation est envisagée pour d'autres finalités. Ainsi, passé la phase d'expérimentation du programme « Sophia » d'accompagnement des patients diabétiques, la CNAMTS a dû obtenir, conformément à l'article 27 de la loi du 6 janvier 1978, une autorisation par décret en Conseil d'Etat pris après avis de la Cnil pour utiliser le NIR dans le cadre de programmes d'accompagnement des personnes atteintes de maladies chroniques²⁴.

Le respect de l'article 27 s'impose également pour procéder au rapprochement de données d'enquête collectées dans le cadre de programmes de recherche en santé et des données du SNIIRAM et du NIR. En effet, la section sociale du Conseil d'Etat a considéré que la circonstance que le NIR faisait l'objet d'un chiffrage dans le cadre du SNIIRAM était sans incidence sur l'applicabilité de l'article 27²⁵. Un décret en Conseil d'Etat pris après avis de la Cnil est donc nécessaire pour autoriser chaque programme de recherche utilisant les données du SNIIRAM et le NIR.

Toutefois, les dispositions combinées du IV de l'article 26 et du III de l'article 27 de la loi du 6 janvier 1978 permettent de procéder par un « *acte réglementaire unique* » à l'autorisation de tous les traitements qui « *répondent à une même finalité, portent sur des catégories de données identiques et ont les mêmes destinataires ou catégories de destinataires* ». Cette possibilité a été mise en œuvre pour la première fois dans le cadre du décret du 9 novembre 2012 mentionné ci-dessus. A cette occasion, la section sociale du Conseil d'Etat a estimé que « *la finalité en vue de laquelle des traitements de données peuvent être autorisés par un acte réglementaire unique doit être définie de manière suffisamment précise par cet acte et correspondre à un objectif suffisamment homogène et*

²³ Le II de l'article 27 permet dans certains cas de les autoriser par arrêté ministériel pris après avis de la CNIL. Mais les traitements comportant des données de santé ne peuvent rentrer dans ces exceptions.

²⁴ Autorisation donnée par le décret n° 2012-1249 du 9 novembre 2012 autorisant la création de traitements de données à caractère personnel pour la mise en œuvre de programmes de prévention et d'accompagnement en santé des assurés sociaux.

²⁵ Avis rendu sur le décret n° 2013-175 du 26 février 2013 portant création d'un traitement automatisé de données à caractère personnel dénommé « NutriNet-Santé », à paraître au rapport d'activité 2014.

circonscrit ». L'acte réglementaire unique ne peut donc porter sur une finalité définie de manière trop générale.

1.2.2.3 La loi du 17 juillet 1978 sur l'accès aux documents administratifs et la réutilisation des données publiques

La loi n°78-753 du 17 juillet 1978 (dite « loi CADA ») a posé les bases du cadre juridique de l'ouverture des données publiques et des informations concernées, en cohérence avec la loi « Informatique et Libertés » du 6 janvier 1978²⁶.

La loi CADA reconnaît à toute personne le droit d'obtenir communication des documents administratifs détenus par une administration ou une personne de droit privé dans le cadre d'une mission de service public. Sont concernés les documents administratifs « (...) *quels que soient leur date, leur lieu de conservation, leur forme et leur support, les documents produits ou reçus, dans le cadre de leur mission de service public, par l'Etat, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission. Constituent de tels documents notamment les dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, directives, instructions, circulaires, notes et réponses ministérielles, correspondances, avis, prévisions et décisions.* (...) » (article 1^{er}).

Sont exclus du champ des données communicables :

- les documents préparatoires à une décision administrative (article 2) ;
- les documents dont la communication porterait atteinte « *à la protection de la vie privée, au secret médical et au secret en matière commerciale et industrielle* » (article 6) : les documents comportant des données à caractère personnel doivent être rendus anonymes avant communication ;
- les documents dont la communication porterait préjudice au bon fonctionnement de l'Etat (secret de la défense nationale, secret des délibérations du Gouvernement, déroulement des procédures judiciaires engagées, etc. tels que listés dans l'article 6).

Outre le droit ainsi défini pour toute personne d'obtenir la communication de documents administratifs, les articles 10 à 19 de cette loi, issus d'une ordonnance du 6 juin 2005 qui transpose une directive 2003/98/CE du 17 novembre 2003, consacrent la liberté de réutilisation des données publiques. Elle est notamment soumise à la condition que ces dernières « ne soient pas altérées, que leur sens ne soit pas dénaturé et que leurs sources et la date de leur dernière mise à jour soient mentionnées » (art. 12).

Lorsque la réutilisation n'est possible qu'après anonymisation des données à caractère personnel, l'autorité détentrice y procède sous réserve que cette opération n'entraîne pas des efforts disproportionnés (art. 40 décret n° 2005-1755 du 30 déc. 2005).

Dans le cadre de la politique d'*open data*, le Conseil d'orientation de l'édition publique et de l'information administrative (COEPIA) a élaboré un Memento sur la protection des informations à

²⁶ La loi CADA a été complétée par diverses directives européennes, décrets et circulaires, pris en compte dans la présentation qui suit.

caractère personnel dans le cadre de l'ouverture et du partage des données publiques, afin de guider les administrations dans la marche à suivre lorsqu'elles envisagent de rendre publics des fichiers contenant des données à caractère personnel.

Les administrations visées par la loi CADA ont un devoir d'ouverture et de partage des données publiques, fondé sur cette loi.

En 2003, la directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003 concernant la réutilisation des informations du secteur public (« directive PSI ») – transposée par l'ordonnance du 6 juin 2005 et le décret n°2005-1755 du 30 décembre 2005 – a permis de compléter la loi « CADA », en y introduisant notamment un Chapitre II relatif à la réutilisation des informations publiques (*cf infra*). Elle a fait l'objet d'une révision par la directive du 26 juin 2013.

La circulaire du Premier ministre et le décret n° 2011-577 du 26 mai 2011 ont par ailleurs arrêté un principe de réutilisation libre, facile et gratuite pour tous les citoyens s'agissant des données publiques de l'Etat.

Au regard des articles 6 et 10 de la loi CADA, une information contenue dans un document administratif non communicable n'est pas réutilisable, sauf si ce document a fait l'objet d'une diffusion publique.

Ainsi, les documents préparatoires et non définitifs de l'administration ne peuvent être communiqués aux citoyens et les informations qui y sont contenues ne peuvent être réutilisées. De même, les informations contenues dans des documents protégés par des secrets légaux (art. 6 loi CADA) ne sont pas réutilisables (secret de la défense nationale, secret des délibérations du Gouvernement, etc.).

Les administrations ne sont pas tenues de communiquer les documents faisant par ailleurs l'objet d'une diffusion publique en vertu du 2^e alinéa de l'article 2 de la loi du 17 juillet 1978, tel que modifié par la loi du 12 avril 2000. Le législateur a considéré que la diffusion publique de ces documents remplissait la fonction de communication dudit document à l'ensemble des citoyens. Pour caractériser la « diffusion publique », la CADA s'attache notamment à l'accessibilité du document tant en termes géographiques, que techniques ou de prix. Depuis 2005, rompant avec une jurisprudence contraire, la CADA a ainsi considéré que des documents administratifs aisément accessibles sur internet, au moment où elle a à connaître de la demande de communication, font l'objet d'une diffusion publique. La CADA s'attache à l'accessibilité pratique du document. La CADA considère qu'une mise à disposition sur internet pendant une période limitée de temps ne suffit pas pour considérer que le document a fait l'objet d'une diffusion publique (conseil n° 20060837 du 16 février 2006). Par conséquent, une information contenue dans un document administratif ayant fait l'objet d'une diffusion publique est réutilisable au regard des dispositions de la loi CADA.

Le droit d'accès portant sur des documents existants, la Commission estime que l'administration n'a pas l'obligation de créer, d'adapter ou de fournir des extraits de documents lorsque cela nécessite des efforts disproportionnés. Elle considère, en revanche, comme des documents existants les données contenues dans des fichiers informatiques qui peuvent en être extraites par un « traitement

automatisé d'usage courant ». La loi du 17 juillet 1978 prescrit en effet la communication des documents administratifs quel qu'en soit le support. La CADA a ainsi donné un avis favorable à la communication de données agrégées relatives à la consommation du Mediator, en considérant que l'extraction des informations sollicitées ne nécessitait pas un traitement des données source de la base du SNIIRAM excédant un usage courant. Elle a relevé également que la communication portait sur des informations anonymes et globales, par année et par département, ne permettant pas, compte tenu de leur niveau d'agrégation, l'identification, même indirecte, des patients ou des médecins concernés (avis n° 20134348 du 21 novembre 2013). Au contraire, la communication d'extraits de la base nationale de pharmacovigilance (BNPV) et de la base relative à la gestion des fiches d'incidents transfusionnels (e-FIT) excède un traitement d'usage courant (conseil n° 20133264 du 10 octobre 2013).

Par ailleurs, la CADA considère que la diffusion publique de données à caractère personnel rendue possible en vertu d'une disposition législative, n'autorise pas leur réutilisation. Dans un avis du 19 avril 2012²⁷, elle a estimé que la publication des données concernant l'activité des médecins sur le site de l'assurance-maladie www.ameli.fr n'autorisait pas pour autant leur réutilisation à des fins d'exploitation commerciale.

✓ ***Les données à caractère personnel ne sont pas des données destinées à être rendues publiques***

L'article 13 de la loi CADA rend possible la réutilisation des données publiques comportant des données à caractère personnel, soit parce que les personnes intéressées y ont apporté leur consentement préalable, soit parce que les données à caractère personnel dans les données publiques ont fait l'objet d'une anonymisation, ou à défaut si une disposition légale ou réglementaire le permet. En tout état de cause, cet article 13 précise que la réutilisation des données publiques comportant des données à caractère personnel demeure soumise aux prescriptions de la loi « Informatique et Libertés » du 6 janvier 1978 modifiée.

Pour des motifs légitimes de protection de la vie privée et de respect du secret statistique, les informations statistiques doivent par ailleurs être publiées dans le respect du secret statistique, tel que défini par la loi du 7 juin 1951, ainsi que de l'article 285 du Traité instituant la communauté européenne.

S'agissant des professionnels de santé, il convient de souligner que les données relatives à leur activité et permettant de les identifier sont des données à caractère personnel au sens de la loi du 6 janvier 1978 modifiée.

1.2.3 Les évolutions envisagées dans le cadre de la proposition de règlement européen

La Commission européenne a adopté le 25 janvier 2012 une proposition de règlement relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données. Ce texte est destiné à se substituer à la directive 95/46/CE et

²⁷ Avis n° 20121581.

réforme donc de manière générale les règles de protection des données. Certaines de ces dispositions sont d'un intérêt particulier pour le sujet des données de santé :

- L'article 4 donne une définition large des données de santé, qui sont « *toute information relative à la santé physique ou mentale d'une personne, ou à la prestation de services de santé à cette personne* ». Le considérant 26 précise : « *Les données à caractère personnel concernant la santé devraient comprendre, en particulier, l'ensemble des données se rapportant à l'état de santé d'une personne concernée; les informations relatives à l'enregistrement du patient pour la prestation de services de santé; les informations relatives aux paiements ou à l'éligibilité du patient à des soins de santé; un numéro ou un symbole attribué à un patient, ou des informations détaillées le concernant, destinés à l'identifier de manière univoque à des fins médicales; toute information relative au patient recueillie dans le cadre de la prestation de services de santé audit patient; des informations obtenues lors d'un contrôle ou de l'examen d'un organe ou d'une substance corporelle, y compris des échantillons biologiques; l'identification d'une personne en tant que prestataire de soins de santé au patient; ou toute information concernant, par exemple, une maladie, un handicap, un risque de maladie, un dossier médical, un traitement clinique ou l'état physiologique ou biomédical de la personne concernée, indépendamment de sa source, qu'elle provienne par exemple d'un médecin ou d'un autre professionnel de la santé, d'un hôpital, d'un dispositif médical ou d'une épreuve diagnostique in vitro* ».

- La santé demeure sur la liste des données dont le traitement est en principe interdit, fixé par l'article 9.1. Toutefois, l'article 9.2 définit désormais une exception spécifique aux données de santé, en des termes assez larges : « *le traitement des données relatives à la santé est nécessaire à des fins liées à la santé, sous réserve des conditions et des garanties prévues à l'article 81* ». Le considérant 42 précise que des exceptions à l'interdiction du traitement des données sensibles sont notamment justifiées « *à des fins de santé publique, en ce compris la protection de la santé, la protection sociale et la gestion des services de santé, notamment pour assurer la qualité et l'efficacité des procédures de règlement des demandes de remboursement et de services dans le régime d'assurance-maladie* ».

- L'article 81 précise les cas et les conditions dans lesquels des traitements de données de santé sont autorisés. Ils doivent être effectués « *sur la base du droit de l'Union ou de la législation d'un État membre qui prévoit des garanties appropriées et spécifiques des intérêts légitimes de la personne concernée* ». Ils doivent être nécessaires, soit aux fins de la médecine préventive, des diagnostics ou des soins (et doivent alors être mis en œuvre par un praticien de la santé soumis au secret professionnel ou par une autre personne soumise à une obligation de confidentialité équivalente), soit pour des motifs d'intérêt général dans le domaine de la santé publique, soit pour des motifs d'intérêt général dans le domaine de la protection sociale, « *particulièrement afin d'assurer la qualité et la rentabilité en ce qui concerne les procédures utilisées pour régler les demandes de prestations et de services dans le régime d'assurance-maladie* ».

Somme toute, la proposition de règlement ne paraît pas modifier en substance le cadre juridique du traitement des données de santé. Elle promeut cependant une approche large de la notion de données de santé et précise les cas dans lesquels ces données peuvent être traitées. Les traitements de données de santé, tels que le SNIIRAM, mis en œuvre par les organismes d'assurance-maladie, apparaissent confortés par ces dispositions. Le projet de règlement souligne davantage, à l'article 81,

le rôle des motifs d'intérêt général comme pouvant justifier le traitement des données de santé, ce qui conforte l'approche de la commission incluant explicitement le bénéfice collectif attendu dans les critères d'autorisation d'accès au SNIIRAM (voir infra 3.5.3.2).

La proposition de règlement doit être approuvée par le Conseil et le Parlement européen dans le cadre de la « procédure législative ordinaire » (ex-procédure de codécision). Le Conseil n'ayant pas encore pu dégager de position commune, il ne pourra probablement entrer en vigueur avant le début de l'année 2015.

1.3 La gouvernance

La gouvernance des données médico-administratives peut s'analyser en quatre fonctions :

- La décision stratégique et la réglementation,
- L'expression des parties prenantes (évaluation, propositions, demandes, orientations),
- La décision de donner ou non accès aux données selon les cas, en application des lois,
- La gestion technique des données et le service de leur mise disposition.

De manière factuelle et pour ce qui concerne le SNIIRAM, la loi inscrite dans le code de la sécurité sociale a délégué à l'assurance maladie la mission de gérer le système d'information inter-régimes de l'assurance maladie. A cette fin, le protocole conclu entre les caisses nationales a également créé un comité d'orientation et de pilotage de l'information interrégimes (COPIIR SNIIRAM), constitué des représentants de l'assurance maladie, des professionnels de santé et de l'Etat. Cet organe de pilotage du SNIIRAM, outre le suivi de son développement, du périmètre des données, organise et met en place l'accès aux données pour les partenaires et les différents acteurs. Le COPIIR est ainsi devenu à la fois l'organe de gestion et le principal organe de gouvernance du système des données médico-administratives du SNIIRAM. Ses décisions sont transcrites dans un arrêté du ministre chargé de la sécurité sociale et soumises à la Cnil avant publication au JO. C'est ce même arrêté qui distingue les droits d'accès permanents et les droits ponctuels et qui a confié au bureau de l'Institut des données de santé (IDS) la mission « d'approuver » les demandes d'accès aux données émanant des organismes de recherche sans but lucratif avant que ces demandes ne soient soumises à la Cnil. Les droits permanents sont régis, après accord de la Cnil :

- Par l'arrêté ministériel issu du vote du COPIIR pour les accès dits permanents ;
- Par la loi pour les ARS dont l'accès a été prévu par la loi HPST, qui a créé les ARS.

Au fil du temps, les versions successives de l'arrêté SNIIRAM, de plus en plus complexes, sont le principal document de référence pour les évolutions du système et les règles d'accès aux données.

La Cnil examine les demandes d'accès ponctuels (i.e. des demandes d'accès ou des demandes d'extraction par des acteurs qui ne disposent pas d'accès permanent) mais ces demandes transitent par des canaux différents selon les cas :

- ou bien la Cnil se prononce seule, par exemple, pour les demandes d'accès au PMSI, en vertu des règles énoncées au chapitre X de la loi « Informatique et Libertés » ;
- ou bien elle prend l'avis d'un comité d'experts (le CCTIRS) si la demande d'accès aux données médico-administratives s'inscrit dans un projet de recherche en santé, en application des dispositions du chapitre IX de la loi « Informatique et Libertés » ;
- Ou bien la demande est soumise d'abord à l'approbation du bureau de l'IDS, comme on l'a vu.

L'architecture actuelle de cette gouvernance est largement considérée comme complexe. Ainsi le rapport Bras qualifie les règles actuelles de « *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. »

Il est vrai que les acteurs sans droit de vote au COPIIR du SNIIRAM ont pu avoir le sentiment de n'être pas assez entendus, que la complexité des règles applicables, notamment pour les appariements de données nécessitant le NIR, a suscité l'exaspération des chercheurs, que la complexité des données elles-mêmes a suscité une demande pour davantage de services aux utilisateurs et que les restrictions d'accès aux données anonymes sont incompréhensibles à l'heure de l'open data.

L'Etat pour sa part n'a sans doute pas réagi assez tôt au besoin d'une doctrine claire associant l'ouverture des données, l'efficacité de leur utilisation et la protection de la vie privée. C'est à l'élaboration d'une telle doctrine que le rapport Bras a tenté d'apporter des réponses.

Plusieurs membres soulignent néanmoins l'avancée constituée par la création de l'Institut des données de santé (créé par la loi en 2004, et mis en place effectivement en 2007). L'IDS est un groupement d'intérêt public, constitué entre l'Etat, l'assurance maladie, les représentants des patients, des professionnels, et des établissements de santé publics et privés, les organismes complémentaires. Ses missions sont « 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'Etat, 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'Etat pris après avis de la Commission nationale de l'informatique et des libertés ».

Il semble nécessaire en tout état de cause de clarifier les principes et les rôles :

- Les finalités des données et les principes d'accès ;
- Les rôles des organes de gouvernance.

Le secret statistique

Lorsque la demande d'accès porte sur des données couvertes par le secret statistique, des règles particulières sont appliquées, ce qui rend le dispositif d'ensemble encore plus complexe ; mais les critères utilisés par le "comité du secret statistique" (abrité par le Conseil national de l'information statistique) pour autoriser cet accès, plus détaillés que ceux mentionnés dans les chapitres IX et X de la loi « Informatique et Libertés », pourraient inspirer une amélioration des textes actuels.

Sont couvertes par le secret statistique les données produites dans le cadre de la loi 51-711 sur l'obligation, la coordination et le secret en matière de statistiques. Il s'agit essentiellement de données collectées ou retraitées²⁸ par l'un des membres du système statistique public, constitué de l'Insee et des services statistiques ministériels (dont la Drees pour le ministère chargé de la santé).

²⁸ Données produites par des administrations, des organismes publics ou des organismes privés chargés d'une mission de service public (par exemple les déclarations de revenus à la DGFIP) et exploitées par le système statistique public pour produire des statistiques (dans l'exemple choisi, des statistiques de revenu des ménages français).

Au cas présent, les données couvertes par le secret statistique concernées par un accès sont principalement soit des données de santé collectées auprès d'un échantillon de ménages par l'Insee ou la Drees (par exemple, celles issues des enquêtes décennales de l'Insee sur la santé) soit des données statistiques (au sens de la loi 51-711) portant sur d'autres domaines (par exemple relatives aux revenus) que des chercheurs souhaiteraient apparier aux données de santé auxquelles ils ont accès.

La loi 51-711 prévoit que les données individuelles couvertes par le secret statistique ne peuvent faire l'objet d'aucune communication avant un délai de 75 ans. Elle prévoit toutefois une dérogation à ce principe général pour des demandes effectuées à des fins de statistique publique ou de recherche scientifique ou historique, toute communication de données individuelles nécessitant dans ce cas une décision de l'administration des archives, l'accord de l'autorité dont émanent les données et un avis favorable du comité du secret statistique.

Le comité du secret statistique est ainsi chargé d'instruire les demandes d'accès aux données couvertes par le secret statistique. Le décret n° 2009-318 du 20 mars 2009 modifié prévoit que le comité du secret statistique « émet son avis en prenant en compte la nature et l'intérêt des travaux pour l'exécution desquels la demande est formulée ainsi que la qualité de la personne ou de l'organisme présentant la demande et les garanties qu'il présente. Il vérifie que le volume des informations demandées n'est pas excessif par rapport aux travaux qui justifient leur communication et que celle-ci ne conduit pas à porter une atteinte excessive aux intérêts que la loi du 7 juin 1951 susvisée a entendu protéger. Dans le cas de demandes portant sur des faits et comportements d'ordre privé, il vérifie que cette demande est effectuée à des fins de statistique publique ou de recherche scientifique ou historique. Dans tous les cas, le comité du secret statistique détermine les conditions dans lesquelles ces renseignements individuels pourront être portés à la connaissance du demandeur. ». Selon les données concernées et les traitements qui leur seront appliqués (en cas d'appariements notamment), le comité peut demander que l'accès aux données et leur traitement se fassent ou pas par le biais d'un centre d'accès sécurisé à distance.

La loi 51-711 dispose que « Le comité est présidé par un conseiller d'État, désigné par le vice-président du Conseil d'État. Il comprend notamment des représentants de l'Assemblée nationale et du Sénat. La composition et les modalités de fonctionnement du comité sont fixées par le décret n°2009-318 du 20 mars 2009 modifié ».

1.4 Cartographie des accès au SNIIRAM et au PMSI

Le Système national d'information inter-régimes de l'assurance maladie (SNIIRAM), issu des données de remboursement de soins de ville et géré par l'assurance maladie, et le Programme de médicalisation des systèmes d'information (PMSI), issu des données de séjours hospitaliers, constitué par l'Agence technique de l'information sur l'hospitalisation (ATIH), font d'ores et déjà l'objet d'une ouverture large.

1.4.1 Les accès aux données de soins de ville

L'accès aux données de santé de l'assurance maladie est organisé selon trois modes conditionnés par l'encadrement juridique (législatif et réglementaire) de la protection des données personnelles et des finalités d'utilisation : 1) les statistiques et les services accessibles à tous, 2) les tableaux de bord de données partagées au sein de l'Institut des données de santé et 3) les données accessibles par les organismes publics sur autorisation de la Cnil, depuis le portail sécurisé du SNIIRAM.

1.4.1.1 Les données à destination du grand public

- **Etudes et statistiques mises en ligne sur www.ameli.fr**

→ Etudes sur le médicament, exemples :

- Etude sur le risque d'embolie pulmonaire, d'accident vasculaire cérébral ischémique et d'infarctus du myocarde chez les femmes sous contraceptif oral combiné en France,
- L'utilisation du Baclofène dans la dépendance contre l'alcool,
- Etude de cohorte sur le risque de cancers chez les personnes exposées à l'insuline glargine,
- Etude de cohorte sur le risque de cancer de la vessie chez les personnes diabétiques traitées par pioglitazone en France.

→ Données statistiques

- Analyse des dépenses (régime général),
- Démographie /activité / honoraires des professionnels de santé libéraux (Inter-Régime),
- Affections de longue durée (ALD),
- CMU,
- Médicament,
- Biologie,
- Liste des produits et prestations.

- **Information sur professionnels de santé et établissements sur www.ameli-direct.fr**

→ Professionnels de santé : l'assuré peut trouver un professionnel, connaître ses coordonnées, le secteur d'activité conventionnel auquel il appartient et ses tarifs, les actes qu'il pratique (ex : échographie, chirurgie de la prostate...), le fait qu'il accepte ou non la carte vitale, ...

→ Etablissements de santé : l'assuré peut se renseigner sur les chirurgies pratiquées et les pathologies traitées dans l'établissement et avoir accès à divers indicateurs : indicateurs de volume et de niveau (nombre d'hospitalisation, durée moyenne de séjour, taux de chirurgie ambulatoire...) / données tarifaires sur le coût de l'hospitalisation / indicateurs de qualité pour les établissements ayant obtenu une certification de la HAS.

1.4.1.2 Les données partagées au sein de l'Institut des données de santé

- **Tableaux de bord accessibles aux membres de l'IDS**

→ sur www.institut-des-donnees-de-sante.fr : reste à charge après remboursement de l'assurance maladie, suivi des dépenses de soins de ville, analyse de l'offre de soins, suivi détaillé sur des axes de gestion du risque, analyse de l'activité des établissements.

1.4.1.3 Les données du SNIIRAM

- **Catégories de données**

→ 15 bases de données thématiques de données agrégées appelées datamarts orientés vers une finalité particulière : suivi des dépenses (DAMIR), analyse de l'offre de soins libérale, biologie, pharmacie, dispositifs médicaux, établissements privés : 71 types d'organismes²⁹ ont accès aux datamarts.

→ un échantillon général des bénéficiaires (EGB) au 100e de la population protégée qui permet de réaliser des études longitudinales et d'analyser le parcours individuel de près de 660 000 bénéficiaires en ville et à l'hôpital : 64 types d'organismes ont accès à EGB-PMSI.

→ une base de données individuelles des bénéficiaires (DCIR) pour réaliser des études plus approfondies sur la consommation des soins : 6 types d'organismes ont accès au DCIR-PMSI (CNAMTS, CCMISA et RSI ; CNSA ; INVS ; HAS ; ANSM ; ARS sur leur champ de compétence régional et sous l'autorité d'un médecin salarié de l'agence).

→ extractions de données individuelles : le CCTIRS émet un avis pour les recherches en santé publique et la création de cohortes nominatives, l'IDS donne son approbation pour les études en santé publique et les cohortes ; ensuite la Cnil autorise ou non les extractions : 38 extractions ont été livrées de 2010 à 2013 à 17 organismes de recherche.

- **Trois types d'accès sont autorisés aux organismes publics selon leurs missions et les finalités d'utilisation des données :**

→ accès permanents : le comité d'orientation et de pilotage de l'information interrégimes (Etat-Assurance maladie-UNPS) approuve les accès permanents des

²⁹ Type d'organisme : ARS compte pour 1, URPS également alors qu'autant que nombre de régions, Caisse locale ou régionale d'assurance compte pour 1 alors que près de 150, etc.

organismes publics lui ayant formulé une demande ; ces accès sont transcrits dans le protocole SNIIRAM et publiés au JO par arrêté pris après avis de la Cnil : **35 types d'organismes ont des accès permanents ce qui représentent plus de 250 organismes nationaux, régionaux et locaux.**

→ accès temporaires: l'IDS approuve les accès temporaires (données agrégées, échantillon EGB) pour les organismes publics à des fins d'études en santé publique : **54 accès temporaires de 3 ans sont ouverts à 30 organismes.**

1.4.2 Les accès aux données du PMSI

1.4.2.1 Restitution des données hospitalières

Les données traitées par l'ATIH sont accessibles via un outil de restitution appelé Système national d'information sur l'hospitalisation (SNATIH).

En ligne sur le site de l'agence, par le menu [Accès aux données > SNATIH] cet outil permet d'effectuer des requêtes afin d'obtenir des résultats mis en forme ou tableaux de bords.

En accès libre, on peut trouver des données statistiques médicales de synthèse ainsi que la cartographie de l'activité des établissements de santé.

En accès avec identifiant pour les acteurs du système de santé : des applications relatives à la performance des établissements de santé, à la maîtrise des dépenses et à l'analyse de l'activité.

Exemples de données en ligne sur SNATIH

- Activité des établissements de santé (PMSI) : nombre de patients traités pour une affection donnée, nombre d'interventions médicales pour une catégorie d'intervention...
- Coûts : coût d'une prise en charge hospitalière, coûts comparés par territoire...
- Financière, comptable et ressources humaines : connaître la situation financière d'un établissement...
- Suivi des dépenses d'assurance maladie pour les soins hospitaliers.
- Qualité : indicateurs de risque d'infection nosocomiale dans un établissement...

1.4.2.2 Diffusion de l'information hospitalière

Cette diffusion a pour objectif de :

- Contribuer à la connaissance de l'activité des établissements de santé et de la nature des pathologies prises en charge,
- Enrichir les travaux d'organisation des soins hospitaliers,
- Prendre et suivre des mesures nationales de politique de santé.

L'ATIH diffuse l'ensemble des données détaillées du PMSI à tout titulaire d'un accord de la Commission nationale de l'informatique et des libertés (Cnil), délivré en fonction de l'utilisation

prévue. Les publics ayant demandé cet accès sont divers : des établissements de santé, des laboratoires de recherche, des organismes complémentaires, des sociétés privées... Ces données sont facturées pour l'ensemble des demandes ne visant pas à la réalisation de travaux de recherche à finalités non marchandes dans le domaine de la santé.

Concernant les données PMSI, l'ATIH a mis en place un système de récupération automatique des données à la demande pour les ARS, et les fournit aux autres acteurs institutionnels (DGOS, DSS, DREES, CNAMTS, Fédérations hospitalières) suivant des périodicités propres à chacune. L'ensemble de ces acteurs disposent naturellement des autorisations Cnil nécessaires.

L'ATIH effectue enfin des traitements ciblés à la demande des services du ministère et d'agences nationales dans le domaine de la santé. *Exemples* : étude sur les petites maternités, sur la prise en charge des patients atteints de la maladie de Parkinson, comparaison des séjours entre régions pour certaines pathologies, analyse de l'absentéisme dans les établissements...

1.5 Cartographie des données de santé

Afin d'alimenter les travaux de la Commission, la mission Etalab a établi une liste détaillée des données de santé. Cette cartographie a été réalisée à partir d'une revue des principaux rapports existants sur les données de santé en France et de recherches menées à partir des sites Internet des principaux gestionnaires de données, complétées lors d'entretiens avec une trentaine d'organisations gestionnaires de données de santé.

Ces entretiens ont permis d'aborder six questions principales :

- La liste des bases gérées ;
- L'origine des données utilisées et les modalités actuelles de partage de ces données ;
- Le format des données et le format de mise à disposition le cas échéant (ex : xls, pdf, csv...) ;
- Les actions préalables nécessaires ou difficultés identifiées en vue d'une mise à disposition ;
- Les actions possibles pour lever les freins ;
- Une échéance de publication.

La cartographie des données de santé ainsi produite identifie plus de 240 jeux et bases de données (voir Annexe 5) et décrit leur niveau d'ouverture actuel. Deux niveaux de granularité sont décrits pour qualifier le niveau d'ouverture de chaque base ou jeux de données :

- Le niveau granulaire (données au niveau le plus fin qu'il est possible d'obtenir en fonction de l'origine de la donnée et du système de collecte) ;
- Le niveau agrégé (données obtenues en regroupant des données granulaires selon une ou plusieurs caractéristiques communes).

Afin de faciliter les recherches des bases et jeux de données, les jeux de données ont été triés selon une typologie ad hoc (certaines jeux/bases de données se retrouvent dans plusieurs catégories). Cette typologie est présentée ci-dessous.

Santé publique et épidémiologie	<ul style="list-style-type: none"> ▪ Etat de santé individuel : facteurs de risques, ALD, diagnostics, données morphologiques et physiologiques (données / résultats cliniques du patient) ▪ Epidémiologie ▪ Environnement par exemple : air, eau, sols
Offre de soins	<ul style="list-style-type: none"> ▪ Infrastructures : MCO, SSR, SLD, EHPAD, professionnels libéraux, etc. ▪ Services/ soins proposés ▪ Personnel : en activité, en cours de formation ▪ Tarifs et honoraires ▪ Infrastructure de recherche ▪ Informations médicales : maladies, parcours, médicaments
Consommation de soins	<ul style="list-style-type: none"> ▪ Activité des établissements de santé: MCO, SSR, SLD, Psy ▪ Consultations : médecins, dentistes, auxiliaires ▪ Biologie ▪ Transports ▪ Autres (LPP...) ▪ Médicaments et dispositifs : pharmacie, liste en sus ▪ IJ, AT, etc. ▪ Activité des établissements médico-sociaux
Finance	<ul style="list-style-type: none"> ▪ Dépenses de soins et de biens médicaux ▪ Dépenses de soins longue durée ▪ Dépenses de prévention ▪ Autres dépenses en faveur du système de soins (subventions, recherche et formation, etc.) ▪ Dépenses administratives et de gestion ▪ Modalités/sources de financement
Performance et opérations	<ul style="list-style-type: none"> ▪ Performance financière ▪ Performance opérationnelle ▪ Qualité (des soins, sécurité, conditions d'accueil, satisfaction des patients, etc.)

La cartographie est riche et volumineuse. Elle est présentée en annexe 6 et est disponible depuis le 2 avril 2014 sur le site www.data.gouv.fr.

1-6 Enseignements du panorama international

Un panorama des initiatives d'open data en santé à l'étranger a également été réalisé pour éclairer les débats de la Commission. Ce panorama présente pour cinq pays (Royaume-Uni, Etats-Unis, Danemark, Canada et Singapour) le type de données mises à disposition du public, les modalités et le contexte de leur ouverture. Quelques cas de réutilisation remarquables de ces données sont également décrits. Le panorama complet est disponible en ligne ; une synthèse des principaux enseignements se trouve en annexe (annexe 7).

Ce panorama identifie trois approches de l'open data en santé dans les pays étudiés :

- Au Canada et au Danemark, les données ouvertes renforcent l'information du patient sur le système de soins :
 - Mise à disposition de données présentant une large couverture, une bonne profondeur (quantité et précision des données) et un volume conséquent ;
 - Volonté de rendre les données intelligibles et accessibles aux citoyens via les services développés.

- Aux Etats-Unis et au Royaume-Uni, l'open data permet une véritable autonomisation du patient dans son parcours de soins :
 - Mise à disposition de données permettant la comparaison des offreurs de soins ;
 - Développement d'outils d'aide à l'orientation dans le système de soins.






- Enfin, Singapour, engagé plus récemment dans l'open data, met l'accent sur l'accompagnement dans la réutilisation des données :
 - Peu de données granulaires ;
 - Services de visualisation et création d'interfaces de programmation ouvertes à la communauté.

Ce panorama montre le caractère récent des démarches d'open data en santé, même dans les pays disposant d'une avance sur la France. Si le principe d'une plus grande ouverture des données de santé ne semble plus faire débat, la démarche d'ouverture – quelles solutions légales, réglementaires, or organisationnelles ; quelles catégories de données ouvrir et avec quelle granularité – reste matière à expérimentation. Le cas récent du programme « Care.data » au Royaume-Uni, même s'il ne concerne pas directement l'open data³⁰, illustre en particulier la difficulté de bien clarifier la frontière entre open data et partage des données personnelles ou réidentifiantes à des fins de recherche, commerciale ou non commerciale.

En effet, le programme « Care.data », lancé par le NHS (le système de santé anglais) a pour objectif de rassembler dans une même base des données provenant du secteur médical et social, dont les données des dossiers médicaux tenus par les médecins ainsi que les données du Ministère de l'éducation sur les enfants, leur handicap, leur prise en charge par les services sociaux, les exclusions de l'école, l'origine ethnique, etc. Il est prévu que les données individuelles ne soient disponibles que sous des règles strictes pour un public ciblé, à des fins de veille, de recherche et de planification des politiques publiques. Cependant, l'annonce de la collecte de ces données dans une même base, faite à travers une campagne de communication qui semblait de petite ampleur par rapport aux enjeux du projet, a soulevé l'indignation d'une partie de la population anglaise. Bien que les règles de communication de ces données soient strictes et réparties en trois niveaux de risque (données anonymes, indirectement identifiantes et réidentifiantes) et que les patients et parents d'élèves aient le droit de refuser que ces détails soient stockés sur la base de données nationale, le lancement de la base a dû être reporté suite aux protestations et à la campagne médiatique alarmante.

Parmi les exemples remarquables d'utilisation de ces données, on peut citer :

³⁰ En effet, les données de cette base sont essentiellement réidentifiantes, tout comme le sont les données du SNIIRAM-PMSI.

Pays	Exemples de données
	<ul style="list-style-type: none"> ▪ <i>Authorization Register</i> : registre contenant les données relatives à tous les professionnels de santé danois autorisés à exercer notamment nom, prénom, date de diplôme, statut d'activité, lieu d'activité ▪ <i>Register of Medicinal Products Statistics</i> : registre des ventes de médicaments par produit, par année, par région et en fonction de l'âge et du sexe de l'acheteur
	<ul style="list-style-type: none"> ▪ Base de donnée du CPCD : montants moyens facturés par type d'intervention, par région et en fonction de l'âge du patient ▪ <i>Canadian Institute of Health Information</i> : temps d'attente moyen et médian de prise en charge par type d'intervention pour 8 interventions principales par région ▪ <i>Canadian Institute of Health Information</i> : temps passé par patient pour les infirmières par hôpital, par région, par année
	<ul style="list-style-type: none"> ▪ <i>NHS Choices</i> : note de satisfaction des patients sur environ 95% des cabinets médicaux ▪ <i>NHS Choices</i> : pourcentage d'employés qui recommanderait la structure dans laquelle ils exercent par établissement de santé ▪ <i>Publication Hospital Prescribing Data</i> : consommation mensuelle de médicaments par molécule à l'hôpital ▪ <i>Public Health England</i> : profil d'activité et d'efficacité sur le cancer par cabinet médical
	<ul style="list-style-type: none"> ▪ <i>Hospital Compare</i> : données de performance opérationnelle sur les hôpitaux telles que les taux de réadmission à 14 et 30 jours par type d'intervention ▪ <i>Hospital Compare</i> : données de qualité sur les hôpitaux telles que des indicateurs de propreté, de communication avec le personnel, de niveau d'information ▪ Base WONDER : nombre d'événements indésirables par type de vaccins et en fonction de l'âge et du sexe du patient
	<ul style="list-style-type: none"> ▪ Plateforme <i>data.gov.sg</i> : historique du nombre de cas de dengue déclarés par quartier ▪ Ministère de la Santé : taux d'occupation hebdomadaires des lits par hôpital

Source : McKinsey, Synthèse du panorama de l'open data à l'étranger (mars 2014).

Partie 2

Enjeux liés à l'ouverture des données de santé

L'informatisation des processus de gestion et (de plus en plus) de soins génère un volume croissant de données numériques. Leur regroupement dans de vastes bases de données structurées ou de simples entrepôts de données permet de disposer d'une « matière première » dont l'intérêt dépasse largement les finalités initiales qui ont conduit à leur constitution.

C'est le cas, par exemple :

- des données de liquidation de l'assurance maladie issues des processus de gestion des remboursements, concentrées au sein du SNIIRAM à des fins de suivi des dépenses pour lesquels des usages diversifiés et de grand intérêt se sont fait jour au fil du temps : santé publique, pharmacovigilance, aide à l'organisation des soins...
- des données issues du PMSI collectées à des fins de connaissance de l'activité de chaque établissement puis de tarification et dont les usages se sont également multipliés : calcul d'indicateurs de qualité des soins, support d'études médico-économiques...

De façon plus prospective, il est facile d'imaginer (le NHS en donne l'illustration avec la *General Practice Research database*) que les données de soins anonymisées et regroupées dans des bases de données pourront être également une source majeure de progrès des connaissances et de création de valeur. D'ores et déjà, plusieurs régions ont pris l'initiative de réunir les données issues des résultats d'antibiogramme individuels au sein de bases de données permettant ainsi de suivre l'évolution des flores bactériennes et leurs résistances aux différentes classes d'antibiotiques. Au-delà, on peut ainsi facilement imaginer le potentiel de connaissance que représenterait une base consolidée et anonymisée de dossier médical partagé (DMP) dont les données seraient structurées voire codées.

Les enjeux de l'accès aux données de santé sont de plusieurs natures qu'il convient de rappeler rapidement. Ils touchent à la démocratie sanitaire, à l'autonomisation des patients, à l'efficacité de l'action publique, aux progrès de la recherche et à la capacité de notre pays à favoriser l'innovation et à rester compétitif en matière d'accueil d'entreprises du champ de la santé.

2.1 Renforcer la démocratie sanitaire

Le premier enjeu est de permettre aux citoyens de peser davantage sur les orientations du système de santé, en fondant leurs contributions sur une connaissance renforcée de son fonctionnement. L'accès aux données qui supportent la décision publique constitue une condition nécessaire pour permettre une participation active des parties prenantes aux instances de la démocratie sanitaire. Ainsi que le souligne le rapport de Pierre-Louis Bras, les données de santé peuvent par exemple servir à documenter des inégalités territoriales de santé ou d'offres de services aussi bien que les pratiques de dépassement d'honoraires.

Ces accès constituent aussi un outil de transparence en permettant aux parties prenantes de construire leurs propres évaluations des politiques publiques.

Parties prenantes du dialogue dans le cadre de la démocratie sanitaire, les professionnels et les établissements de santé, par le biais des organisations qui les représentent, doivent pouvoir également accéder aux données les concernant collectivement.

Les associations de patients et d'usagers du système de santé mettent aussi en avant la possibilité de jouer leur rôle de lanceur d'alerte, par exemple sur les questions de sécurité sanitaire, ou encore la possibilité d'aider plus efficacement leurs adhérents en matière d'accès à l'assurance en leur garantissant notamment que les éventuelles exclusions ou surprimes seront évaluées au regard des données actualisées de survie et d'invalidité les concernant.

Enfin, l'accès aux données permet de construire un partage des connaissances entre acteurs du système de santé (Assurance maladie obligatoire, organismes complémentaires et offreurs de soins) et représentants des usagers fondé sur une base d'informations commune qui servira donc de référence et facilitera le dialogue entre les parties prenantes. Ce partage des connaissances est indispensable pour le débat démocratique, l'évaluation de la réponse aux besoins de santé et la recherche collective d'une plus grande efficacité de notre système de santé.

2.2 Renforcer l'autonomisation du patient

Le deuxième enjeu touche les patients citoyens au plan individuel³¹. Il s'agit de faciliter l'émergence de nouveaux services d'information sur l'offre et les parcours de soins répondant aux besoins des patients. Au-delà de la connaissance « à plat » de l'offre, il s'agit surtout de permettre aux citoyens de prendre des décisions plus informées en matière de choix des offreurs de soins auxquels ils recourent. La mise à disposition d'informations doit ainsi contribuer à réduire l'asymétrie d'information entre patients et professionnels de santé, offrant la possibilité d'une forme de concurrence par la qualité constitutive d'une dynamique d'amélioration continue de la qualité du système de santé.

L'accès effectif à l'information du citoyen doit être garanti par la publicité faite sur cet accès mais aussi par les possibilités de démultiplication de l'information et de réutilisation par des tiers.

2.3 Accroître l'efficacité de l'action publique

Les voies d'amélioration permises par l'accès aux données sont multiples. Elles touchent aussi bien à la qualité des soins, la promotion des bonnes pratiques, qu'au bon usage des ressources.

D'ores et déjà, la diversité des études menées à partir du SNIIRAM dans le champ de la santé publique illustre les bénéfices que l'on peut tirer de son exploitation. L'étude relative au MEDIATOR a ainsi montré la possibilité de mesurer le suivi en vie réelle d'un médicament.

D'autres exemples peuvent être donnés :

³¹ Par ailleurs, les droits reconnus par la loi « Informatique et Libertés » aux personnes sur la collecte et l'utilisation de leurs données personnelles s'inscrivent également dans le cadre d'un renforcement de l'autonomie du patient.

- la possibilité d'un suivi amélioré de la performance des acteurs du système de santé. Il s'agit bien sûr de leurs performances individuelles mais aussi et surtout de leur performance collective dans leur capacité à mieux coordonner leurs interventions (médecine de parcours). La possibilité d'assurer un suivi longitudinal et d'analyser l'ensemble du processus de soins sur les différents épisodes du parcours du patient est ici essentielle ;
- une meilleure information des professionnels de santé, indépendante de celle donnée par les industriels. L'information donnée par les industriels sur leurs produits est essentielle mais elle ne doit pas être la seule source de données accessible aux professionnels. La diffusion large de l'information permet de diversifier les sources de données comme les travaux qu'elles rendent possibles ;
- la possibilité d'engager un cercle vertueux d'amélioration de la qualité des pratiques par la pression des pairs et le retour d'information entre producteurs et utilisateurs de données. De nombreuses études l'ont montré, le principal levier d'action de la diffusion publique d'information sur la qualité des pratiques tient au souci de « bonne réputation professionnelle » vis-à-vis des patients mais aussi vis-à-vis des pairs ;
- l'opportunité pour les administrations de développer un nouveau canal de communication et d'échanges avec leurs usagers, par lequel ces derniers peuvent contribuer à améliorer les données utilisées par l'administration, renforcer leur visibilité par des *datavisualisations* ou simplement par une diffusion supplémentaire, et de ce fait d'améliorer la perception des citoyens vis-à-vis de leur administration.

Les améliorations attendues concernent également les institutions et leurs relations avec les professionnels de santé. On peut notamment anticiper une réduction des demandes de reporting et un dialogue enrichi et plus pertinent entre les institutions publiques et les professionnels sur la base de données à la fois plus riches et partagées.

Plus généralement, l'accès aux données, de nature et d'origine diverses, permettra à chaque acteur du système, quel que soit son rôle (offreur de soins, régulateur, financeur...), d'améliorer l'efficacité de son action, par une meilleure connaissance et un plus grand partage des informations disponibles.

- C'est le cas pour l'action publique pour laquelle la transparence pourrait contribuer à un décloisonnement entre institutions (caisses d'assurance maladie, Etat, agences sanitaires, agences régionales de santé, acteurs locaux...) favorable à l'efficacité collective.
- C'est le cas pour les professionnels de santé et acteurs de terrain, en renforçant la logique du retour d'information (qui, pour mémoire, constituait l'un des objectifs fondateurs du SNIIRAM) et en l'élargissant à une meilleure connaissance globale du système de soins.
- C'est le cas pour les organismes complémentaires d'assurance maladie : en tant que financeurs et porteurs du risque pour la part de dépense sur laquelle ils interviennent, l'accès permanent à une information statistique plus large et plus riche que les seules données de leur propre portefeuille est essentiel. Cette information leur permet de mieux apprécier les risques qu'ils couvrent, d'accroître les possibilités de régulation et de réduction des restes à charge, d'améliorer les services rendus aux assurés.

Enfin, la pluralité des utilisations de ces données enrichira la documentation de ces bases et poussera à l'amélioration de leur qualité et de leur capacité à répondre aux besoins des acteurs du système de santé et à ses usagers.

2.4 Permettre aux professionnels de santé de renforcer la connaissance de leur exercice et l'amélioration de sa qualité

L'information du patient sur le système de soins, l'autonomisation du patient dans son parcours, l'accompagnement dans la réutilisation des données sont les trois principales approches identifiées de l'Open Data en santé. Ces choix ont été influencés principalement par la structuration des systèmes et des données dans chaque pays.

L'exploitation des données par les professionnels de santé apparaît nécessaire à la réalisation de la démocratie sanitaire, dans les faits, comme un pendant natif au renforcement de l'autonomisation des patients.

En effet, l'accès aux données qu'ils produisent individuellement permettra aux professionnels de santé de renforcer la connaissance de leur exercice, donc l'amélioration de sa qualité.

Cette réflexion s'appuie sur l'état des données sur les professionnels de santé dans cinq pays réalisés pour la commission :

		CANADA	DANEMARK	USA	UK	SINGAPOUR
Notions sociodémographiques	Diplômes, parcours universitaire	oui	oui (réutilisables)	oui	oui	oui
	Expérience professionnelle, formation continue	oui	oui	oui	oui	oui
	Situation géographique	oui	oui	oui	95% des praticiens	oui
Performance	Professionnels des hôpitaux	/	/	oui	pas de données sur les hôpitaux privés	/
	Activité professionnelle	oui	oui	oui	oui	oui
	Qualité	Quelques enquêtes de satisfaction	Questionnaires de satisfaction sur la totalité des hôpitaux exploitables et réutilisables	oui	100% des cabinets médicaux, 148 indicateurs + nombreuses enquêtes	Quelques jeux de données, non accessibles en Open data seulement en consultation
	Tarifs	Estimation des coûts	Seulement pour l'hôpital. Non réutilisables	Seulement quelques bases (Medicare)	oui	oui
Traitement	Réutilisation	Non sauf pour la qualité	Non réutilisables, non exploitables	Non exploitables car en ligne, non réutilisables si but commercial	oui, réutilisables	Non, seulement consultation en ligne
	Profondeur des données	Faible	Forte	Moyenne à forte (30% sur les médicaments)	Forte	limitée
	Agrégation des données sur les professionnels de santé	oui	Données agrégées en open data sur les consultations du secteur primaire	Non	/	oui

Réalisé à partir des documents présentés à la commission, ce tableau synthétique, même s'il n'est pas totalement exhaustif, permet d'appréhender la problématique de « l'open data et les professionnels de santé ». Il permet d'identifier facilement les points forts que possède la France dans l'élaboration de son propre modèle : la première partie du tableau est déjà accessible au patient, mais pas toujours facilitée, d'où l'importance du respect de la notion d'accès.

Il apparaît également que la qualité, même si elle est évaluée dans presque tous les pays étudiés, n'est : soit pas consultable sous forme granulaire (Singapour) ; soit résumée à quelques enquêtes de satisfaction (Canada). Seul le Royaume-Uni dispose de plusieurs bases de données sur la qualité. Une d'entre elles établit un score sur 148 critères répartis dans quatre domaines (clinique, organisation, expérience patient, qualité de prise en charge).

Ce tableau met en évidence également la difficulté d'obtenir des données équivalentes selon le mode d'exercice du professionnel de santé : en ville, à l'hôpital, public ou privé. Or, seul un objectif d'un traitement identique pourrait garantir la démocratie sanitaire, sans discrimination entre les modes d'exercice et activités.

La consultation des données sur la qualité permet d'apprécier sa vertu pédagogique, dès l'instant où les professionnels de santé se les sont appropriées, car elles sont le reflet de leur propre exercice.

En conséquence, les représentants des Ordres des professions de santé estiment que l'Open data des données sur les professionnels, y compris les données qualité, ne peut que motiver leur adhésion à un processus d'amélioration de la qualité de leurs pratiques professionnelles au bénéfice de la santé publique, sous réserve qu'ils contribuent eux-mêmes de façon prééminente à la définition de ces critères et que ceux-ci ne soient pas établis de façon univoque.

2.5 Développer la recherche

Un matériel essentiel de la recherche en santé est l'exploitation de données de santé et de données susceptibles d'avoir un lien avec la santé. C'est pourquoi les chercheurs ont besoin de pouvoir utiliser des données de natures très variées : clinique, génétique, biologique, environnementale, professionnelle, sociale, de consommation, économique, de parcours de soin et de consommation de soin.

La recherche en santé est productrice et utilisatrice de données populationnelles. Par le biais de publications scientifiques, mais également de collaborations nombreuses, la communauté des chercheurs est exercée à la mutualisation et aux échanges de méthodes permettant une utilisation optimale des bases de données.

La perspective d'utiliser plus largement des données médico-administratives dans un objectif de production de connaissances utiles à la santé publique s'est amplifiée récemment. Elle provient essentiellement de l'amélioration, de la centralisation et du chaînage des systèmes d'informations de l'assurance maladie (SNIIRAM) et de l'administration hospitalière (PMSI). La fiabilisation de ces données et leur ouverture progressive aux chercheurs a généré de très nombreuses perspectives de recherches, dont les retombées pour la santé publique sont importantes. La communauté des chercheurs en santé y a ainsi identifié des opportunités évidentes de gain de compétitivité en matière de production scientifique.

Avec les progrès rapides des méthodes de séquençage et d'autres instruments d'analyse biologique, ces perspectives s'étendent à la recherche biomédicale dans son ensemble (comme le confirment des expériences étrangères, notamment au Danemark). Le couplage entre les données de recherche biomédicale et les bases médico-administratives, voire avec d'autres sources de données statistiques populationnelles, constitue un élément stratégique clé de la compétitivité internationale à moyen terme de la recherche française.

2.6 Encourager l'innovation et le développement industriel

A l'évidence, ces deux activités sont de plus en plus consommatrices de données.

L'innovation dans le champ des outils et services numériques s'appuie sur l'existence de données dont elle trouve des usages nouveaux (innovations d'usage). Mettre à disposition des innovateurs un maximum de données « brutes » anonymes et sans limite d'usage constitue un *booster* de l'innovation. Le domaine de la santé constitue de ce point de vue une priorité affirmée par tous les rapports récents de prospective économique et industrielle.

Grâce à l'open data, un écosystème organisé et régulé peut se créer, source de valeur pour la société (emplois de service, qualité de vie, bien-être) et d'emplois pour la filière numérique.

A plus grande échelle encore, toutes les branches économiques, à commencer par les industries de santé, ont aussi besoin de données. Le développement de nouveaux médicaments ou dispositifs passe par des étapes d'évaluation irréalisables sans accès à des données de santé. Mettre à disposition des industriels des bases de données riches et structurées constitue pour notre pays un enjeu essentiel pour maintenir son attractivité vis-à-vis d'industries à la recherche permanente du territoire le plus propice à leur développement. Comme l'a exprimé à plusieurs reprises le Conseil Stratégique des Industries de Santé, la collaboration entre les industries et la recherche publique dans l'analyse des grandes bases de données et leur appariement avec des données générées par la recherche biologique, clinique et de santé publique constitue une priorité pour le développement de partenariats public/privé indispensables au développement de l'innovation biomédicale.

Partie 3

Attentes et éléments de doctrine

Les développements qui précèdent démontrent qu'il est de l'intérêt général de favoriser une plus grande fluidité de la circulation des données de santé entre acteurs du système de santé, avec les citoyens et patients, dans le respect de la vie privée et des différents secrets légaux (statistique, industriel et commercial, médical). Ainsi, la commission recommande qu'un principe d'ouverture par défaut des données anonymes, sauf exception motivée, soit appliqué. Forte de cette conviction, la commission considère qu'il convient de donner accès sans restriction (open data) aux données de santé, y compris la possibilité de les réutiliser, dès lors qu'elles ne sont pas porteuses de risque de réidentification. Le risque de mésusage n'est pas considéré comme un motif de non publication des données – mais il conviendrait d'inciter les utilisateurs à rendre compte des méthodes utilisées pour traiter les données, règle qui a cours lorsque ces données sont utilisées dans la publication de travaux scientifiques. Une action soutenue doit être conduite en conséquence pour rendre accessibles des ensembles de plus en plus nombreux de données de santé, y compris par transformation de données considérées à l'origine comme indirectement nominatives.

La préservation du secret des données personnelles de santé justifie en revanche des précautions pour l'accès aux données de santé dès lors qu'elles sont porteuses de risques de réidentification des patients. Il ne s'agit pas de fermer l'accès aux données mais de l'organiser en se fondant sur une évaluation objective et explicite des risques de réidentification mis en regard des bénéfices attendus.

Des restrictions d'accès aux données nominatives des professionnels de santé se justifient pour des raisons de protection de la vie privée. En outre, les exigences du secret industriel et commercial peuvent induire des précautions particulières en matière de publication de données relatives aux professionnels et aux établissements de santé. Là encore, il ne s'agit pas de fermer l'accès aux données, mais de l'organiser, dans le respect de ces contraintes, pour répondre à la demande forte d'information de la part des citoyens.

Afin de favoriser la mise en œuvre dans la durée des orientations énoncées ci-dessus et d'atteindre les objectifs cités en Partie 2 relatifs à l'innovation, la recherche, l'efficacité de l'action publique et l'autonomisation des patients, la Commission recommande la mise en place d'une fonction d'administrateur des données de santé. Ce dernier aurait pour mandat de concourir, dans les domaines de la santé et du médico-social, aux missions de l'administrateur général des données telles que définies dans l'annonce en Conseil des Ministres du 21 mai 2014 :

- Organiser une meilleure circulation des données au sein de l'administration ;
- Lancer des expérimentations pour éclairer la décision publique en matière de politique ;
- Diffuser les outils, méthodes et culture de la donnée au sein des administrations, et au service de leurs objectifs respectifs.

Cet administrateur des données de santé pourrait rendre compte périodiquement de son action à l'instance d'expression des parties prenantes présentée infra au 3.5.2.

Les orientations qui sont proposées ci-dessous nécessiteront, dans certains cas, des modifications législatives et/ou réglementaires.

3.1 L'accès aux données anonymes ou rendues anonymes

3.1.1 Doctrine de l'open data en santé

L'ouverture des données de santé anonymes vise quatre objectifs stratégiques : contribuer à la démocratie sanitaire, permettre l'autonomisation du patient, renforcer l'efficacité de l'action publique et stimuler la croissance et l'innovation.

Pour atteindre ces objectifs, l'ouverture des données anonymes de santé au public doit suivre les principes généraux suivants, en cohérence avec la loi CADA, la directive européenne *Public Sector Information* et la politique du gouvernement rappelée dans la première partie du rapport :

- L'ouverture de toutes les données anonymes relatives à la santé et au système de santé français doit être organisée, sans préjuger de l'utilisation qui sera faite de ces données ;
- La qualité des données, leur complétude ou leur complexité ne constituent pas des critères devant déterminer l'ouverture ou non des données de santé anonymes ;
- L'ouverture des données granulaires doit être privilégiée. Si les données granulaires ne sont pas anonymes, ou sont protégées par un secret prévu par la loi (ex : secret industriel et commercial, secret statistique), l'ouverture des données agrégées doit être organisée (dans la mesure des capacités des organisations à produire ces agrégats s'ils n'existent pas déjà) ;
- Les enquêtes et recherches menées sur fonds publics devront dès leur conception prévoir l'ouverture des données granulaires ou anonymisées.

3.1.2 Attentes sur les données anonymes à ouvrir

Cette partie présente la liste des données anonymes dont l'ouverture en open data a été discutée par la Commission en réunion plénière et lors d'ateliers thématiques.

En parallèle, pour enrichir le présent document, Etalab a également organisé le recueil de demandes émanant de contributions citoyennes. Les données demandées dans ce cadre sont présentées en annexe.

La Commission demande dans les plus brefs délais l'ouverture des données suivantes en open data, c'est-à-dire gratuitement, en format ouvert à tous et réutilisable.

- Tableaux de bord de pilotage pour le médico-social (ANAP) ;
- Tableaux de bord du SNATIH : tableaux d'analyse de l'activité, d'analyse des recettes et de l'offre de soin(ATIH) ;
- Données en volume de consommation des médicaments et DMI en SUS dans les hôpitaux (ATIH) ;
- Base de l'outil national d'inventaire et de pilotage du patrimoine hospitalier OPHELIE (ATIH et DGOS) ;
- Enquête sur les événements indésirables liés aux soins (DREES) ;

- Thésaurus des interactions médicamenteuses (ANSM) ;
- Base E-Fides : déclaration des intérêts des professionnels de santé couverts par le Sunshine Act (ANSM) ;
- Répertoire des établissements pharmaceutiques (ANSM) ;
- Répertoire des dispositifs médicaux (ANSM) ;
- Répertoire des spécialités (ANSM et DGS) ;
- Base sur la qualité des eaux de baignade (DGS) ;
- Répertoire des essais cliniques des médicaments (ANSM) ;
- Base de données des résultats de certification des établissements de santé (HAS) ;
- Base QUALHAS pour les indicateurs désignés dans l'arrêté du 5 mars 2013 (HAS) ;
- Les résultats d'accréditation des médecins (HAS) ;
- Enquête ESPS (IRDES) ;
- Baromètre Santé (INPES) ;
- Base de l'observatoire des services d'information en santé, données agrégées (ATIH/DGOS) ;
- Base de l'observatoire de recomposition (ATIH/DGOS) ;
- Base CERVEAU de suivi de l'activité des urgences en Ile-de-France (ARSIF)³² ;
- Extraits de la base ARHGOS de suivi des autorisations d'activité (ARSIF) : données concernant les arrêtés d'autorisation relatifs aux activités de soins ;
- Base Allocation de Parent Isolé (ARSIF) ;
- Données transmises par le CNOP en vertu de l'article 23 de la loi n°2011-2012 (« données anonymes relatives aux médicaments qui sont hébergées dans le cadre du dossier pharmaceutique ») ;
- Données sur la qualité des eaux de piscine (ARS) ;
- Base SISEAUX sur la qualité des eaux de robinet et embouteillées (ARS).

En outre, **la Commission demande l'ouverture en open data des bases et jeux de données listés dans le tableau ci-dessous, tout en reconnaissant que des actions préalables, également décrites ci-dessous, doivent être mises en œuvre, justifiant quelques délais.**

Quatre types d'actions préalables sont identifiés :

- Produire de manière automatisée une extraction à façon de la base initiale, afin d'éviter un risque de ré-identification ou d'éviter de révéler des informations portant atteinte au secret industriel et commercial.
- Vérifier que la base ne comporte pas de risque de ré-identification (ex : DAMIR de la CNAMTS).
- Mener une concertation avec les professionnels de santé producteurs de la donnée afin de définir les modalités d'ouverture en open data (exemples : données financières sur les honoraires, répertoires et référentiels nominatifs pour lesquels l'accord préalable des professionnels est nécessaire).
- Modifier des dispositions légales, réglementaires ou contractuelles contraignant l'ouverture des données (exemples : référentiels nominatifs à publication obligatoire pour lesquels un droit de réutilisation doit être spécifié ; harmonisation des obligations légales en matière de partage des données financières entre établissements de santé publics et privés ; révision de contrats ou conventions restreignant le partage d'informations comme pour les tableaux de bord de l'Institut des données de santé).

³² L'ARS Ile-de-France s'assurera au préalable que les conditions d'anonymisation de cette base sont compatibles avec les recommandations du groupe de travail sur la ré-identification.

Gestionnaire	Donnée	Actions préalables
ANSM	Base de pharmacovigilance	Extraction de la base excluant les champs libres où des informations à caractère personnel ou ré-identifiantes peuvent figurer.
ANSM	Base « déclaration de vente des médicaments » (base « TAXE »)	Extraction de la base avec agrégation afin d'éviter la révélation d'informations protégées par le secret industriel et commercial
Toutes ARS	Données sur le temps d'attente aux urgences	Extraction à réaliser à partir des divers outils informatiques utilisés par les ARS, dont certains sont encore en phase d'expérimentation
Toutes ARS	Données sur les lits disponibles par spécialité en établissement	Extraction à réaliser à partir des divers outils informatiques utilisés par les ARS, à accompagner d'une présentation de la méthodologie utilisée.
ARSIF	Contrats pluriannuels d'objectifs et de moyens (CPOM)	Extraction à définir pour mettre en ligne les indicateurs, contenus et évaluations réalisés dans le cadre des CPOM <i>Rappel : les modèles de CPOM sont disponibles sur le site du ministère de la santé et les éléments sur les financements alloués font l'objet d'arrêtés publiés</i>
ARSIF	Données des Répertoires Opérationnel des Ressources	Ouverture limitée dans un premier temps aux éléments relatifs au PAERPA et Territoire de soins numériques pour les ARS volontaires. Mener une concertation avec les professionnels de santé pour une publication complète à moyen terme.
ATIH	Base ANCRE	Extraction de données agrégées dans un premier temps. Action recommandée : compléter les obligations réglementaires des établissements publics et privés (ESPIC) en matière de communication des comptes financiers.
ATIH	Données sur les prix des médicaments et DMI en sus dans les hôpitaux (ATIH)	Extraction de données agrégées afin d'éviter la révélation d'informations protégées par le secret industriel et commercial
ATIH	DIPI : données sur le suivi des dossiers d'instruction des projets d'investissement immobilier	
ATIH	DIPISI : Dossier d'instruction des projets d'investissement en système d'information dans le cadre du projet Hôpital Numérique	
ATIH	Tableaux de bord du SNATIH : performance des établissements	- A court terme, tenter l'ouverture des données avec une option de retrait sur demande expresse des établissements
ATIH	Tableaux de bord du SNATIH : données de charges par établissement	- A moyen terme, renforcer les obligations réglementaires des établissements en matière de diffusion des données financières
ATIH	Tableau de bord du SNATIH : analyse des coûts	

HAS	Base QUALHAS	Mener une concertation avec les établissements de santé concernant l'ouverture des indicateurs non inclus dans l'arrêté du 5 mars 2013
HAS	Base de certification des établissements de santé	Mener une concertation avec les professionnels et établissements de santé concernés
HAS	Base d'accréditation des médecins	
IRDES	Base HYGIE	Extraction de données agrégées à réaliser pour éviter tout risque de ré-identification
CNAMTS	DAMIR Suivi des dépenses	Vérifier s'il n'y a aucun risque de ré-identification
CNAMTS	Base Etablissement Référentiel (BERF)	Mener une réflexion sur l'ouverture des référentiels comprenant des données nominatives (adresse des cabinets, diplôme des médecins,...)
CNAMTS	Indicateurs ROSP des médecins	Mener une concertation avec les professionnels de santé sur la publication des ROSP
IDS	Tableau de bord de l'IDS	<ul style="list-style-type: none"> - Confirmer que l'article L.161-36-5 du Code de la Sécurité Sociale ne constitue pas une contrainte au partage et à la réutilisation des tableaux. - Mener une analyse de faisabilité technique - Obtenir l'accord des membres de l'IDS
ASIP Santé	Répertoire Partagé des professionnels de santé (RPPS)	Mener une réflexion sur la publication des données nominatives (adresse des cabinets, diplôme des médecins,...)
ASIP Santé	Référentiel des Acteurs Santé Sociaux	
Centre National Hospitalier d'information sur le médicament	Base Thériaque	Revoir le principe actuel selon lequel la base est consultable, mais le téléchargement et la réutilisation des données sont des prestations payantes
DGS	Base de données relatives à la transparence des avantages accordés aux professionnels de santé	Modifier le Décret n° 2013-414 du 21 mai 2013 qui ne prévoit pas la possibilité de réutilisation des données issues de cette base.

3.1.3 Recommandations pour pérenniser le mouvement de publication en open data des données en santé

Les réunions plénières de la Commission, et plus encore les rencontres et ateliers thématiques organisés en parallèle, ont permis de lancer un mouvement dynamique autour de l'open data en santé, concrétisé par l'ouverture des données suivantes :

- Indicateurs Hospi Diag (ANAP) ;
- Base BIOLAM 2010-2012 (CNAMTS) ;
- Base Medic'AM 2008-2013 (CNAMTS) ;
- Tableau sur la démographie et le conventionnement des professionnels de santé en 2012 (CNAMTS) ;

- Dénombrement des bénéficiaires de la CMU en 2012 (CNAMTS) ;
- Dépenses annuelles d'assurance maladie en 2012 et 2013 (CNAMTS) ;
- Produits de santé remboursés par l'assurance maladie en 2006-2012 (CNAMTS) ;
- Activités des médecins (actes) par région en 2010 (CNAMTS).

De plus, l'ATIH a partagé un échéancier pour l'ouverture de données supplémentaires, qui définit et priorise les actions préliminaires à engager par les services des organisations concernées :

Données	Date d'ouverture	Action(s) préalable(s)
Statistiques du SNATIH : statistiques issues des bases nationales de données PMSI MCO, de la base PMSI-SSR, de l'agrégation des RAPSS (HAD) et de l'agrégation des RPSA et R3A (PSY)	Fin avril 2014	A référencer sur data.gouv.fr
Tableaux de bord du SNATIH : tableaux d'analyse de l'activité, d'analyse des recettes et de l'offre de soin	Septembre 2014	Automatiser l'extraction pour une mise à jour à fréquence régulière
Données en volume de consommation des médicaments et DMI en SUS dans les hôpitaux	Septembre 2014	Automatiser l'extraction pour une mise à jour à fréquence régulière

Afin de pérenniser cette dynamique positive, et de répondre aux attentes exprimées par les membres de la Commission, celle-ci souhaite que des actions soient menés dans les domaines suivants.

Sur les données médico-administratives et en particulier les données financières :

La cartographie des données montre que la France dispose d'une exceptionnelle richesse d'informations médico-administratives, actuellement inaccessibles aux citoyens à quelques exceptions près – principalement des tableaux fortement agrégés mis à disposition par l'Assurance Maladie et l'ATIH. Une dynamique positive est lancée avec l'ouverture des données Hospi Diag par l'ANAP. Pour pousser plus loin cet effort de transparence, un groupe de travail réunissant notamment les professionnels, établissements de santé publics et privés et les représentants des patients et usagers du système de santé, pourrait être constitué à court-terme (et s'inscrire ensuite dans le cadre de la nouvelle gouvernance proposée en fin de rapport) pour élaborer « une offre open data » ambitieuse dépassant les freins actuels n'ayant pas de justification légale.

Sur les données relatives à la qualité et la sécurité des soins :

La cartographie des données de santé indique que la France dispose sur ce thème de données nettement moins complètes, et surtout moins ouvertes, que les pays anglo-saxons. Tout en reconnaissant la différence entre notre modèle et celui des Etats-Unis ou du Royaume-Uni, la Commission recommande l'ouverture en open data des indicateurs existants, générés par les établissements de santé et certaines Agences Régionales de Santé pour suivre les processus

qualité/sécurité engagés ; en particulier les indicateurs de résultats, plus importants encore du point de vue des patients que les indicateurs de processus.

A titre d'exemple, la Commission recommande l'ouverture en open data de la base QUALHAS de la HAS, ainsi que l'ouverture en open data d'indicateurs de suivi, tels que les cas déclarés de maladies nosocomiales. La Commission encourage aussi la finalisation de la base I-Satis, créée par la DGOS et gérée par l'ATIH, contenant des indicateurs de perception des soins délivrés par les professionnels de santé.

Sur les référentiels des professionnels de santé :

Les données de référentiels sont fondamentales dans toute démarche d'open data. Elles constituent le socle à partir duquel l'exploitation de nombreuses autres données devient possible. Elles sont essentielles pour réaliser l'ambition exposée dans la partie 2 de ce rapport. Les rencontres et ateliers organisés en marge de la Commission ont montré qu'il existe une multiplicité de référentiels des professionnels et établissements de santé. La Commission considère que le besoin de les harmoniser, reconnu par de nombreux acteurs, ne peut constituer un frein à leur ouverture en open data. Elle recommande au contraire leur ouverture la plus exhaustive comme un moyen d'accélérer ce travail d'harmonisation.

Si la publication de certains référentiels nominatifs est requise par la loi, ces données ne sont pas réutilisables et sont donc actuellement hors périmètre de la démarche de l'open data. La Commission recommande, comme le faisait le rapport Bras, la levée des freins légaux, réglementaires et techniques à la mise en open data des référentiels. A cette fin, la Commission recommande qu'un groupe de travail associant notamment le Ministère de la santé, les représentants des professionnels et établissements de santé et la CNAMTS, formule des propositions concrètes. Il s'agirait par exemple d'élargir le champ des données du répertoire partagé des professionnels de santé (RPPS) disponibles en open data (exemples : inclure les adresses des cabinets médicaux, les diplômes, un indicateur du nombre d'années de pratique), d'ouvrir les Répertoires Opérationnels des Ressources des ARS, ainsi que les accréditations des professionnels de santé. L'ouverture progressive des données de la rémunération sur objectifs de santé publique (ROSP) pourraient également être discutée dans ce groupe de travail.

Sur les données de recherche et les données d'enquête :

La cartographie des données et les échanges en marge de la Commission ont révélé la nécessité de mener une réflexion spécifique pour renforcer l'ouverture en open data de données de recherche et données d'enquête dans le domaine de la santé. Il s'agit en effet de :

- Clarifier quelles données d'enquête et de recherche qualifient comme des données administratives au sens de la loi « CADA » ;
- Partant des recommandations du groupe de travail sur la ré-identification, proposer des solutions concrètes et standardisées pour éviter les risques de ré-identification et respecter le secret statistique afin de permettre l'ouverture de données de recherche et d'enquête proches du niveau granulaire ;
- Définir une méthode permettant de standardiser la quantification de l'effort à fournir pour une ouverture en open data : actions à réaliser pour supprimer le risque de ré-identification, respecter le secret statistique, redresser les données, générer les méta-

données... ; il s'agira de trouver des solutions permettant de limiter le coût additionnel supporté par les équipes de recherche pour ouvrir leurs données ;

- Proposer des solutions pour valoriser l'investissement des chercheurs pour ouvrir les données issues de leur travail en open data (comme le programme européen « SHARE »).

La Commission recommande qu'une réflexion soit lancée sur cette question. Etant donné qu'elle se pose également dans les autres catégories de données publiques (ex : transport, logement...), une telle réflexion pourrait impliquer les principaux acteurs publics et privés de la recherche et de la statistique.

La recherche en santé publique est favorable à l'ouverture des données rendues anonymes produites par les entités dont elle est responsable. Il faut toutefois rappeler qu'une partie des financements obtenus par la recherche se fait à travers la monétisation de la transmission de données produites. Elle permet souvent de participer à la maintenance d'un système d'information. De plus, certains organismes publics finançant la recherche en santé, comme ceux du Plan d'Investissement d'Avenir, demandent que les projets s'autofinancent, ce que l'ouverture des données de recherche peut rendre plus difficile encore. Une ouverture des données qui exigerait la gratuité doit donc être accompagnée d'une réflexion sur les sources alternatives de financements pour le producteur.

Sur l'extension du réseau de correspondants Open Data aux établissements publics et opérateurs sous tutelle du ministère de la santé :

La Commission reconnaît que la mise en œuvre de ces recommandations demandera, de la part de certaines institutions gestionnaires des données concernées, un effort initial de préparation des données en vue de leur ouverture. Qu'il s'agisse d'un travail de mise en format ouvert de données déjà rendues publiques, ou d'automatisation de nouvelles extractions anonymes ou agrégées pour respecter les bornes légales à l'ouverture des données publiques, cet effort pourra se traduire par une augmentation ponctuelle de travail pour les équipes techniques concernées.

Pour accompagner les administrations dans l'ouverture de leurs données et promouvoir l'open data, la mission Etalab travaille avec un réseau de 13 coordinateurs auprès des Secrétariats généraux des ministères, et près de 250 correspondants Open Data dans les principaux services centraux. Les établissements publics sous tutelle ainsi que les opérateurs sont initialement coordonnés par les correspondants Open data rattachés au Secrétariat Général du ministère de tutelle – en l'occurrence celui de la Santé. Afin d'accompagner les établissements publics et opérateurs sous tutelle du ministère de la santé dans l'open data, la Commission formule les recommandations suivantes :

- La direction générale de chacune de ces structures se charge élaborer une feuille de route de l'ouverture de ses données, comprenant un échéancier d'ouverture des données en open data qui précise les actions préalables à l'ouverture et la fréquence de mise à jour ;
- Les directions du Ministère de la santé sont chargées de répondre aux interrogations des établissements publics et opérateurs sur lesquels elles exercent une tutelle concernant les feuilles de route proposées, et s'assurent de la production effective d'une feuille de route par structure ;

- Chaque établissement public et opérateur sous tutelle nomme un correspondant Open Data avec lequel le coordinateur Open Data du Ministère pourra travailler sur les aspects techniques devant permettre la mise en œuvre de la feuille de route ; ce correspondant sera également responsable de la mise à jour de la cartographie des données de santé pour son institution, sous le pilotage du coordinateur Open Data du Ministère.

De façon générale, la Commission recommande que le coordinateur Open data du Ministère accompagne en tant que de besoin les établissements publics et opérateurs sous tutelle dans le développement de leur stratégie d'ouverture des données publiques, et continue à animer les groupes de travail sur l'open data santé tel que les 3 ateliers qui ont été organisés dans le cadre de la commission (voir en annexe 8 les compte rendus des ateliers). Dans ce sens, la commission recommande que le mandat et le rôle du correspondant open data du secrétariat général soit renforcé et que des dispositions soient prises pour que celui-ci puisse consacrer plus de temps à l'ouverture des données publiques de santé.

La Commission recommande en outre que le principe de l'open data soit inscrit en amont de tout nouveau projet financé sur fonds publics dans le domaine de la santé. Ainsi, par exemple, l'Institut Cancer (INCa) a déjà prévu la publication en open data des tableaux de bord issus de la cohorte cancer. Il est ainsi fortement recommandé que la possibilité de télécharger les données granulaires soit prévue dans l'élaboration de nouveaux outils tels que le Répertoire Opérationnel des Ressources ou de nouveaux sites internet, tel que celui de la CNSA. Ce principe est ainsi recommandé à l'ANSM, qui travaille actuellement à la refonte de la base de pharmacovigilance.

3.1.4 Actions en cours et à poursuivre pour permettre l'ouverture en open data de données supplémentaires à partir du SNIIRAM-PMSI

La manière la plus simple de mettre à disposition des données anonymes issues de bases indirectement nominatives est évidemment de produire des tableaux statistiques répondant par des données agrégées aux questions que se posent les différentes catégories d'utilisateurs. La limite de l'exercice est que la demande est potentiellement infinie, alors que la production de statistiques agrégées est coûteuse et répond souvent mal ou trop lentement aux besoins particuliers des uns et des autres, notamment aux besoins des chercheurs ou à des questions d'une grande technicité ou exigeant le recours à plusieurs sources de données.

Une solution est de rendre plus « agile » la production de ces tableaux statistiques à l'aide d'outils techniques, par exemple en mettant à la disposition des utilisateurs des bibliothèques de requêtes en libre service ; à la limite cela revient à laisser interroger la base à l'aide d'outils logiciels paramétrables fournissant des réponses anonymes sans que l'auteur de la requête ait lui-même accès aux données. Le même résultat peut être obtenu, plus classiquement, en organisant des collaborations (un « guichet ») entre une équipe d'experts attachés à la base de données et les experts « métier » des demandeurs, pour fabriquer les sorties agrégées adaptées aux besoins, sachant que la mobilisation d'agents pour des travaux à façon a un coût et que là aussi on ne répond pas toujours aux besoins les plus pointus.

L'autre solution explorée par le Groupe de travail « risques de réidentification » (dont le rapport détaillé peut être trouvé en annexe 9) consiste en la production d'un grand nombre de « jeux » de données rendus anonymes, tels que pour chacun de ces jeux, le nombre de personnes présentant

des caractéristiques identiques soit suffisant (10 au moins dans le test) et que parmi ces personnes il y en ait toujours un nombre suffisant (3 au moins dans le test) qui aient des maladies différentes. On a démontré que c'est possible et que les critères d'anonymat peuvent être vérifiés. Des précautions doivent être prises cependant pour qu'il ne soit pas possible de reconstituer ainsi des jeux de données identifiants. Une sécurité supplémentaire peut être obtenue en utilisant des techniques d'échantillonnage, si la règle de tirage est secrète.

3.2 L'enrichissement à poursuivre du SNIIRAM et du PMSI

Les données du SNIIRAM et du PMSI sont issues des séjours hospitaliers et des remboursements de l'assurance maladie. L'appariement de ces deux bases, depuis 2009, a déjà permis de faire considérablement progresser la capacité d'études sur les parcours des patients. L'expérimentation récente d'appariement avec les causes de décès de l'INSERM constitue un pas supplémentaire.

Néanmoins, les seules informations non médicales relatives aux patients sont leur âge, leur sexe et leur adresse, ainsi que leur statut de bénéficiaire ou non de la CMUC et de l'aide à la complémentaire santé (ACS).

La commission estime qu'il est nécessaire, pour développer les études et la recherche et approfondir en particulier la connaissance des mécanismes au fondement des inégalités sociales de santé, que soit accru le périmètre descriptif des patients et que soient accessibles, soit directement dans le SNIIRAM, soit par appariement avec des bases de données sociales ou fiscales, des variables socio-économiques (revenu, catégorie socioprofessionnelle, diplôme, niveau d'instruction...).

Au-delà de cette attente forte, d'autres types de données actuellement absentes du SNIIRAM pourraient à terme avoir vocation à s'y trouver : diagnostics en ville, résultats d'analyses médicales ou d'imagerie, informations médico-sociales (données des maisons départementales des personnes handicapées et des conseils généraux)...

La commission est naturellement consciente qu'un équilibre entre l'apport que représentent ces données pour la recherche et le coût de leur recueil (notamment pour ceux qui saisissent l'information à l'entrée, le cas échéant) doit être trouvé.

3.3 Les données des patients porteuses de risque de réidentification

Il convient de rappeler en préambule que les besoins de recours à des données porteuses de risques de réidentification sera d'autant plus réduit qu'une action efficace aura été menée pour accroître le champ des données anonymes à partir de retraitements des données décrits plus haut (micro-agrégation, floutage, etc....).

Pour autant, il restera des situations justifiant le recours aux données les plus fines et les plus complètes comportant dès lors un risque de réidentification. Dans ce cas, le principe posé est celui d'un accès régulé et contrôlé protégeant effectivement le droit constitutionnel à la vie privée. Pour l'essentiel, dans le cadre des travaux de la commission, il s'agit d'organiser les accès au SNIIRAM-PMSI qui constitue aujourd'hui la principale base de données à laquelle les accès doivent être facilités.

La communauté des chercheurs est particulièrement intéressée par la possibilité d'avoir accès à des données porteuses de risque de ré-identification dans la mesure où, par définition, ces données individuelles permettent d'effectuer les traitements statistiques les plus précis et les plus puissants.

Il ressort des travaux menés sur l'analyse du risque de ré-identification qu'au-delà des données clairement anonymes, il existe une zone grise pour laquelle il est d'autant moins possible de se prononcer sur l'absence de risque de ré-identification que l'on ne connaît pas, par exemple, les autres données disponibles entre les mains des tiers.

Le constat de cette situation ne doit pas conduire à fermer l'accès aux données, mais à en organiser l'accès de telle sorte que les risques soient évalués et mis en regard des bénéfices attendus. La commission souligne que l'analyse de risque doit être objectivée (par exemple en termes de matrice probabilité/conséquence) et que les restrictions ou les contraintes apportées dans l'accès aux données doivent être dimensionnées en conséquence. Cinq principes doivent donc gouverner l'accès aux données de santé relatives aux patients :

- L'objectivation et l'évaluation des risques permettant de graduer les exigences de sécurité;
- La prise de mesures proportionnées de gestion des risques à travers la mise en œuvre de conditions d'accès et de régulation des personnes habilitées à accéder aux données;
- La traçabilité des accès et des requêtes ;
- L'évaluation des bénéfices collectifs attendus ;
- L'effectivité des sanctions en cas d'action illicite.

Il convient de souligner que ni l'évaluation des risques, ni l'évaluation des bénéfices attendus ne constituent des données figées. L'évaluation du rapport bénéfice/risque comporte une dimension qualitative dont l'appréciation peut varier dans le temps mais aussi selon les acteurs. Toutefois les membres de la Commission partagent une conception selon laquelle l'utilité économique ne saurait justifier la divulgation de données personnelles de santé détenues par les administrations.

Au-delà des exigences liées à l'instruction propre à chaque demande d'accès, une obligation générale de transparence sur les objectifs, les protocoles, les méthodologies d'analyse des données et les résultats obtenus doit servir de base à une publication annuelle des demandes d'accès et des réponses apportées.

La commission considère par conséquent important de placer le dispositif de régulation des accès sous une gouvernance qui permette aux parties prenantes de s'assurer que l'accès aux données de santé est bien mis en œuvre au service de tous les utilisateurs dans le respect des principes posés.

Dans ce cadre, l'organisation proposée repose sur trois éléments :

- La distinction entre des accès liés aux missions permanentes et des accès - temporaires ou permanents - liés à des projets, fondée sur l'analyse des missions justifiant pour les premiers de façon manifeste et impérieuse cet accès ;
- Un niveau opérationnel de mise à disposition de l'accès aux données et/ou de mise à disposition d'échantillons ;
- Le renforcement de l'effectivité des sanctions en cas d'accès ou d'usage illicite des données.

La mise en œuvre de cette organisation est décrite plus précisément plus bas dans le 3.5.

En outre, la commission a souhaité rappeler l'exigence de simplification et de lisibilité du futur dispositif d'accès aux données, y compris dans l'articulation de l'examen des demandes d'accès avec les comités de protection des personnes³³, notamment si ces derniers devaient faire l'objet d'une réforme.

3.4 Les données relatives aux professionnels et établissements de santé

3.4.1 Les professionnels de santé libéraux

Si les données nominatives relatives aux professionnels de santé n'ont pas la même sensibilité que celles des patients, il n'en demeure pas moins qu'il existe des limites précises à leur publication :

- La protection de la vie privée ;
- Le respect de l'article 13 de la loi du 17 juillet 1978 qui précise les conditions de réutilisation de données personnelles détenues par une administration : « soit lorsque la personne intéressée y a consenti, soit si l'autorité détentrice est en mesure de les rendre anonymes, ou à défaut d'anonymisation si une disposition législative ou réglementaire le permet » ;
- Le secret en matière commerciale et industrielle (art. 6 II de la loi du 17 juillet 1978), qui, selon la jurisprudence de la CADA, protège l'activité libérale des professionnels de santé.

Il ressort assez clairement des échanges au sein de la commission que les données relatives aux honoraires (avec la question du choix de l'indicateur), à la nature et au volume de l'activité, aux pratiques de prescription et de façon plus générale à la qualité de la pratique constituent les champs prioritaires de publication des données.

Le statut des données susceptibles de répondre à cette demande est variable, certaines données étant d'ores et déjà disponibles et publiées (par exemple, les honoraires), d'autres sont disponibles dans d'autres contextes et non publiées (par exemple, les indicateurs de la rémunération sur objectifs de santé publique – ROSP), d'autres enfin seraient à extraire des bases de données. Ces différents statuts suggèrent une démarche de publication progressive et par étapes.

En premier lieu, il conviendrait de permettre la libre réutilisation des données déjà publiées par la CNAMTS. Si le risque de mésusage ne saurait s'opposer à cette orientation, il est nécessaire en revanche d'essayer de le prévenir en prévoyant la publication des métadonnées nécessaires à la compréhension des données ainsi que des exemples de bons et mauvais usages.

La publication de données permettant l'identification de la nature de l'activité des professionnels de santé (type d'activité, nature des actes pratiqués...) constitue un deuxième objectif de nature à donner une information non contestable et utile aux patients.

³³ L'article L.1123-6 du code de la santé publique (introduit par la loi du 9 août 2004) précise qu'"avant de réaliser une recherche biomédicale sur l'être humain, le promoteur est tenu d'en soumettre le projet à l'avis de l'un des comités de protection des personnes compétents pour le lieu où l'investigateur ou, le cas échéant, l'investigateur coordonnateur, exerce son activité".

S'agissant de données relatives à la qualité des pratiques, il est nécessaire de permettre l'émergence consensuelle d'indicateurs porteurs de sens tant pour les professionnels que pour les patients. Pour ce faire, un lieu de concertation (par exemple au sein d'un groupe de travail ad hoc puis de l'instance d'expression des acteurs décrite en 3.5.2) devrait être mis en place.

Il est naturellement nécessaire que l'assurance maladie adresse aux professionnels de santé leurs données individuelles avant publication afin de leur permettre de rectifier d'éventuelles erreurs.

Dans les limites posées par la protection de la vie privée, la publication de ces données pourra requérir une disposition juridique permettant de déroger aux dispositions de la loi du 17 juillet 1978.

3.4.2 Les établissements de santé

Le champ des établissements de santé donne lieu depuis de nombreuses années à une abondante publication statistique.

Les actions de publication de données initiées par les journalistes puis mises en œuvre par la puissance publique ont permis au cours des 15 dernières années d'élargir le champ des données à des sujets plus sensibles sur la qualité et les coûts.

En dehors de la question relative au secret industriel et commercial, il n'existe pas de limite à la publication des données (sous réserve des secrets protégés par la loi). Au même titre que pour les professionnels de santé, les informations relatives à la qualité et au coût des soins (suppléments pour diverses prestations, ticket modérateur...) apparaissent prioritaires.

3.4.3 Les professionnels de santé exerçant au sein des établissements de santé

Les données relatives à l'activité et à la qualité des pratiques individuelles des professionnels de santé salariés constituent en revanche une zone largement inconnue des statistiques.

Si le fait que le patient est réputé se confier à un établissement et à une équipe peut justifier l'absence de données individuelles, cette raison ne s'applique plus dès lors que la décision du patient est orientée par le choix d'un professionnel, par exemple pour une intervention chirurgicale ou une consultation de spécialiste. La mise en place d'un recueil de données individualisées par praticien, au moins sur certains champs, constitue une première étape indispensable.

3.5 Un niveau de gouvernance qui assure le fonctionnement du dispositif

Rappelons que, comme cela a été indiqué dans la première partie, la gouvernance des données médico-administratives peut s'analyser en quatre fonctions :

- La décision stratégique et la réglementation ;
- L'expression des parties prenantes (évaluation, propositions, demandes, orientations) ;
- La décision de donner ou non accès aux données selon les cas, en application des lois ;
- La gestion technique des données et le service de leur mise disposition.

3.5.1 La décision stratégique et la réglementation

Cette fonction relève par nature des pouvoirs publics.

3.5.2 L'expression des parties prenantes

La seconde fonction est fondamentale pour assurer que les administrations et les gestionnaires des données sont à l'écoute des besoins des utilisateurs.

Si les principes posés précédemment en termes d'ouverture des données anonymes et d'évaluation du bénéfice/risque de l'accès aux données de santé ont un caractère pérenne, l'évaluation qui en sera faite concrètement a un caractère contingent. Les débats ne sauraient être clos au terme des travaux de la présente commission. Il convient donc de maintenir une instance dans laquelle la mise sous tension des producteurs de données ayant vocation à être mises en open data et la doctrine en matière d'accès aux données réidentifiantes et sa déclinaison opérationnelle sont régulièrement interrogées au regard des attentes des parties prenantes.

S'appuyant sur l'expérience de l'IDS qui a montré l'intérêt d'un lieu d'expression des parties prenantes, les membres de la Commission insistent sur le fait qu'une telle instance ne trouve sa justification qu'à la condition qu'elle soit dotée d'une véritable capacité à agir à la fois sur les orientations et sur la mise en œuvre des accès aux données de santé. A cette fin, il pourrait être proposé que les ministres chargés de la sécurité sociale et de la santé transmettent chaque année au Parlement un rapport sur la politique de mise en œuvre du dispositif, et que ce rapport soit précédé d'un avis de cette instance.

En outre, elle serait chargée de proposer aux ministres chargés de la sécurité sociale et de la santé, des choix d'orientation et de priorité dans les programmes d'amélioration des bases de données. A cet effet, elle pourrait s'appuyer sur un comité scientifique composé d'utilisateurs experts et de gestionnaires des données afin de l'éclairer sur les améliorations à apporter.

Elle serait rendue destinataire des rapports d'utilisation des données produits par les administrations publiques dotées d'accès permanents ainsi que des rapports d'instruction des demandes d'accès temporaires et des décisions de la Cnil.

Compte tenu des questions traitées, cette instance devrait rassembler l'ensemble des parties prenantes : professionnels et établissements de santé, organismes gestionnaires des bases de données, chercheurs, industriels, représentants des patients et usagers du système de santé, organismes d'assurance maladie complémentaire, Etat. Il est proposé, pour faciliter la lecture de la suite du présent rapport de dénommer cette instance, comité d'orientation.

3.5.3 La décision de donner ou non accès aux données

Faciliter l'accès effectif aux données de santé passe d'abord par l'identification des organismes qui pourraient bénéficier d'un accès permanent. Cette mesure de rationalisation, déjà en vigueur pour certains organismes, ne se justifie cependant que si les missions de l'organisme le nécessitent de façon manifeste et impérieuse. Dans les autres cas, il conviendra de se limiter à une autorisation d'accès spécifiquement liée à un projet, qui peut d'ailleurs avoir un caractère récurrent ou relever de démarches génériques.

Les travaux sur les risques de réidentification ont toutefois mis en évidence la possibilité de constituer des bases de données qui, sans être strictement anonymes, ne comporteraient qu'un très

faible risque de réidentification. Ces travaux permettent de proposer un troisième mode d'accès aux données.

3.5.3.1 Les accès liés aux missions permanentes

La nécessité a déjà conduit à permettre un accès direct à certains organismes publics dont les missions le justifiaient à l'évidence. Si ces accès ont fait consensus, s'agissant par exemple des agences sanitaires, les modalités de l'instruction et l'absence de règles claires ont entraîné des délais et des restrictions qui justifient d'adopter une procédure simplifiée.

Il est proposé de remplacer le cadre actuel qui distingue des fondements légaux et réglementaires variés en indiquant que les accès permanents des services de l'Etat et de ses opérateurs ou d'autres utilisateurs permanents des données sont décidés par le ministre chargé de la santé, après avis de la Cnil.

La décision d'autoriser tel ou tel organisme est fondée sur l'analyse de ses missions et sur des garanties apportées en termes de respect de la confidentialité, d'organisation des accès, et de respect d'un référentiel de sécurité. Au sein de chaque organisme, le directeur habilite nominativement les agents placés sous son autorité – et/ou des sous-traitants agissant selon ses instructions – dont les accès aux données sont enregistrés et soumis au contrôle de la Cnil.

Il est important que les organisations bénéficiant d'un droit d'accès permanent rendent compte périodiquement de l'utilisation des bases de données et des conditions d'organisation des accès. Comme ils bénéficient d'un accès plus large aux bases de données, il ne serait pas acceptable que leur mobilisation insuffisante prive la collectivité des bénéfices attendus.

3.5.3.2 Les accès liés à des projets

Si des progrès réels ont été réalisés dans la période récente en matière d'ouverture des données du SNIIRAM-PMSI, il est nécessaire de faciliter davantage l'accès à ces données, au bénéfice de la communauté des utilisateurs.

S'agissant de l'accès aux données porteuses de risques élevés de réidentification, le principe posé est celui d'un accès régulé et contrôlé. Pour l'essentiel, dans le cadre des travaux de la commission, il s'agit d'organiser les accès à la base complète SNIIRAM-PMSI – ou PMSI seul – qui constituent aujourd'hui les principales bases de données auxquelles les accès doivent être facilités.

Il est proposé de simplifier le cadre actuel d'instruction en créant un circuit unifié et simplifié de décision, confiant à la Cnil la délivrance des autorisations d'accès dit ponctuel³⁴ après avis d'une instance technique chargée d'évaluer :

- Le bénéfice collectif attendu, le caractère collectif du bénéfice s'appréciant au regard de la capacité de l'étude envisagée à produire un bénéfice qui dépasse les intérêts particuliers du demandeur ;
- La pertinence des traitements au regard de la finalité poursuivie et le besoin effectif d'accéder aux données ;

³⁴ Pour les projets de longue durée comme les cohortes, l'autorisation d'accès peut avoir un caractère permanent.

- Les garanties données par le demandeur³⁵ quant à ses compétences;
- La transparence du demandeur sur les méthodes employées ;
- L'engagement à adresser en retour les résultats obtenus et les données secondaires constituées dans la limite des règles de diffusion propres à l'open data (absence de réidentification possible...).

L'étape d'instruction technique pourrait être écartée lorsqu'une recherche est réalisée à la demande d'une autorité publique et qu'elle a validé le protocole d'étude, la Cnil restant naturellement le décisionnaire ultime. D'autres procédures simplifiées visant à rendre la procédure plus rapide pourront être mises en place sur le modèle déjà existant des méthodologies de référence, notamment pour l'accès à des ensembles de données à faible risque de réidentification. Seraient également concernées par cette procédure, les demandes d'industriels visant à alimenter un dossier type requis par l'Administration.

Par ailleurs, la Commission est favorable à la demande émise de longue date par les chercheurs, de disposer d'un cadre juridique rénové de l'emploi du NIR pour faciliter les appariements de données, notamment des données du SNIIRAM avec des données individuelles recueillies auprès des patients dans le cadre d'enquêtes ou de suivi de cohortes.

La création d'une étape d'instruction systématique par un comité technique vise à objectiver l'analyse des demandes d'accès en la plaçant sur le seul champ de l'expertise.

Dans ces conditions, il est proposé que ce comité soit exclusivement composé d'experts et de scientifiques désignés par les responsables d'autorités compétentes telles que l'INSERM, le HCSP, l'ANSM, la HAS, etc. Un travail d'instruction de la demande devra être mené en amont avec les gestionnaires des bases de données (CNAMTS et ATIH) ainsi qu'avec les représentants des agences (ANSM, InVS) disposant d'une expertise particulière sur le contenu des bases. S'agissant des membres du comité technique, la commission souligne la nécessité d'une gestion rigoureuse des situations éventuelles de conflits d'intérêt, ainsi que la nécessité d'une motivation précise des avis émis.

Les travaux de la commission ont toutefois conduit à observer que l'évaluation du bénéfice collectif ne pouvait pas relever du seul comité technique.

Il est habituel de considérer qu'en matière de recherche académique, le progrès des connaissances constitue en soi un bénéfice collectif et que la nature de ce progrès dépend de la qualité de la recherche. Dans cette hypothèse, l'évaluation que réalisera le comité technique portera essentiellement sur la pertinence des traitements au regard de la finalité poursuivie et le besoin effectif d'accéder aux données sans qu'il soit besoin de réinterroger les finalités, l'avis émis par l'instance technique étant ensuite directement transmis à la Cnil.

En revanche, en dehors du cas de la recherche académique motivée par le progrès des connaissances, la finalité qui justifie l'accès aux données doit être évaluée au regard du bénéfice collectif spécifique attendu. La commission considère que cette évaluation doit être éclairée par le

³⁵ Y compris les chercheurs issus de l'Union européenne et des pays présentant un niveau de protection adéquat des droits des personnes.

comité technique mais qu'il doit appartenir à l'instance regroupant l'ensemble des parties prenantes (le comité d'orientation) de la réaliser, celle-ci étant la plus à même de prendre en compte les enjeux collectifs et sociétaux de l'étude nécessitant l'accès aux données. En pratique, la commission propose que le président du comité d'orientation ou son représentant assiste aux travaux du comité technique avec la possibilité de demander le transfert de l'avis à émettre en matière de bénéfice collectif au comité d'orientation. L'avis du comité d'orientation (incluant les avis motivés de ses membres) serait ensuite transmis à la Cnil accompagné de l'avis du comité technique pour ce qui concerne les autres critères d'évaluation.

Compte tenu des liens qui devront exister entre le comité technique et le comité d'orientation, la commission considère qu'ils devront être fonctionnellement articulés autour d'un guichet unique, constitué *a minima* d'un secrétariat commun aux deux instances et de préférence situé au sein d'une structure commune ; compte tenu de l'expérience acquise, l'Institut des données de santé, sous réserve d'en revoir notamment sa composition, pourrait être la structure d'accueil de ce dispositif.

3.5.3.3 L'accès aux données à très faible risque de ré-identification

S'agissant des données à faible risque d'identification, la proposition faite est de constituer différents jeux de données, à partir du SNIIRAM et du PMSI, visant à répondre aux besoins exprimés par les demandeurs d'accès aux données. A ce titre, il est proposé d'étudier la constitution d'un échantillon de données de grande taille tel qu'on ne puisse y reconnaître personne de manière certaine et dont l'accès soit ouvert à des organismes publics ou privés dont les missions et les compétences le justifient. Le principe de l'accès leur étant acquis, une limitation s'impose cependant en raison d'un risque résiduel tenant au caractère très particulier de certains parcours de soins³⁶ ; il conviendra qu'ils souscrivent un certain nombre d'engagements sur le respect des règles d'accès, d'usage des données et de publication des résultats.

Il appartiendrait à la Cnil après avis du comité technique décrit au paragraphe précédent de valider le niveau de risque faible. Dès lors, l'autorisation d'accès à tel ou tel organisme pourrait relever d'un arrêté ministériel, après avis du même comité technique pour ce qui concerne les acteurs autres que les services de l'Etat et ses opérateurs.

3.5.4 Un niveau opérationnel de mise à disposition des données

La mise en place d'une organisation facilitant les accès est essentielle, sauf à créer un droit virtuel. Quelle que soit l'Administration responsable d'organiser ces accès, elle doit satisfaire à un cahier des charges garant de la qualité du service rendu au demandeur et du respect de la sécurité et de la confidentialité des données. Un service d'extraction de données devrait être également offert incluant la mise à disposition de données rendues anonymes. Ce service à façon pourrait s'inscrire dans un modèle économique fondé sur la facturation des prestations concernées.

³⁶ Dans l'échantillon, si la règle de tirage est secrète et si certaines données sont rendues suffisamment imprécises, on ne pourra jamais affirmer qu'une personne est reconnaissable même si elle présente (âge, sexe, parcours de soins) des caractéristiques uniques.

Les prestations doivent être servies dans le cadre d'un centre de service dédié et identifiable afin notamment de permettre au comité d'orientation (décrit en 3.5.2) d'évaluer la qualité du service rendu et d'en publier les résultats.

En outre, les chercheurs de l'Inserm ont émis le souhait de pouvoir constituer une plateforme d'accès aux données, respectant les mêmes garanties de sécurité et de traçabilité.

En tout état de cause, la centralisation et la consolidation des données issues des systèmes d'information de gestion continueront d'être réalisées par les organismes (ATIH et CNAMTS) qui en sont chargés aujourd'hui. Les représentants des professionnels de santé et des établissements de santé doivent continuer d'être associés au suivi de ces procédures dans le cadre d'un comité de gestion.

3.5.5 Le renforcement de l'effectivité des sanctions

Qu'il s'agisse des accès permanents ou temporaires, l'existence d'un dispositif crédible de contrôle et de sanction constitue un élément dissuasif essentiel pour prévenir les accès non autorisés ou l'utilisation des données à des fins autres que celles prévues par la Cnil. Plus qu'un éventuel renforcement des sanctions, il s'agit de mettre l'accent sur l'effectivité des sanctions et donc des contrôles opérés sur l'effectivité des mesures de sécurité (serveur et conditions d'accès) et la conformité entre les traitements opérés et les traitements autorisés par la Cnil

Conclusion

Les travaux de la Commission ont permis de mesurer le consensus existant sur les impacts positifs d'une plus grande ouverture et d'une meilleure utilisation des différents types de données produites par le système de santé, qui permettent d'attendre de nombreux bénéfices en termes de démocratie sanitaire, de renforcement de l'autonomie des patients, de développement de la recherche et de l'innovation, d'efficacité de l'action publique et d'amélioration des pratiques professionnelles.

La principale limite à cette ouverture étant la nécessité de garantir la protection de la vie privée des patients, la Commission a cherché les principes permettant de garantir ces deux objectifs.

La nécessité de protéger la vie privée des personnes est d'autant plus partagée que les données de santé constituent des données particulièrement sensibles. Même en supprimant tout identifiant direct dans ces données, il faut se prémunir contre les risques de réidentification des personnes dans les bases médico-administratives, qui peut survenir dès lors qu'un tiers se trouverait avoir par ailleurs connaissance de certains traits spécifiques du parcours de soins de ces personnes.

La Commission considère qu'il est nécessaire d'adopter des attitudes différentes en fonction du niveau de risque objectivement mesuré, qui constitue le seul critère pouvant justifier les différences de procédure d'accès. Il est proposé, au sein de chaque classe de risque, d'unifier et de simplifier les procédures d'accès.

Pour les données strictement anonymes, c'est-à-dire, à risque nul d'identification, l'accès doit être totalement libre, sans restriction de réutilisation des données publiques. Le champ de la santé doit ainsi résolument s'engager dans la voie de l'open data. A ce titre, la Commission prône le principe d'ouverture par défaut des données publiques anonymes de santé, sauf exception motivée.

La Commission rappelle que la qualité des données, leur complétude ou leur complexité ne constituent pas des critères devant déterminer l'ouverture ou non des données de santé anonymes. L'ouverture des données granulaires doit être privilégiée et les enquêtes et recherches menées sur fonds publics devront dès leur conception prévoir l'ouverture des données granulaires anonymisées.

Les réunions plénières de la Commission, et plus encore les rencontres et ateliers thématiques organisés en parallèle, ont permis de lancer un mouvement dynamique autour de l'open data en santé, concrétisé par l'ouverture d'un certain nombre de jeux de données. La définition d'une liste de jeux de données dont la commission demande l'ouverture dans les plus brefs délais, et l'identification d'actions préalables nécessaires à mettre en œuvre, à court terme, pour l'ouverture d'un certain nombre de jeux de données supplémentaires, sont listés dans le rapport.

La commission recommande de continuer ces travaux à travers la constitution d'un groupe de travail pour renforcer l'ouverture des données médico-administratives et élaborer « une offre open data » ambitieuse dépassant les freins actuels n'ayant pas de justification légale.

Des problématiques spécifiques ayant été soulignées pour les données d'enquêtes et de recherche, la Commission recommande qu'un groupe de travail soit également constitué afin de définir des solutions au niveau législatif, technique et économique pour ouvrir ces données à tout public.

Enfin, la Commission recommande que les directions générales des établissements publics et des opérateurs renforcent leurs actions concernant l'open data à travers la définition d'une feuille de route et d'un échéancier pour l'ouverture des données. Cette feuille de route pourrait être élaborée avec l'appui de correspondants open data dans les organisations concernées.

A ce titre, des travaux techniques ont été menés pour constituer des jeux de données anonymes à partir du PMSI. Ces travaux devront se poursuivre et être étendus à l'ensemble du SNIIRAM.

Pour les données présentant un très faible risque de ré-identification du fait de la granularité limitée des informations mises à disposition ou de la non exhaustivité de la population décrite (échantillon), une procédure simplifiée pourrait être retenue dès lors que la Cnil, après avis d'un comité technique, en aura validé les caractéristiques. L'autorisation d'accès à tel ou tel organisme pourrait relever d'un arrêté ministériel, après avis de la même instance, pour ce qui concerne les acteurs autres que les services de l'Etat et ses opérateurs. Des propositions ont été formulées pour créer de nouveaux jeux de données répondant à cet objectif.

Enfin, pour les données détaillées, emportant de ce fait un risque plus important de ré-identification, il est proposé de simplifier le circuit des demandes d'accès des organismes privés ou publics en créant un canal unique d'autorisation délivrée par la Cnil :

- après avis d'un comité technique pour ce qui concerne la recherche académique ;
- après avis d'un comité technique et d'un comité d'orientation s'il ne s'agit pas de recherche académique.

Enfin, s'agissant de l'autorisation d'accès à ces données par les administrations, les opérateurs de l'Etat, les ARS ainsi que d'autres utilisateurs permanents des données, il est recommandé de confier cette décision au ministre en charge de la santé après avis de la Cnil.

L'étape d'instruction technique n'aurait pas lieu d'être lorsqu'une recherche est réalisée à la demande d'une autorité publique et qu'elle a validé le protocole d'étude, la Cnil restant naturellement le décisionnaire ultime. D'autres procédures simplifiées visant à rendre la procédure plus rapide pourront être mises en place sur le modèle déjà existant des méthodologies de référence notamment, ou pour l'accès à des ensembles de données à faible risque de ré-identification. Seraient également concernées par cette procédure, les demandes d'industriels visant à alimenter un dossier type requis par l'Administration.

Une attention particulière a été portée aux données relatives aux professionnels et établissements de santé. L'exigence de transparence sur des sujets aussi importants pour les patients que le reste à charge et la qualité des soins délivrés justifie la mise à disposition de données individuelles. Si cette mise à disposition est maintenant bien acquise dans le champ des établissements de santé, elle pose plus de difficultés s'agissant des professionnels de santé rendus ainsi directement et personnellement exposés. La Commission propose donc une approche graduelle et concertée permettant des avancées immédiates, par exemple sur la réutilisation des données tarifaires ou sur les données permettant d'identifier le profil d'activité d'un professionnel et confiant à la concertation entre les parties prenantes la responsabilité de faire émerger progressivement des indicateurs rendant compte de façon fiable et pertinente de la qualité des soins délivrés. La Commission appelle de ses vœux un lancement rapide de cette concertation. Les représentants des professionnels appellent néanmoins à la prudence sur ce sujet, la constitution et la publication

d'indicateurs individuels nécessitant, dans certains cas, une certaine pédagogie pour leur interprétation ou leur appropriation.

L'effectivité des mesures proposées dépend de la qualité de l'organisation qui sera mise en place pour leur donner une réalité : modalités d'accès aux données, traçabilité des accès, mais aussi respect des données d'instruction des demandes, degré d'ouverture réelle en fonction des réponses apportées aux demandes d'accès, capacité à prendre en compte des questions nouvelles...

Ceci doit conduire à prêter une attention particulière aux moyens qui seront attribués tant pour l'instruction des demandes d'accès que pour la mise à disposition effective des données. Dans un contexte de contrainte économique forte pour l'ensemble des acteurs, le choix du modèle économique est un sujet sur lequel la Commission souhaite insister même s'il n'était pas inclus dans sa saisine.

La question de la gouvernance, garante de la mise en œuvre effective des principes posés, a fait l'objet de longs débats. Les propositions qui sont formulées visent à répondre aux préoccupations légitimement exprimées par les différentes parties prenantes. De façon résumée : garantir au chercheur un accès réel aux données au nom des exigences de la seule recherche, mais aussi garantir aux parties prenantes (patients et usagers, professionnels de santé, industriels, assureurs...) un examen ouvert et loyal de leur demande d'accès au vu des seules règles posées.

Ces préoccupations conduisent la commission à recommander une organisation sous la forme de deux comités disposant de compétences propres, fonctionnellement articulés autour d'un guichet unique constituant *a minima* un secrétariat commun aux deux instances, et de préférence regroupés au sein d'une même structure :

- Un comité technique composé d'experts et de scientifiques dont la diversité et l'indépendance doivent être assurées, désignés par les responsables d'autorités compétentes telles que l'INSERM, le HCSP, l'ANSM, la HAS... Cette instance étudie les demandes d'accès au regard des critères mis en avant par la Commission (bénéfice collectif, pertinence des traitements, références de l'équipes...). Son avis est ensuite transmis à la Cnil ;
- Un comité d'orientation, réunissant l'ensemble des parties prenantes, constituant le lieu de supervision du dispositif et d'animation des débats qui doivent se poursuivre au-delà des travaux de la Commission, notamment pour ce qui concerne l'accès aux données nominatives d'activité ou de tarification des professionnels de santé. La Commission recommande que ce comité puisse être également saisi, à l'initiative de son président ou de son représentant, de l'évaluation du bénéfice collectif attendu des projets soumis par d'autres acteurs que ceux de la recherche académique, afin de garantir une meilleure prise en compte des enjeux collectifs et sociétaux induits par la demande. Son avis est ensuite transmis à la Cnil, l'avis du comité technique étant joint pour ce qui concerne l'évaluation des autres critères.

La mise en œuvre de ces recommandations suppose diverses mesures législatives et réglementaires dont les premières devront pouvoir trouver leur place dans la loi de santé publique en cours de préparation.