



Commission open data en santé

RAPPORT

Remis à Madame Marisol TOURAINÉ,

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'efficience 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 Lenne, 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 devraient 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.gouv.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) réglemente 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, numéros clausus...).

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 personnels 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 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 poursuite, 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 chiffrement 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 2e 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'efficience 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 transcris 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 éléver 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 glargin,
- 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, CCMSA et RSI ; CNSA ; INVS ; HAS ; ANSM ; ARS sur leur champ de compétence régional et sous l'autorité d'un médecin salarié de l'agence).

➔ exactions 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'efficience 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 également être 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'efficience 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 efficience 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'efficience 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'efficience 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 réalisé 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 règlementaires.

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, Etabal 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	<ul style="list-style-type: none"> - 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	<ul style="list-style-type: none"> - 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 métadonnées.

- 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 Etablab 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 d'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 réidentification 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 réidentification 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 règlementaire 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 règlementaires 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 (décris 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 réidentification, 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 réidentification. 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 sociaux 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.

Commission open data en santé

Annexes du RAPPORT

Juillet 2014

Annexes

Annexe 1 : Composition de la commission open data

Annexe 2 : Lettre de mission

Annexes 3a et 3b : politique du gouvernement français en matière d'open data (feuille de route du gouvernement, décisions prises en CIMAP)

Annexe 4 et 4 bis : compléments sur l'état des lieux juridique

Annexe 5 et 5 bis : compléments sur les accès au SNIIRAM et au PMSI

Annexe 6 : Cartographie des données de santé

Annexe 7 : Panorama international et principaux enseignements

Annexe 8 : Compte-rendu des ateliers thématiques

Annexe 8 bis : Synthèse de la consultation publique

Annexe 9 : rapport du groupe technique « risques de réidentification »

Annexe 10 : Contributions des membres

Annexe 1
Composition de la commission

Philippe	BURNEL	DSSIS	Délégué à la stratégie des systèmes d'information de santé
Christian	BABUSIAUX	IDS	Président
Christian	BAUDELOT	ENS	Sociologue
Max	BENSADON	ATIH	Directeur adjoint
Kat	BORLOGAN	Five by Five	Co-fondatrice
Yann	BOURGUEIL	IRDES	Directeur
Chantal	CASES	INED	Directrice
David	CASTILLO	FHP	Responsable des études économiques et des SI
Brigitte	CONGARD-CHASSOL	SNITEM	Directeur Affaires Médicales, Economiques et Export, Responsable Qualité
Patrick	CORNE	UNPS	Masseur-Kinésithérapeute
Mathieu	COUSINEAU	UNOCAM	FNMF - responsable pôle assurance santé
Laurent	CYTERMANN	Conseil d'Etat	Rapporteur général adjoint
Bénédicte	DELAUNAY	Université de Tours	Professeur de droit public, membre de la CADA
Mathieu	ESCOT	UFC QUE CHOISIR	Chargé de mission santé
Martial	FRAYSSE	CNOP	Président du conseil régional des pharmaciens d'Ille-de-France
Yves	GAUBERT	FHF	Responsable pôle Finances
Pierre-Yves	GEOFFARD	PSE - ENS	Directeur de l'Ecole d'Economie de Paris
Claude	GISSOT	CNAMTS	Directeur de la stratégie, des études et des statistiques
Jean-François	GOGLIN	FEHAP	Conseiller Systèmes d'Information de Santé
Marcel	GOLDBERG	INSERM	Chercheur
Evelyne	GUILLET	UNOCAM	CTIP - directrice santé
Chrystel	JOUAN-FLAHAUT	LEEM	Directeur médical du Leem
William	JOUBERT	UNPS	Médecin - vice-président UNPS
Pierre	LEPORTIER	UNPS	Pharmacien - vice-président UNPS
Frédérique	LESAULNIER	CNIL	Responsable pôle santé
André	LOTH	DREES	Directeur de projet
Jacques	LUCAS	CNOM	Vice-président
Bertrand	LUKACS	APHP	Chirurgien urologue
Cécile	MALGUID	UNOCAM	FFSA - responsable des études santé
Odile	MARQUESTAUT	FNEHAD	-
Jacques	MARZIN	DISIC (SGMAP)	Directeur
Jean-Roch	MEUNIER	MEDICEN PARIS REGION	Délégué Général
Jean-Paul	MOATTI	INSERM ; Aviesan	Directeur ITMO santé publique
Michel	PINAULT	COEPIA	Président
Philippe	RANQUET	DAJ	directeur des affaires juridiques des Ministères sociaux
Grégoire	REY	CépiDc-INSERM	Directeur

Emmanuel	REYRAT	FNCLCC	Directeur du département des systèmes d'information
Christophe	RICHARD	SYNTEC	Medical director Santeos
Christian	SAOUT	CISS	Président d'honneur du CISS
Bruno	SARFATI	SYNTEC	Vice-Président Cegedim
Henri	VERDIER	Etalab (SGMAP)	Directeur
Franck	VON LENNEP	DREES	Directeur de la recherche, des études, de l'évaluation et des statistiques
Mahmoud	ZUREIK	ANSM	Directeur de la stratégie et des affaires internationales

Annexe 2

Lettre de mission



[Lettre de mission.pdf](#)

Annexe 3 : Politique du gouvernement français en matière d'open data :

Feuille de route du gouvernement français sur l'ouverture et le partage des données publiques (3a), extraits des relevés des décisions prises sur l'open data en Comité Interministériel pour la Modernisation de l'Action Publique (3b), et Vademecum

Annexe 3a – Feuille de route du gouvernement français sur l'ouverture et le partage des données publiques

L'ouverture des données publiques, liberté publique et levier d'innovation

L'ouverture des données publiques est fondée sur la loi n° **78-753** du 17 juillet 1978 relative au droit d'accès aux documents administratifs, ainsi définis : « (...) 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}).

Le droit d'accès aux documents administratifs a été reconnu comme une « liberté publique » par le Conseil d'État (CE, 29 avril 2002, U., n° 228830). En 1997, le gouvernement a élargi le principe en décidant la mise en ligne gratuite des « données publiques essentielles ». 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, transposée par l'ordonnance du 6 juin 2005 et le décret du 30 décembre 2005, a facilité la réutilisation des documents existants détenus par des organismes du secteur public. Le décret n° **2011-577 du 26 mai 2011 relatif à la réutilisation des informations publiques détenues par l'Etat et ses établissements publics administratifs** a posé le principe de la gratuité du droit à réutilisation.

Le 21 février 2011, le décret n° 2011-194 portant création d'une mission « Etab » chargée de la création d'un portail unique interministériel des données publiques créait la structure en charge d'animer et d'héberger la diffusion de ces données publiques.

Avec la révolution numérique - qui facilite la production, la diffusion et le traitement de données de masse - l'ouverture et le partage des données publiques sont devenus un puissant levier pour :

- renforcer la confiance des citoyens grâce à une plus grande transparence de l'action publique (par exemple, avec la mise en ligne d'informations sur les comptes publics, la pollution ou les délits constatés) ;
- permettre de nouvelles formes de co-production avec la société et soutenir l'innovation sociale (comme le projet Handimap qui a permis, grâce aux données des villes de Rennes et de Montpellier, de développer une application de calcul d'itinéraire pour handicapés moteurs) ;
- améliorer le fonctionnement administratif (comme en témoigne la très forte utilisation des données publiques par l'administration elle-même) ;

- améliorer l'efficacité de l'action publique en développant de nouveaux modes d'organisation et de nouveaux processus de travail (comme le suivi des accidents de la route par la sécurité routière permettant d'améliorer l'aménagement de la voirie) ;
- soutenir le dynamisme économique, en créant de nouvelles ressources pour l'innovation et la croissance (comme en témoignent les jeunes entreprises innovantes primées par Dataconnexions) ;
- contribuer à l'influence et au rayonnement international de la France (initiative *Open Government Partnership* -OGP-, initiative sur la transparence de l'aide internationale -IATI- ou sur la transparence des industries extractives - EITI).

[L'engagement du Gouvernement en faveur de la transparence de l'action publique et de l'ouverture des données](#)

Dès le premier Conseil des ministres, le 17 mai 2012, le Président de la République a fait signer, à tous les membres du gouvernement, une charte de déontologie rappelant notamment leur « devoir de transparence », leur « scrupuleux respect des dispositions garantissant l'accès des citoyens aux documents administratifs » et leur engagement à « mener une action déterminée pour la mise à disposition gratuite et commode sur Internet d'un grand nombre de données publiques ».

En lançant le chantier de la modernisation de l'action publique, le Premier ministre a réaffirmé le 31 octobre 2012 les principes d'ouverture et de partage des données publiques prévus par la circulaire du Premier ministre du 26 mai 2011, notamment le droit pour tout citoyen de pouvoir réutiliser librement et gratuitement les données publiques, ou l'encadrement strict des redevances de réutilisation de données publiques.

En intégrant la mission Etabl au SGMAP, il a constitué une force globale d'innovation, associant la conception des systèmes d'information de l'Etat, l'ouverture des données publiques, l'évaluation des politiques publiques et l'innovation en matière de politiques publiques.

Dès le premier comité interministériel pour la modernisation de l'action publique (CIMAP), le 18 décembre 2012, le gouvernement a pris trois décisions importantes pour l'ouverture des données publiques :

- « Réaffirmer le principe de gratuité de la réutilisation des données publiques et l'étendre, en concertation, aux collectivités, aux délégataires de service public, à la sphère sociale et aux autorités administratives indépendantes » ;
- « Lancer, dès 2013, des travaux sur la transparence de l'action publique et des institutions, notamment en ouvrant le débat sur la mise à disposition de données en matière d'éducation, de risques environnementaux, de transparence du système de santé, de l'offre de transport, de logement, de lisibilité des prestations sociales et des dépenses publiques » ;
- « Confier au SGMAP, en lien avec le ministère du budget, la mission d'évaluer les modèles économiques de chaque redevance existante, notamment en auditant les coûts et les recettes associés. Le SGMAP doit remettre ses propositions au Premier ministre au printemps 2013. »

L'ouverture des données publiques est engagée. Outre les informations partagées (mais pas toujours réutilisables) par le service public de l'accès au droit (Légifrance), le cadastre ou encore Géoportail, de nombreuses bases de données sont mises en libre accès par les administrations, les établissements publics et les collectivités locales.

La mission Etabl a développé une plateforme nationale d'hébergement et de partage de données publiques ouvertes, gratuites et réutilisables (www.data.gouv.fr). Elle coordonne un réseau de

correspondants Open data et de fournisseurs de données qui ont permis la mise en ligne de plus de 350.000 jeux de données. La mission Etalab a également travaillé à structurer l'écosystème des réutilisateurs de données (startups, associations, ...), notamment par le biais de l'organisation du concours Dataconnexions, qui lui a permis, en deux éditions, de récompenser 11 projets particulièrement prometteurs.

L'ambition du Gouvernement en matière d'ouverture des données publiques

Le gouvernement entend donner sa pleine mesure à cette politique, et en faire un levier de modernisation de l'action publique et de soutien au dynamisme économique.

1. Travailler à l'ouverture de jeux de données stratégiques

Le gouvernement souhaite favoriser l'ouverture de données stratégiques et de qualité, qu'elles permettent de créer des services utiles au quotidien (comme pour le transport, l'environnement, la santé, la consommation énergétique, le tourisme...), qu'elles soient réclamées par les citoyens en raison de leur caractère démocratique ou pratique (budgets et dépenses détaillées par administration...), ou qu'elles répondent à une exigence légale ou réglementaire (publication des attributaires de marchés publics conformément à l'article 138 du code des marchés publics).

Pour ce faire, la mission Etalab du SGMAP organisera, dès 2013, six débats thématiques et ouverts, afin d'identifier et de publier de nouveaux jeux de données. Ces débats seront confiés à une commission composée par le SGMAP et le ministère concerné, pouvant bénéficier d'un rapporteur ou d'un secrétariat assuré par Etalab ; ils donneront lieu à un rapport, remis au Premier ministre et au ministre intéressé, qui formuleront des recommandations en termes d'ouverture cohérente des données et préciseront les bénéfices attendus.

Dans ce cadre, le gouvernement étudiera les possibilités :

- d'élargissement de la stratégie d'ouverture des données publiques à tous les établissements publics administratifs, aux délégataires de service public, aux collectivités territoriales et aux autorités administratives indépendantes ;
- d'extension de la politique d'ouverture des données publiques aux missions de service public à caractère industriel et commercial, aux entreprises publiques et à la sphère associative.

Le Premier ministre a par ailleurs demandé à tous les ministres de préciser leur stratégie d'ouverture des données publiques dans les programmes ministériels de modernisation et de simplification (PMMS), à transmettre d'ici la mi-mars.

2. Faciliter et améliorer le processus d'ouverture des données publiques dans les ministères, les établissements publics, et les collectivités territoriales

Etalab ouvrira, dès 2013, une nouvelle version de la plateforme www.data.gouv.fr. Tout en conservant sa vocation de diffuser largement les données brutes, celle-ci deviendra à terme une vitrine grand public de l'ouverture et du partage des données publiques. Elle proposera en particulier des outils techniques utiles pour s'approprier ces données (outils de visualisation de données, de traitement statistique, etc.).

Une attention toute particulière sera apportée à l'indexation des données, à leur traitement sémantique et aux liens entre ces données. Etalab, en collaboration avec les collectivités locales et les autres producteurs de données publiques, travaillera à répertorier et à présenter l'ensemble des

données publiques librement accessibles et gratuitement réutilisables, quelle qu'en soit la source. De même, Etalab poursuivra avec les fournisseurs de données la réflexion sur l'importance des formats ouverts et la possibilité de les faire converger, et s'attachera à construire une plus grande interopérabilité entre la plateforme www.data.gouv.fr et les systèmes d'information de l'État, des collectivités territoriales ou des institutions européennes.

Par ailleurs, Etalab lancera également dès 2013 le chantier de la diffusion de données par des interfaces programmables (« API »).

[3. Accompagner les réutilisations innovantes et la communauté des réutilisateurs](#)

L'État encourage la réutilisation la plus large des données publiques afin que cette politique d'ouverture soutienne une stratégie d'innovation, y compris sociale, et de croissance économique.

- Le service public est le premier utilisateur de données publiques de qualité**

Les données publiques sont produites dans le cadre d'une mission de service public et le service public est fréquemment le premier bénéficiaire de leur ouverture. La production, la collecte et la publication de données par l'administration doivent répondre aux besoins d'un service public de qualité. Etalab doit donc être à la disposition des administrations pour leur permettre de saisir ces opportunités. En particulier, elle organisera des sessions de formation, pourra monter des ateliers ad-hoc, pour analyser le potentiel de certaines applications de l'ouverture des données publiques (lutte contre le chômage, par exemple), et publiera des notes d'analyses de cas exemplaires.

- L'Etat favorise la réutilisation des données publiques par les entreprises pour créer des services innovants**

Etalab continuera à soutenir les entreprises innovantes, l'économie sociale et solidaire ainsi que les grands groupes industriels, dans leurs projets de réutilisation de données publiques. Elle organisera en 2013 trois concours Dataconnexions, permettant de repérer et de faire connaître des réutilisations exemplaires, et soutiendra l'organisation d'Assises nationales de l'open data, en lien avec le ministère chargé du redressement productif, le ministère chargé des PME et le ministère chargé de l'économie sociale et solidaire.

- L'Etat soutient l'engagement des citoyens et des associations autour des données publiques au profit d'une action publique moderne et renouvelée**

Le soutien à l'engagement des citoyens, pour coproduire certaines informations, concevoir de nouveaux services ou aider les acteurs publics à relever certains défis, fait partie intégrante des missions d'Etalab, qui organisera en particulier des concours mobilisant les citoyens autour des données et des objectifs partagés par l'État.

- Suivre et encourager la recherche en sciences de la donnée**

Les « Big Data », la transformation des chaînes de valeur, les stratégies et les organisations fondées sur la « culture de la donnée » caractérisent les formes actuelles de la révolution numérique. Pour suivre cette transformation rapide, Etalab accentuera sa coopération avec le monde de l'enseignement supérieur et de la recherche et se dotera, avant la fin de l'année 2013, d'un collège d'experts rassemblant des experts technologiques, des spécialistes de data sciences, des juristes, des économistes ou des réutilisateurs (entrepreneurs ou responsables associatifs), à moyens constants.

[4. La transparence et la culture de la donnée au service de la modernisation de l'action publique](#)

L'ouverture des données publiques est un outil majeur au service de la transformation du service public, pour en améliorer la qualité et en permettre l'évaluation par tous ses usagers. Le SGMAP veillera à construire, chaque fois que cela sera pertinent, des référentiels quantitatifs, si possible en temps réel, pour les évaluations des politiques publiques, afin qu'elles permettent de construire des outils utiles au pilotage de ces politiques publiques. Comme prévu par le CIMAP du 18 décembre 2012, le SGMAP travaillera à la conception d'innovations dans les services publics en ayant recours aux données publiques.

5. L'évaluation des redevances existantes

À l'occasion du CIMAP du 18 décembre 2012, le Premier ministre a confié au SGMAP, en lien avec le ministère du Budget, une mission d'évaluation des modèles économiques des redevances existantes, notamment en auditant les coûts et les recettes associés. Ces recommandations seront remises au premier semestre 2013.

Etalab et le SGMAP accompagneront les administrations qui commercialisent de l'information publique dans l'évolution de leurs modèles économiques. En effet, le respect du principe d'accès libre et gratuit du citoyen à des données publiques réutilisables n'empêche pas de concevoir des stratégies de monétisation de services à haute valeur ajoutée ou de création de plateformes.

6. Évolutions du cadre juridique et administratif

Un vade-mecum rappelant le cadre juridique de la mise à disposition des données publiques sera diffusé aux ministères, afin de faciliter leur engagement dans la démarche d'ouverture de leurs données.

A l'occasion, notamment, de la révision de la directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003 concernant la réutilisation des informations du secteur public, le gouvernement lancera des travaux de concertation, visant à actualiser si nécessaire les définitions légales encadrant le droit d'accès et de réutilisation des données publiques.

7. L'action internationale de la France

Enfin, le gouvernement attache une importance toute particulière au rayonnement de la France sur le sujet de la transparence, de l'ouverture des données publiques (« Open Data ») et de la gouvernance ouverte (« Open Government »). Etalab prendra part au pilotage des projets européens de plateformes de données ouvertes communautaires (notamment le site ouvert au début de l'année 2013 par la Commission européenne (<http://open-data.europa.eu/open-data/>)).

La France étudie par ailleurs l'opportunité de rejoindre l'*Open Government Partnership* (« OGP »), pour engager une coopération fructueuse avec les autres gouvernements engagés dans cette voie. Elle se rapprochera du partenariat pour examiner les conditions de son adhésion. La France s'engagera par ailleurs largement dans la coopération multilatérale autour des approches de transparence et de gouvernement ouvert.

Annexe 3b – Décisions en CIMAP relatives à l'ouverture des données publiques (extraits des relevés de décisions)

CIMAP du 18 décembre 2012

Une administration ouverte pour encourager la citoyenneté active

Décision n°32. Réaffirmer le principe de gratuité de la réutilisation des données publiques et l'étendre, en concertation, aux collectivités, aux délégataires de service public, à la sphère sociale, aux autorités administratives indépendantes : dès 2013, la transposition du projet de directive européenne sur la réutilisation des données publiques offrira cette occasion.

Décision n°33. Développer la transparence de l'action publique et stimuler une citoyenneté active. Dès 2013, des travaux seront lancés sur la transparence de l'action publique, en ouvrant le débat sur la mise à disposition des données en matière d'éducation, de risques environnementaux, de système de santé, d'offre de transport et de logement et enfin de prestations sociales. L'implication des citoyens, des usagers et des agents dans la conception des politiques publiques et l'élaboration des services publics sera favorisée.

Décision n°34. Favoriser l'innovation collective autour des services publics : dix “services numériques de demain” seront conçus et testés dans une démarche de “laboratoire”.

CIMAP du 2 avril 2013

Poursuivre l'ouverture des données publiques à fort impact économique et démocratique

Conformément aux décisions du CIMAP de décembre, la feuille de route du Gouvernement en matière d'ouverture et de partage des données publiques a été rendue publique au cours du séminaire sur le numérique du 28 février 2013. Elle poursuit trois objectifs principaux : l'ouverture et le partage de nouvelles données à fort impact économique et démocratique, la création d'un “écosystème” cohérent de l'ouverture des données publiques et la mise en place d'une ouverture “par défaut”, intégrée dès la conception des systèmes d'information.

De nombreux nouveaux jeux de données ont d'ores et déjà été partagés sur le site data.gouv.fr :

- dans le domaine de la sécurité, les crimes et délits et les faits constatés par les services de police et de gendarmerie en 2012 ;
- de nombreuses données culturelles, telles que la liste des mobiliers classés monument historique par région ou les aides à la presse ;
- dans le champ de la santé, des données ont été mises à jour, telles que les honoraires des médecins libéraux par spécialité et lieu d'exécution de janvier à octobre 2012 ou les dépenses de santé remboursées par l'assurance maladie par région de janvier à octobre 2012 ;
- dans le secteur de la consommation, les cotations des fruits et légumes par marché et par produit sont réactualisées à un rythme hebdomadaire.

Enfin, la mission d'évaluation des modèles économiques des redevances existantes a été lancée et rendra ses conclusions à la fin du mois de juin.

Au-delà de ces premiers acquis, le Gouvernement se mobilise pour concrétiser cette feuille de route :

- Les premiers débats thématiques ouverts (logement, santé, enseignement supérieur) permettant de créer les conditions d'une politique d'ouverture des données publiques utile aux citoyens, aux entreprises et à la modernisation de l'action publique seront lancés au premier semestre 2013 ;

- Un vade-mecum de l'ouverture des données publiques, rappelant le cadre juridique et technique de la mise à disposition des données publiques, sera mis à disposition des administrations avant juin 2013 ;
- L'étude technique des modalités de l'adhésion au "Partenariat pour une gouvernance ouverte" (Open Government Partnership) est en cours.

Pour franchir une étape supplémentaire, le CIMAP arrête les décisions suivantes :

Décision n° 13 : Le Gouvernement publiera dès 2013 de nouveaux jeux de données sur data.gouv.fr dans des domaines stratégiques de l'action publique.

Ceci concerne notamment : la sécurité, avec la publication des zones de vigilance accrue en matière de sécurité routière ; la santé, avec la mise en ligne de données sur les permanences de soins ambulatoires et la couverture maladie universelle (CMU) et des indicateurs de qualité et de sécurité des soins ; la vie économique, notamment grâce à la publication de données produites par la Banque publique d'investissement (BPI), la direction des entreprises de la Banque de France et l'observatoire des PME ; la recherche, avec les dépenses en recherches et développement, ainsi que le nombre de chercheurs par région, dans les administrations publiques et privées ; dans le champ environnemental, les données relatives à la qualité de l'eau agrégées par l'ONEMA.

Décision n° 14 : Le Gouvernement prendra systématiquement en compte l'exigence de partage des données publiques dans les réformes qu'il présente.

Ainsi, le projet de loi de décentralisation et de réforme de l'action publique prévoit que les collectivités territoriales qui disposent d'un site Internet mettront les données publiques concernant leur territoire à disposition des citoyens ; l'amélioration des réseaux de transport en Île-de-France sera accompagnée de progrès de "l'information-voyageur" par l'ouverture de données de la SNCF et de la RATP ; le projet de loi sur l'enseignement supérieur et la recherche donnera un plus large accès aux données fiscales aux enseignants-chercheurs.

Décision n° 15 : Le Gouvernement améliorera l'accès aux données ouvertes par la création d'un catalogue national permettant de trouver l'ensemble des données publiques librement accessibles; l'élaboration d'une **nouvelle version de la plateforme www.data.gouv.fr** : un travail collaboratif sera conduit pour faire évoluer le site et en faire un lieu de citoyenneté et d'innovation. Le nouveau site sera opérationnel à l'automne 2013.

CIMAP du 17 juillet 2013

L'ouverture des données publiques s'intensifie

Depuis le CIMAP du 2 avril 2013, le Gouvernement a mis à disposition plus de 200 nouvelles séries de données portant sur des domaines stratégiques de l'action publique, comme par exemple : l'aide publique au développement (données détaillées projet par projet), la transparence démocratique (résultats du référendum du 7 avril 2013 relatif au projet de création d'une collectivité territoriale d'Alsace), la transparence de l'action de l'Etat (attributaires de marchés publics), l'égalité entre les femmes et les hommes (principaux chiffres clés en 2012, panorama régional des inégalités), l'éducation (résultats du diplôme national du brevet, classement des performances des lycées), la santé (catégories majeures de diagnostic par établissement de santé), le tourisme (ensemble des établissements touristiques classés) ainsi que la liste de l'ensemble des interventions économiques auprès des entreprises étudiées dans le cadre de la mission d'évaluation de cette politique.

Par ailleurs, sur la base du rapport de M. Mohammed Adnène Trojette, magistrat à la Cour des comptes, relatif au modèle économique des redevances d'utilisation des données publiques, qui sera

remis au Gouvernement et rendu public, le Gouvernement élaborera un plan d'action partagé relatif aux redevances existantes.

Décision n°16 : A la suite de la signature par le Président de la République de la charte du G8 pour l'ouverture des données publiques, le 18 juin 2013, le Premier ministre demande à chacun des ministres de participer à l'élaboration du plan d'action national prévu par la charte pour la fin 2013.

Décision 17 : Le SGMAP (mission Etalab) coordonnera avec l'ensemble des administrations concernées la transposition dans le droit national de la directive européenne du 26 juin 2013 concernant la réutilisation des informations du secteur public.

Décision n°18 : Les trois premiers débats thématiques sur l'accès aux données publiques sont effectivement programmés :

- Le ministère de l'égalité des territoires et du logement (METL) confie au conseil national de l'habitat (CNH), organe consultatif largement ouvert aux parties prenantes du logement, le soin de mettre en place, avec l'appui de la mission Etalab, un groupe de travail pluraliste (composé de membres du CNH et de représentants du monde des réutilisateurs de données publiques et de spécialistes de l'open data) qui commencera ses travaux au mois de septembre 2013 et qui remettra ses préconisations au METL au début de l'année 2014.
- Le ministère des affaires sociales et de la santé lancera un débat thématique sur l'ouverture des données publiques de santé. Nourri des conclusions de la mission confiée à M. Pierre-Louis Bras, ce débat sera officiellement lancé au mois de septembre 2013 pour une remise de conclusions au plus tard dans les six mois. Il sera porté par une commission pluraliste associant les parties prenantes pertinentes et ouvert aux contributions citoyennes.

CIMAP du 18 Décembre 2013

Open data et Gouvernement ouvert : faire de l'administration une plateforme pour l'innovation économique et sociale Le Gouvernement réaffirme son engagement dans le projet d'open data et de Gouvernement ouvert, levier de confiance démocratique, de stimulation de la croissance, d'innovation et de modernisation de l'action publique.

Le Gouvernement ouvre ce jour un nouveau portail d'open data, à l'adresse data.gouv.fr. Ce nouveau portail ouvre la voie d'une coproduction des données d'intérêt public. Il rassemble des données publiques et des données produites par les forces vives de la société, et met en avant les réutilisations et enrichissements de ces données apportés par la société civile (citoyens, associations...), les innovateurs, les chercheurs, les journalistes, etc.

Pour accélérer l'ouverture et le partage des données publiques, le Gouvernement clarifie sa doctrine en matière de redevances sur la réutilisation de données publiques :

Décision n°24 : Le Gouvernement réaffirme le principe de gratuité de la réutilisation des données publiques et décide de ne plus autoriser la création de nouvelle redevance.

Décision n°25 : Sur la base du rapport remis par M. Trojette relatif au modèle économique des redevances de réutilisation des données publiques et après examen des différentes catégories de données publiques soumises à redevance, le Gouvernement décide de supprimer plusieurs redevances. Les redevances supprimées sont celles perçues sur :

- les données relatives aux dotations globales de fonctionnement versées aux collectivités locales par la direction générale des collectivités locales (ministère de l'Intérieur) ;
- les données comptables brutes gérées par la direction générale des finances publiques – balance des comptes définitifs des collectivités et établissements publics locaux recensés à compter de l'arrêté

définitif des comptes 2013. Les fiches de situation financière des collectivités locales restent consultables sur le site [collectivités-locales.gouv.fr](http://collectivites-locales.gouv.fr)

- le parcellaire agricole de l'Agence de services et de paiement (ministère de l'agriculture, de l'agroalimentaire et de la forêt) ;
- les données cadastrales de l'Institut national de l'origine et de la qualité (ministère de l'agriculture, de l'agroalimentaire et de la forêt) ;
- les données brutes diffusées par l'Institut français du cheval et de l'équitation ;
- les données relatives aux adresses postales des établissements scolaires (ministère de l'éducation nationale) ;
- les données de l'Office national d'information sur les enseignements et les professions – ONISEP (ministère de l'éducation nationale) ;
- les données de la CADA
- les redevances sur la réutilisation professionnelle des informations suivantes diffusées par la DILA : ASSOCIATIONS, DEPOTS DES COMPTES DES ASSOCIATIONS, JORF, LEGI, KALI, JADE, CONSTIT, CASS-INCA, CAPP, CNIL, CIRCULAIRES.

Par ailleurs, le ministère de l'économie et des finances fournira en open data, sur une base hebdomadaire, la base des prix des carburants.

Décision n°26 : Le Gouvernement précise sa **doctrine en matière d'exceptions au principe de gratuité** :

- aucune redevance ne saurait être exigée sur des données résultant des missions de service public des administrations générales. Il est demandé aux administrations d'accroître leurs efforts en matière de publication de données leur permettant de mener à bien leurs missions, de motiver leurs décisions, de mesurer leurs activités ou leur bilan. Le site data.gouv.fr, est en mesure d'accueillir ces données ;
- le respect du secret statistique et donc l'anonymisation des informations pouvant éventuellement permettre d'identifier des personnes est un principe central du droit français. Certaines redevances ont été instaurées pour financer cette tâche, qui est indispensable avant ouverture de la donnée. Le Gouvernement demande au SGMAP d'apporter tout le soutien possible aux services publics concernés, afin d'identifier les moyens de réduire ces coûts sans dégrader la qualité de cette anonymisation ;
- enfin, le Gouvernement réaffirme que les opérateurs dont la mission même est de produire des données doivent rechercher des modèles économiques leur permettant de faire face à un paysage économique en profonde reconstitution. Conformément aux conclusions du rapport Trojette, il leur demande d'engager, dans les meilleurs délais, avec l'appui du SGMAP et du ministère du Budget, une réflexion sur les évolutions de leurs modèles économiques. Il leur demande de rechercher des modèles stimulant l'innovation autour de leurs données, favorables aux entrepreneurs innovants, et soutenables à l'heure de l'économie numérique, de la production de nombreuses données par les citoyens eux-mêmes, et des stratégies de plateformes.

Afin de mener cette transition - qui vise à, maintenir et améliorer la qualité des informations publiques, à garantir la pérennité du service public et à développer des stratégies numériques pérennes, efficaces et innovantes, le Gouvernement demande au SGMAP d'accompagner au mieux ces opérateurs, tel l'IGN (Institut géographique national), le SHOM (Service hydrographique et océanographique de la marine) ou Météo France, en lien avec les ministères de tutelle pour définir avec eux des trajectoires d'investissement dans l'innovation compatibles avec le redressement des finances publiques.

Le Gouvernement fait le choix d'investir durablement au service de la transition numérique de l'Etat et s'apprête à lancer les premiers appels à projets du Programme d'investissements d'avenir « Transition numérique et modernisation de l'action publique », doté de 150 M€.

Décision n°27 : Le Gouvernement lancera les premiers appels à projets co-financés par le **Programme d'investissements d'avenir « Transition numérique et modernisation de l'action publique »**, au premier semestre 2014. Les projets seront évalués par un comité d'experts indépendants et sélectionnés dans le cadre d'une gouvernance interministérielle mise en place avec le Commissariat général à l'Investissement et le SGMAP. Ils feront l'objet d'un suivi technique et financier régulier et donneront lieu à une évaluation ex post.

Annexe 4
La conservation, la transmission et l'hébergement des données de santé

Tout responsable de traitement de données à caractère personnel est soumis à une obligation générale de sécurité, définie par l'article 34 de la loi du 6 janvier 1978, selon laquelle « *le responsable du traitement est tenu de prendre toutes précautions utiles, au regard de la nature des données et des risques présentés par le traitement, pour préserver la sécurité des données et, notamment, empêcher qu'elles soient déformées, endommagées, ou que des tiers non autorisés y aient accès* ». Toutefois, compte tenu de la sensibilité particulière des données de santé, le législateur a instauré, dans le cadre de la loi du 4 mars 2002 relative aux droits des malades, deux régimes spécifiques destinés à préciser et à renforcer les obligations pesant sur les personnes assurant le traitement de ces données :

- le premier dispositif, prévu par l'article L. 1110-4 du code de la santé publique, régit la conservation sur support informatique et la transmission par voie électronique des données médicales détenues par les professionnels et les établissements de santé ;
- le second, qui figure à l'article L. 1111-8 du même code, encadre le dépôt de données de santé par les professionnels et les établissements de santé auprès de personnes tierces, qualifiées d'hébergeurs.

S'agissant de la conservation et de la transmission des données médicales, l'article R. 1110-1 dispose qu'elles doivent s'effectuer dans le respect de « *référentiels* » pris par arrêté du ministre de la santé après avis de la CNIL, qui fixent le niveau de sécurité requis. L'utilisation de la carte de professionnel de santé ou (depuis la loi HPST) d'un dispositif équivalent est obligatoire pour accéder aux données médicales informatisées et pour les transmettre par voie électronique.

S'agissant de l'hébergement de données de santé, la loi réserve cette activité à des prestataires agréés par le ministre de la santé, après avis de la CNIL et d'un comité d'agrément¹. L'agrément permet notamment de s'assurer de la qualification des personnels, de la sécurité des solutions techniques et de la mise en œuvre d'une politique de confidentialité. En outre, l'hébergement de données de santé ne peut se faire qu'avec le consentement exprès de la personne concernée. Les établissements et les professionnels de santé ne peuvent déroger à cette exigence de consentement que si l'accès aux données est limité à l'établissement ou au professionnel qui les a déposées. 64 sociétés ont été agréées à ce jour².

On peut souligner que le régime de l'hébergement de données de santé est une « *règle relative aux services* »³ au sens de la directive n° 98/34/CE du Parlement européen et du Conseil du 22 juin 1998 prévoyant une procédure d'information dans le domaine des normes et réglementations techniques et des règles relatives aux services de la société de l'information. En vertu des articles 8 et 9 de la directive, tout projet de modification significative de ce régime⁴ doit donc être communiqué à la Commission européenne et son adoption ne peut intervenir moins de trois mois avant la réception du projet par la Commission.

¹ Ce comité comprend un membre de l'IGAS, deux représentants des associations de patients, deux représentants des professionnels de santé et trois personnalités qualifiées compétentes en matière d'éthique et de droit, de sécurité des systèmes d'information et dans le domaine économique et financier.

² ASIP santé, mars 2014.

³ Article 1.5 : une règle relative aux services est « *une exigence de nature générale relative à l'accès aux activités de services visées au point 2 et à leur exercice, notamment les dispositions relatives au prestataire de services, aux services et au destinataire de services, à l'exclusion des règles qui ne visent pas spécifiquement les services définis au même point* ».

⁴ Selon l'article 8.1, la communication d'une modification d'une règle technique s'impose si les Etats « *apportent au projet de règle technique, d'une manière significative, des changements qui auront pour effet de modifier le champ d'application* ».

d'en raccourcir le calendrier d'application initialement prévu, d'ajouter des spécifications ou des exigences ou de rendre celles-ci plus strictes ».

Annexe 4 bis
Les dossiers de santé personnalisés

Un dossier de santé personnalisé rassemble sur un même traitement informatique des données à caractère personnel concernant un patient issues de l'activité de différents professionnels de santé. Trois systèmes coexistent aujourd'hui, qui ont chacun leur cadre légal et réglementaire : le dossier médical personnel (DMP) ; le dossier pharmaceutique (DP) ; le service « Web Médecin ».

1) Le dossier médical personnel

La création du dossier médical personnel a été prévue par la loi du 13 août 2004 relative à l'assurance-maladie. Il a vocation à recevoir l'ensemble des informations à caractère médical concernant à une personne (recensement des actes et consultations effectués par chaque professionnel de santé, éléments diagnostiques et thérapeutiques, résumés de séjour en établissement de santé) et à les conserver tout au long de sa vie.

Le régime du DMP est marqué par le souci de respecter le secret médical et l'exigence de consentement de la personne concernée. La création du DMP, qui peut être effectuée par tout professionnel de santé, nécessite l'accord de la personne. Celle-ci doit en outre donner son autorisation d'accès pour chaque professionnel de santé. L'accès au DMP ne peut être donné, même avec l'accord de la personne, à des personnes autres que les professionnels de santé exerçant en ville ou en établissement. Il est notamment interdit lors de la conclusion d'un contrat relatif à une complémentaire santé ou nécessitant l'évaluation de l'état de santé de la personne, ainsi que dans le cadre de la médecine du travail. Il ne peut être dérogé à l'exigence de consentement à l'accès que si la personne est hors d'état d'exprimer sa volonté et en présence d'une situation comportant un risque immédiat pour sa santé ; la personne peut s'opposer de manière expresse à ce que les professionnels de santé accèdent au DMP dans cette hypothèse.

La CNIL a autorisé une première phase de déploiement généralisé du DMP par une délibération n° 2010-449 du 2 décembre 2010. Plusieurs garanties pour la personne concernée ont été précisées dans le cadre du dossier de demande d'autorisation et de la décision de la CNIL : exigence d'une copie papier du document par lequel le professionnel de santé ayant ouvert le DMP atteste avoir informé la personne et recueilli son consentement ; remise préalable à la personne d'une brochure d'information, précisant notamment que son refus de créer un DMP n'aura pas de conséquences sur ses remboursements ainsi que les possibilités d'accès de la personne au DMP depuis son ordinateur personnel ; droit de la personne de demander que certains documents ne soient pas mentionnés dans le DMP ou soient « masqués » à certains professionnels de santé.

Le DMP est hébergé par un hébergeur agréé de données de santé (cf. annexe 1). Suite à un appel d'offres lancé en 2009 par l'ASIP santé, maître d'ouvrage du DMP, un groupement composé de la société Santeos et de filiales d'Atos et de La Poste a été retenu.

Plusieurs éléments du cadre juridique du DMP restent en suspens. Ni le décret prévu par l'article L. 1111-21 du code de la santé publique, qui devait notamment fixer les conditions d'accès aux différentes catégories d'information, ni celui relatif à l'identifiant national de santé, numéro distinct du NIR et non signifiant et qui devait être utilisé pour le DMP et le DP, n'ont été adoptés. Dans sa délibération du 2 décembre 2010, la CNIL a admis que le cadre juridique ne soit pas complètement fixé dans l'attente des retours d'expérience, mais a estimé que la deuxième phase de déploiement du DMP devrait se faire dans un cadre juridique achevé. Des dispositions législatives modifiant ponctuellement le cadre actuel seront inscrites dans la loi de santé publique et un décret d'application sera publié en aval.

A ce jour, près de 450 000 DMP ont été créés⁵.

2) Le dossier pharmaceutique

La création d'un dossier pharmaceutique a été prévue par une loi du 30 janvier 2007. L'objet du dossier pharmaceutique est plus restreint que le DMP : il ne comporte que les informations relatives aux médicaments et dispositifs médicaux dispensés. Si son ouverture est subordonnée, comme pour le DMP, au consentement de la personne attesté par la remise d'un écrit, son utilisation est ensuite soumise à des conditions moins strictes : alors que pour le DMP, le patient doit donner son accord à l'accès de chaque professionnel de santé (système qu'on peut qualifier « d'*opt-in* »), pour le DP, tout pharmacien est en principe tenu de le consulter et de l'alimenter, sauf opposition de la personne (système « d'*opt-out* »).

Le DP permet notamment aux pharmaciens de prendre connaissance des traitements pris par la personne et déceler des risques d'interactions médicamenteuses indésirables. Dans le cadre d'une convention signée le 3 novembre 2011 entre le Conseil national de l'ordre des pharmaciens, maître d'ouvrage du dispositif, et l'ANSM, il sert également à la diffusion des messages de rappels de lots de médicaments (fonction « DP-Rappels »). Enfin, l'article 23 de la loi du 29 décembre 2011 relative au renforcement de la sécurité sanitaire du médicament et des produits de santé a autorisé le ministre de la santé, l'ANSM et l'INVS à accéder aux données anonymes contenues dans les DP, dans une optique de veille sanitaire.

Le décret n° 2012-1131 du 5 octobre 2012 a étendu le DP aux pharmaciens exerçant dans les pharmacies à usage intérieur des établissements de santé. En outre, une expérimentation relative à l'accès au DP de médecins exerçant dans certains établissements de santé a été lancée pour trois ans par le décret n° 2013-31 du 9 janvier 2013 ; elle concerne notamment les médecins anesthésistes-réanimateurs, les urgentistes et les médecins exerçant dans les structures de médecine gériatrique. En revanche, l'alimentation du DMP par le DP, prévue par la loi, est subordonnée à l'adoption du décret relatif aux conditions de mise en œuvre du DMP (cf. point précédent).

Fin 2012, 97 % des officines de pharmacie étaient raccordées au DP et 24 145 000 DP avaient été créés, soit un taux de couverture de 36 %. Le taux de couverture est supérieur à 50 % pour les plus de 60 ans⁶.

3) Le service « Web Médecin »

Comme le DMP, le service Web Médecin a été créé par la loi du 13 août 2004 relative à l'assurance-maladie. Contrairement au DMP et au DP, il s'agit d'un service exclusivement destiné aux médecins, qui n'est pas directement accessible aux patients⁷. Il permet aux médecins de consulter l'historique de remboursements de leurs patients sur les douze mois précédent la consultation. Le patient doit donner son consentement à cette consultation, consentement qui est réputé obtenu lorsque le médecin utilise à cette fin la carte Vitale.

La gestion technique du dispositif est assurée par la CNAMTS.

⁵ 445 157 le 26 mars 2014, selon le site www.dmp.gouv.fr

⁶ CNOP, *Le dossier pharmaceutique. Rapport d'activité de l'année 2012*, avril 2013.

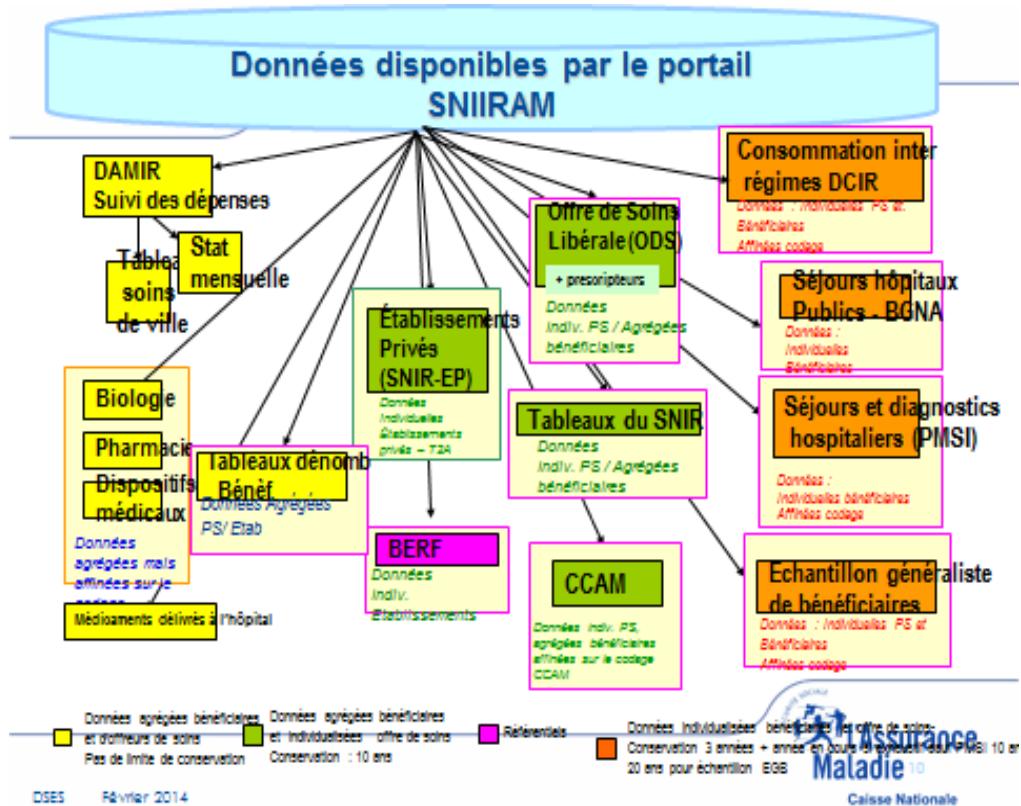
⁷ Le patient dispose cependant, comme pour tout traitement de données à caractère personnel, d'un droit d'accès qui s'exerce par demande adressée à la caisse d'assurance-maladie dont il relève.

Annexe 5

Arrêtés SNIIIRAM du 19 juillet 2013 et du 14 février 2014

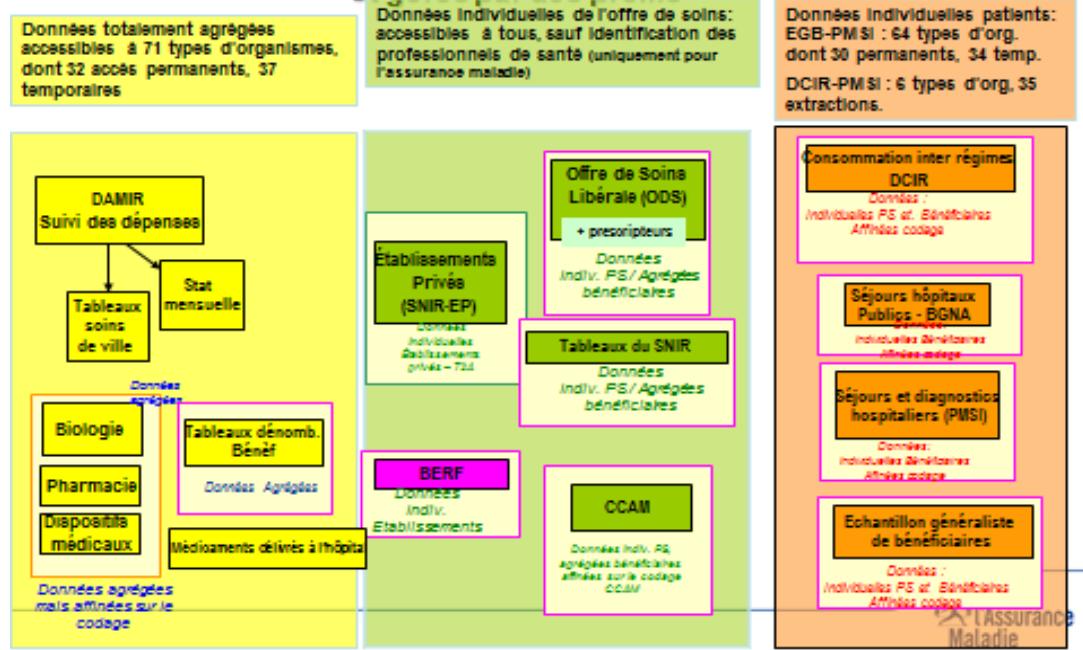
	Suivi des dépenses	Suivi de l'activité offreuse de soins	de par de	Echantillon généraliste des bénéficiaires	Données individuelles exhaustives
Caisse des régimes de base d'assurance maladie + CNSA	X	X	X		X
Institut de veille sanitaire (InVS)	X	X	X		X
Agence nationale de sécurité du médicament (ANSM)	X	X	X		X
Haute autorité de santé (HAS)	X	X	X		X
Agences Régionales de Santé (ARS)	X	X	X		X (sur périmètre régional)
Ministères (santé, sécurité sociale et finances) et services déconcentrés	X	X	X		(en cours DREES)
Haut conseil pour l'avenir de l'assurance maladie (HCAAM)	X	X	X		
Agence de la biomédecine	X	X	X		
Institut national du cancer (INCa)	X	X	X		
Agence technique de l'information sur l'hospitalisation (ATIH)	X	X	X		(en cours ATIH)
Union nationale des professions de santé (UNPS)	X	X	X		
Institut national de la santé et de la recherche médicale (INSERM)	X	X	X		
Centre national de la recherche scientifique (CNRS)	X	X	X		
Institut de Recherche et Documentation en Economie de la Santé (IRDES)	X	X	X		
Centre technique d'appui et de formation des centres d'examens de santé (CETAF)	X	X	X		
Fonds CMU	X	X	X		
Observatoire des drogues et toxicomanies	X	X	X		
Institut des données de santé (IDS)	X	X	X		
Union Nationale des Organismes Complémentaires d'Assurance Maladie (UNOCAM) <i>Membre de l'IDS</i>	X	X	X		
Fédérations constitutives de l'UNOCAM (FNMF, CTIP, FFSA) signataires de la charte d'engagement entre l'assurance maladie obligatoire et les assureurs maladie complémentaires	X	X	X		
Fédérations hospitalières (FHF, FHP, FEHAP) <i>Membre de l'IDS</i>	X	X	X		
Collectif interassociatif sur la santé CISS <i>Membre de l'IDS</i>	X	X	X		
Syndicats de professionnels libéraux membres de l'UNPS	X	X			
Unions régionales des professions de santé (URPS)	X	X			
Etablissements membres des fédérations hospitalières	X	X			
Associations adhérentes au CISS	X	X			
Organismes complémentaires contributeurs de données membres des fédérations constitutives de l'UNOCAM signataires de la charte d'engagement entre l'assurance maladie obligatoire et les assureurs maladie complémentaires	X	X			

Annexe 5 bis
Cartographie des produits SNIIRAM



Cartographie des types d'accès

Caractéristiques des accès SNIIR-AM définis par un arrêté et générés par des profils



Annexe 6

Cartographie des données de santé

L'étude réalisée par le SGMAP est disponible à l'adresse :

<https://www.data.gouv.fr/fr/dataset/cartographie-des-bases-de-donnees-publiques-en-sante>

Annexe 7

Panorama international et principaux enseignements

On trouvera en document joint une synthèse de l'étude :



[Panorama de l'open data en santé – note de synthèse.pdf](#)

La présentation détaillée du panorama se trouve à l'adresse :



[Open data en sante-panorama international 12 mars 2014.pdf](#)

Annexe 8
La cartographie des données de santé

Annexe 7a – Liste des institutions rencontrées pour réaliser la cartographie des données de santé

La liste des institutions rencontrées est exhaustive. Toutes les personnes rencontrées au sein de chaque institutions n'ont cependant pas pu être mentionnées ici. Le SGMAP remercie l'ensemble des personnes rencontrées pour le temps qu'ils nous ont consacré.

Institution	Interlocuteur	Position	Date
DREES	Franck Von Lennep	Directeur	05-févr
DREES	André Loth	Directeur de projet	05-févr
DREES	M. Badouraly	Responsable SI	05-févr
IDS	Richard Decottignies	Directeur adjoint	05-févr
INSERM	Grégoire Rey	Directeur du CépiDC	06-févr
ATIH	Max Bensadon	Directeur adjoint	07-févr
INPES	Pierre Ardewison	Directeur des affaires scientifiques	07-févr
IRDES	Yann Bourgeuil	Directeur	10-févr
INSERM	Marcel Goldberg	Chercheur / expert	11-févr
HAS	Catherine Grenier	Chef du service indicateurs pour l'amélioration de la qualité et de la sécurité des soins	17-févr
HAS	Sabine Cohen-Hygounenc	Responsable de la mission d'information sur la qualité des prises en charge dans les établissements de santé à destination des usagers	17-févr
HAS	Etienne Doutreleau	Adjoint au secrétaire général	17-févr
HAS	Georges Dutreix	Chef de projet SSI	17-févr
HAS	Hervé Nabarette	Chef du service Qualité de l'information médicale	17-févr
HAS	Ariane Sachs	Service évaluation et amélioration des pratiques	17-févr
ASIP Santé	Jeanne Bossi	Secrétaire Générale	17-févr
ASIP Santé	Anne Monnier		17-févr
ANSM	Mahmoud Zureik	Directeur de la stratégie et des relations internationales	18 février et 28 mars
CGIET	Robert Picart	Référent Santé au CGIET	19-févr
	Fabrice Mattatia		19-févr
INVS	Anne Brant-Bisson	Directrice générale adjointe	21-févr
INVS	Javier Nicolau	Chef du projet informatique	21-févr
INVS	Anne Doussin	Directrice du département des maladies chroniques	21-févr
INSEE	Thierry Aouizerate	Adjoint chef du dpt info service	24-févr
INSEE	Pascale Breuil	Chef de l'unité des Etudes démographies et sociales	24-févr
INCA	Philippe-Jean Bousquet	Responsable du département Veille, Observation et Documentation	24-févr
INCA	Fabien Calvo	Directeur général adjoint, directeur de la recherche	24-févr
CNSA	Juliette Bloch	Directrice scientifique	25-févr
CNSA	Jean-Luc Bernard	Directeur adjoint SI	25-févr
CNSA	Hamid Boukhlof	Directeur des SI	25-févr
DGS	Marie-Christine Favrot	Adjoint au directeur général, chef des politiques de santé	25-févr
DGS	Anne Périllat	Adjointe au chef du bureau du médicament	25-févr
DGS	Alain Fontaine		25-févr

Institution	Interlocuteur	Position	Date
ARSIF	Patrick Olivier	Directeur de la stratégie	10-mars
ARSIF	David Clavereau	Directeur de projet e-santé	10-mars
FHF	Yves Gaubert	Responsable pôle finances	28-févr
DSSIS	Philippe Burnel	Délégué à la stratégie des systèmes d'information en santé	05-mars
CNAMTS	Claude Gissot	Directeur de la stratégie, des études et des statistiques	11-mars
CNAMTS	Hélène Caillol		11-mars
ANAP	Christian Anastasy	Directeur	11-mars
ANAP	Gilles Bontemps	Adjoint directeur	11-mars
DSS	François Godineau	Adjoint au directeur	05-mars
DSS	Abraham Hamzaoui	Chef de bureau des SI	05-mars
Malakoff Mederic	Laurent Borella	Directeur santé	19-mars
Malakoff Mederic	Laurence D'haultfoeuille		19-mars
DGOS	Gaël Raimbault	adjoint sous directeur pilotage	26-mars
DGOS	Yannick Leguen	sous directeur pilotage	26-mars
FHP	David Castillo	Direction de la prospective économique, médicale	20-mars
FNORS	Alain Trugeon	ORS Picardie	31-mars
ARS PACA	Norbert Nabet	Directeur Général adjoint	08-avr
Réseau Quêtelet	Roxanne Silberman	Secrétaire Générale	14-avr
DGOS	Franck Jolivaldt	Mission SI des acteurs de l'offre de soins	26-mars
ARSIF	Laurent Trevillier	Pôle SI	10 mars et 18 avril
ARSIF	David Clavereau	Chef de projet e-santé	10 mars et 18 avril

Les compte-rendus des ateliers se trouvent dans les documents joints :



[Annexe 8b : Compte-rendu des ateliers thématiques – Hôpital.pptx](#)



[Annexe 8c : Compte-rendu des ateliers thématiques - Consommation de soins.pptx](#)



[Annexe 8d : Compte-rendu des ateliers thématiques - Offre de soins.pptx](#)

Annexe 8 bis

Synthèse de la consultation publique



[Synthèse de la consultation publique.pdf](#)

Annexe 9

Rapport du groupe de travail « risques de réidentification »

Conclusions du groupe de travail sur le « Risque de réidentification » dans les bases de données médico-administratives (pour la Commission open data)

A la suite du rapport Bras⁸ qui lui a été remis le 3/10/2013, la ministre, Madame Marisol Touraine, a demandé à Franck Von Lennep, directeur de la recherche, des études, de l'évaluation et des statistiques (DREES), de diligenter une expertise technique sur la sécurité des données concernant le risque de ré-identification des personnes à partir de données [supposées] anonymes.

L'étude a été réalisée de novembre 2013 à avril 2014 par un groupe de travail animé par André Loth (DREES) et composé de : Alireza Banaei et Max Bensadon (ATIH), Hélène Caillol (CNAMTS), Nora et Frédéric Cappens (Institut Telecom Bretagne), Emmanuelle Denis (DSS), Françoise Dupont (INSEE et CASD), Noémie Jess (DREES), Jean-Pierre Le Gléau (expert), Grégoire Rey (INSERM-CépiDc).

Y ont également contribué (pour le test de jeux de données anonymes) Maxime Bergeat (INSEE), Dominique Blum (expert) et les équipes techniques du CASD.

Le comité de pilotage, présidé par Franck von Lennep, s'est réuni 3 fois. Ont participé à tout ou partie de ces réunions Max Bensadon (ATIH), Philippe Burnel (DSSIS), Anne Coat (ANSSI), Claude Gissot (CNAMTS), Nora et Frédéric Cappens (Institut Télécom Bretagne), François Godineau (DSS), Philippe Cuneo et Michel Isnard (INSEE), Jean-Pierre Le Gléau (expert), André Loth (DREES), Christian Saout (CISS) ainsi que Marc-André Beaudet (expert de la CNIL, en observateur).

1/ L'objet de l'étude

En substance la mission confiée au groupe de travail était

- d'identifier et évaluer les risques de ré-identification dans les bases médico-administratives considérées (SNIIRAM, PMSI...) notamment pour les jeux de données issus de ces bases et ayant donné lieu à une diffusion relativement large (données hospitalières du PMSI, échantillon généraliste des bénéficiaires) ou dont l'anonymat faisait débat (datamart des consommations inter-régimes de l'assurance maladie) ;
- d'établir une ligne de démarcation entre
 - o jeux de données anonymes (pouvant donc être mis en open data ou publication sans restriction)
 - o jeux de données présentant un risque de ré-identification et devant pour cette raison être en accès restreint...
- De recommander des moyens pour élargir, dans cette offre de données, la part et le volume des données en accès libre ;
- De préciser pour les données en accès restreint (parce que comportant des risques de ré-identification) sous quelles conditions techniques les personnes habilitées à accéder à ces données devraient pouvoir le faire afin de limiter le risque.

⁸ Rapport sur la gouvernance et l'utilisation des données de santé (sur les risques de ré-identification dans les bases de données médico-administratives, voir notamment pp. 26 à 30 et 56 à 58 du rapport).

Le groupe a entendu des chercheurs et parcouru la littérature sur ce sujet. Il a été surpris et déçu de constater qu'il n'existe pas vraiment encore de doctrine établie et que la plupart des responsables de bases de données, publiques et privées mais aussi les responsables des autorités de contrôle (les équivalents de la CNIL) en sont encore à tâtonner. Le groupe des 29 CNIL européennes vient ainsi de définir une liste de trois conditions suffisantes pour qu'un jeu de données puisse être qualifié d'anonyme mais ces trois conditions ne sont pratiquement jamais réunies et on est renvoyé dès lors à une évaluation au cas par cas⁹ : la veille technologique et scientifique sur ces questions devra être poursuivie demain, à l'initiative des gestionnaires des bases mais aussi dans les instances de gouvernance et dans les instances de contrôle afin que la doctrine s'affirme (et évolue quand c'est nécessaire).

2/ L'évaluation du risque de ré-identification dans le cas des données médico-administratives de santé

Il est entendu qu'il n'est question ici que des bases de données dé-identifiées ou « pseudonymisées », dites parfois, à tort, anonymes, ou de « jeux » de données issus de ces bases.

Le risque de ré-identification des personnes dans ces bases ne vient pas tant du pseudonyme utilisé¹⁰ que de la possibilité pour des tiers de reconnaître, dans des jeux de données, des personnes sur lesquelles ils disposent d'informations par ailleurs.

Classiquement (cf. le Modèle d'évaluation harmonisée du risque –MEHARI- du Club de la sécurité de l'information français), le risque, dépend à la fois :

- de la probabilité ou potentialité que l'événement se réalise, malgré des mesures dissuasives ou préventives (potentialité généralement évaluée sur une échelle de 1, faible, à 4, très fort) ;
- de l'impact (noté de 1 à 4 également) qu'a sa réalisation, compte tenu d'éventuelles mesures compensatoires ou réparatrices...

Risque = Potentialité x Impact

S'agissant de l'**impact** d'une divulgation non souhaitée de données personnelles de santé, le groupe a considéré qu'il était *très fort*, s'agissant de données classées comme sensibles dans le droit français et européen, et protégées par le droit constitutionnel à la vie privée. Des exemples de conséquences potentiellement très dommageables pour la personne concernée peuvent être cités (portant sur la carrière professionnelle ou l'image ou la vie de famille notamment).

Le fait que certaines personnes puissent rendre publiques elles-mêmes des informations sur leur propre santé ne change rien à l'évaluation de l'impact pour les victimes d'une divulgation non souhaitée. En outre, une fois l'information divulguée, le mal est fait et on ne peut ni remettre l'information dans la boîte ni considérer qu'une compensation financière réparera le préjudice.

Si on s'accorde à évaluer l'impact comme très fort, cela conduit à refuser en l'espèce une conception utilitariste de l'analyse bénéfice/risque où on admettrait que l'utilité économique pour le plus grand nombre compense le préjudice subi par quelques uns (supposés peu nombreux) dont la vie privée

⁹ <http://www.cnil.fr/linstitution/actualite/article/article/le-g-29-publie-un-avis-sur-les-techniques-danonymisation/>

¹⁰ Dans les bases SNIIRAM et PMSI, le pseudonyme est le NIR chiffré deux fois de manière irréversible. Une bonne organisation (par exemple des clés de chiffrement secrètes ou une table de correspondance, confiées à plusieurs personnes différentes extérieures à la base de données) évite le risque que le pseudonyme permette de remonter à la vraie identité de la personne, identité qui est parfaitement inutile pour les traitements ordinaires. Dans les cas où remonter à l'identité des personnes est nécessaire (pour rendre possible des appariements avec d'autres jeux de données, d'enquête par exemple), le principe de cloisonnement évite que les gestionnaires de la base connaissent les vraies identités. Ce principe de gestion séparée des données directement identifiantes (nom, NIR, numéro de téléphone...) est celui qui est posé par le projet de règlement européen sur la protection des données personnelles (art. 81 et 83).

serait divulguée. Les membres du groupe ont préféré une autre conception de l'analyse bénéfice-risques, consistant à maximiser le niveau de bénéfice pour un niveau de risque maîtrisé. Cela permet de conclure à l'intérêt :

- de rendre les données personnelles de santé accessibles – notamment aux chercheurs ou à des bureaux d'études –dans des conditions convenables (rapidité, assistance...)- mais avec l'accord de la CNIL et en contrôlant le risque de divulgation ;
- d'ajuster le niveau des précautions au niveau du risque ;
- et de maximiser la production et la diffusion des données anonymes issues -par divers procédés- des données individuelles brutes.

S'agissant de la **probabilité de ré-identification** et de l'efficacité de la dissuasion ou des mesures préventives, le groupe a considéré sans s'y appesantir que la diffusion de données agrégées issues des bases médico-administratives ne présentait pas en général de risque de ré-identification et qu'il n'y avait donc pas lieu d'en limiter l'accès à telle ou telle catégorie de personnes ou d'organismes comme c'est encore actuellement le cas, et qu'il y avait en outre certainement des moyens d'accroître la richesse et la pertinence des statistiques mises en ligne.

Comme on le verra, le groupe a également travaillé sur la possibilité d'étendre la production et l'usage de jeux de données individuelles -jeux exhaustifs ou échantillons- ayant pour caractéristiques que le risque de ré-identification y soit a priori inexistant et que ces données puissent donc elles aussi être publiées et réutilisées par tous.

Cependant, il lui était d'abord demandé de se prononcer sur la politique de diffusion des données médico-administratives individuelles par le ministère de la santé et les agences ou établissements publics placés sous sa tutelle. S'agissant des données individuelles à vocation exhaustives du SNIIRAM, du PMSI et de l'ensemble SNIIRAM-PMSI, il confirme l'évaluation concluant à un niveau élevé du risque :

- le risque que présente une diffusion mal contrôlée des données hospitalières du PMSI exhaustif (dans les conditions et avec les précautions insuffisantes¹¹ qui caractérisent son mode de diffusion actuel) ;
- le risque que cela entraîne pour l'Échantillon généraliste des bénéficiaires (EGB) au 1/97, accessible à un nombre important d'organismes, où on trouve les parcours de soins de 600 000 personnes sur (potentiellement) 20 ans. Cet échantillon a été malheureusement compromis par la diffusion trop large du PMSI lequel permet comme on l'a vu d'identifier des personnes hospitalisées et de les retrouver, avec les mêmes caractéristiques dans l'EGB ;
- le risque que présenterait un accès trop large au datamart de consommations inter-régimes (DCIR) exhaustif de l'assurance maladie, si par exemple il devenait ouvert à toutes les personnes ayant le statut de chercheurs, même sans possibilité de croisement des variables sensibles et même sans croisement avec le PMSI (ce qui lui retirerait d'ailleurs une bonne part de son intérêt).

Dans tous ces cas, le nombre de personnes présentant les mêmes caractéristiques au regard des quasi-identifiants (date en mois des soins, durée d'hospitalisation, arrêt de travail, âge en années...)

¹¹ Les calculs de D. Blum, confirmés par l'ATIH, donnent les ratios de 89 % et 100 % de séjours où une personne est seule à présenter les caractéristiques recherchées (respectivement sur l'ensemble des séjours de 2008 et sur les séjours des patients hospitalisés au moins 2 fois dans l'année). Ces calculs ont été effectués en 2011 sur une base présentant le même type de « floutage » que les bases diffusées au cours des années récentes (imprécision volontaire sur la date de sortie et code géographique correspondant à au moins 1000 habitants). La diffusion à plusieurs centaines d'exemplaires de copies de cette base complète, sous la forme de disques, la diffusion systématique du fichier de chaînage des séjours, l'existence avérée de copies sauvages (et l'intérêt d'en obtenir en raison de leur valeur marchande et pour éviter des procédures et un paiement), l'impossibilité pratique d'empêcher ces copies ou de contrôler le respect des engagements pris par les titulaires des autorisations de la CNIL ont été les principaux arguments cités.

est très petit, tendant rapidement vers la valeur 1 si les données sont chaînées dans le temps : **tous les parcours de soins sont différents**.

Si on prend l'exemple de l'INSEE, celui-ci ne prend pas le risque de diffuser un fichier exhaustif présentant ce type de risques autrement que via un dispositif sécurisé à des personnes dont le besoin de connaître les données a été préalablement vérifié et évalué (c'est également ce qui est prévu pour les données fiscales).

Dans cette évaluation de la probabilité de ré-identification, il convient de distinguer trois temps :

- ✓ Le *comptage* du nombre de « personnes dans une case » (personnes présentant les mêmes caractéristiques au regard des quasi identifiants que sont l'âge, le sexe, le lieu de résidence, le lieu, la nature et les dates de soins, la date de décès éventuellement). Il présente un caractère objectif. Ce nombre est fortement réduit jusqu'à devenir vite égal à un pour toute la population dès que les données relatives à une même personne sont chaînées¹² ;
- ✓ Le *classement* des quasi-identifiants par degrés de notoriété n'est guère contestable non plus. Les dates de naissance et de décès par exemple sont aisément connues de tous ; les dates et lieu d'hospitalisation ou les dates d'arrêt de travail restent assez faciles à connaître, même si pour une personne donnée, le nombre de personnes susceptibles de les connaître est généralement petit (proches, employeur, assureur). En revanche, les dates de consultations médicales ou de soins par des auxiliaires médicaux sont plutôt mal connues et vite oubliées ;
- ✓ L'*estimation* de la probabilité que ces informations donnent lieu à une divulgation par les personnes qui en ont eu connaissance est plus subjective (« à dire d'experts ») :
 - cette probabilité augmente évidemment avec le nombre de personnes mises dans le secret, directement ou indirectement, et dépend donc de la procédure de sélection des demandes d'accès ;
 - elle est réduite avec l'emploi de dispositifs techniques permettant le confinement des données, la traçabilité des accès et des requêtes et l'interdiction des copies, alors qu'au contraire la diffusion des données au moyen de copies et d'extractions augmente la probabilité de divulgation par négligence et/ou malveillance ;
 - elle est réduite aussi avec le caractère dissuasif des sanctions pénales pour divulgation illicite, que le groupe a jugé
 - assez efficace dans le cas d'institutions telles que les assurances (secret de pratiques illicites difficile à garder, souci d'image...)
 - mais peu efficace devant d'autres « attaquants » (par exemple adversaires politiques, malveillance dans un cadre familial ou professionnel, presse à scandale, démonstration qu'on peut le faire ou simple curiosité...) surtout si la probabilité d'être pris est très faible (pas de traçabilité pour les données extraites). On notera que les cas de divulgation des données peuvent alors rester confinés à un petit cercle.

En tout état de cause, l'argument « jusqu'ici tout va bien » ne dispense pas, selon le groupe de travail, de se conformer aux standards professionnels en vigueur et ne peut justifier la continuation du mode actuel de diffusion des données du PMSI et de l'EGB, ni l'ouverture incontrôlée des accès au DCIR, qui sont porteuses de risques pour les personnes concernées et qui seraient de nature, en cas d'incidents publics, à saper la confiance dans le système d'information.

Il appartient à la CNIL de faire savoir quelle sera désormais son attitude concernant la diffusion des données du PMSI, mais le groupe de travail recommande au ministère de la santé de revoir sa politique de diffusion des données qui est jugée trop restrictive dans le cas des données agrégées que chacun s'accorde à classer comme anonymes et trop laxiste dans le cas des données du PMSI.

Pour mémoire, le groupe de travail ne s'est pas penché sur le cas des usages opérationnels des données (par exemple dans les caisses d'assurance maladie pour les feuilles de soins ou dans les

¹² Même sans chaînage des données, certains cas facilitent la réidentification des personnes (patients soignés loin de chez eux, décès...).

ARS pour les données du PMSI) : ces usages sont ceux qui ont justifié initialement le recueil des données et sont donc par hypothèse nécessaires au service public. Le groupe rappelle seulement que dans ces cas aussi l'accès aux données doit être encadré (personnes habilitées individuellement et tenues au secret professionnel¹³, tenue de registres, accès sécurisé...).

3) Une réponse graduée selon le risque de ré-identification

L'offre que le groupe de travail propose de faire à ce stade (travaux à poursuivre, en liaison avec les services de la CNIL et la future instance de gouvernance s'articule autour de trois axes :

a) une offre pour tous les publics en accès direct sur internet

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. Le groupe n'a pas exploré cette piste faute de temps mais elle ne pose pas de problème de principe : seulement de faisabilité technique et de coûts. La limite de l'exercice est que la demande est potentiellement infinie, alors que la production de statistiques agrégées a un coût 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 ré-identification » 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 ($K=10$ au moins dans le test) et que parmi ces personnes il y en ait toujours un nombre suffisant ($L=3$ 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).

Les fichiers qui seront mis à disposition de cette façon doivent être absolument anonymisés, pour tout type de consultant qui n'a d'ailleurs pas besoin de se faire connaître (assureur, employeur, voisins, famille...).

Le groupe de travail est parti du fichier PMSI et a construit, par généralisation ou suppression de variables, des jeux répondant aux critères d'anonymisation : pour toute combinaison de variables quasi identifiantes, il y a au moins 10 individus. Parmi ces individus, il y a au moins 3 types de « maladies » différentes.

À titre d'exemple, on a construit un fichier comprenant pour chaque séjour d'hospitalisation : le sexe et l'âge (en 18 groupes d'âge) de la personne hospitalisée, son mode d'entrée et de sortie

¹³ Le groupe rappelle qu'il n'est pas besoin d'être médecin pour respecter et faire respecter le secret professionnel sur des données personnelles de santé.

(en 2 groupes : « domicile » ou « autre ») et la durée d'hospitalisation (12 modalités) et le groupe homogène de malades (indicateur de la maladie).

Un autre fichier donnant un peu moins de précision sur la durée d'hospitalisation (2 modalités), mais comprenant la région de résidence du malade semble aussi répondre aux critères. D'autres exemples ont été étudiés et un grand nombre de jeux répondant aux critères d'anonymisation peut a priori être mis à disposition de la même manière,

Des fichiers analogues pourront, selon la même méthode, être construits à partir du SNIIRAM.

Le groupe ne donne pas de recommandation définie sur le nombre minimal de personnes présentant les mêmes caractéristiques que doit comporter un jeu de données pour être considéré comme anonyme (k-anonymat). Certains membres du groupes préconisaient d'examiner les seuils $K=3$ et $L=3$ qui préservent en théorie l'anonymat en admettant qu'il serait peut-être nécessaire de se fixer sur un seuil un peu plus élevé ($K=5$, $L=3$ par exemple). Les travaux devront être poursuivis à ce sujet, car certains membres du groupe considéraient à l'inverse que ces valeurs sont trop basses pour garantir l'anonymat.

Il est mis à l'étude aussi la proposition que des fichiers plus détaillés sur les caractéristiques médicales du séjour (médicaments en sus, DMI, diagnostics et actes...) soient fournis mais sans indication du nom de l'établissement et sous la forme de sondage (au 1/10 voire 1/3). Certes, les données médicales ne sont pas considérées comme un quasi identifiant (c'est l'information que « l'assaillant » est supposé *rechercher* dans la base : s'il en disposait déjà, il n'aurait pas besoin de la base) mais elles doivent être appauvries elles aussi dans les jeux de données anonymes parce que la partie médicale du fichier de résumé des sorties anonymisées (RSA) est très *discriminante* (chaque RSA est à cet égard différent des autres et on pourrait, si on conservait cette information commune dans des jeux de données par ailleurs appauvris reconstituer le jeu d'origine).

La technique de sondage confère en tout état de cause une sécurité supplémentaire et peut être appliquée partout où un fichier exhaustif n'est pas indispensable, à la condition bien sûr qu'on ne parte pas d'un fichier où chaque personne a déjà des caractéristiques uniques et que la règle de tirage soit gardée secrète..

Le caractère anonyme de ces fichiers, notamment l'impossibilité de remonter au fichier d'origine, devra être validé par des instances à désigner (on pense notamment à un avis de la CNIL sur le mode de construction de ces fichiers et non sur chaque fichier pris isolément).

La piste du « bruitage » (ajout ou substitution de données rendant « fausses » les données individuelles mais conservant des valeurs telles que moyenne, médiane, écart type etc.) n'a pas été suivie, notamment parce qu'elle nous a paru supposer une utilisation prédéterminée du jeu de données alors qu'en open data, les usages sont à la main des utilisateurs. Elle suppose également une compétence forte des utilisateurs pour en maîtriser les limites d'utilisation.

b) une offre plus limitée, destinée à un public ciblé : un nouvel EGB ?

S'agissant des données présentant un risque de présomption d'identification, la proposition faite est de constituer au moins 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 soins¹⁴ : il conviendra qu'ils souscrivent un certain nombre

¹⁴ 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 elle présente (âge, sexe, parcours de soins) des caractéristiques uniques.

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 d'une instance technique de valider le niveau de risque. Dès lors, l'accès à tel ou tel organisme (chercheurs, organisations professionnelles, administrations) pourrait relever comme aujourd'hui d'un arrêté ministériel.

Un fichier spécifique pourrait être construit à partir du SNIIRAM, chaîné avec le PMSI. Ce fichier ne serait pas exhaustif, mais issu d'un sondage (par exemple au 1/10¹⁵). Il est important que le mode de sondage et les paramètres retenus pour celui-ci demeurent strictement confidentiels. Il sera alors impossible à quiconque de savoir si telle ou telle personne est présente dans ce fichier.

Si l'on découvre alors une personne ayant exactement les caractéristiques d'un individu (et d'un seul) présent dans ce fichier, on ne pourra pas en conclure à son identification, car :

- on ne sait pas si la personne visée est ou non présente dans le tirage ;
- il n'y a qu'une certaine probabilité (dans l'exemple, une chance sur dix) pour que la personne repérée dans le fichier soit la personne connue.

Cette remarque doit cependant être tempérée par le fait qu'une accumulation de ressemblances avec une personne connue pourrait, au fil de l'addition des millésimes diffusés, au fur et à mesure de l'enrichissement annuel de la base , conduire à une probabilité d'identification proche de l'unité.

C'est pourquoi, ce type de fichier devrait être :

- rendu imprécis dans certaines de ses dimensions (telles que dates et lieu de soins, adresse, âge, date de décès ?) ce qui le rendrait inutilisable pour le suivi précis des parcours de soins mais très utile pour des études de pharmaco-épidémiologie par exemple...
- mis en accès limité, avec identification de la personne qui y a accès et engagement de sa part à n'utiliser le fichier qu'à des fins déclarées à l'avance, lesquelles n'incluent pas la tentative d'identifier une personne (modèle du réseau Quetelet ou de l'accès actuel des chercheurs INSERM aux données de l'EGB).

c) une offre restreinte pour les fichiers permettant une identification indirecte des malades

Comme cela a été vu, certains fichiers, utiles pour la recherche, sont extrêmement détaillés et permettent une identification indirecte de nombreuses personnes.

Les fichiers PMSI et EGB ayant été assez largement diffusés, ils ne devraient dorénavant être disponibles que moyennant des procédures d'accès très strictes.

Les personnes demandant un accès doivent justifier de la raison qui leur fait effectuer cette demande. Elles doivent démontrer les garanties de sérieux pour elles-mêmes et pour leur environnement professionnel. Au moment de leur accès, une trace des opérations effectuées doit être conservée. Enfin, un organisme doit vérifier que les sorties effectuées à partir de ces fichiers ne mettent pas en danger la protection de la vie privée des personnes.

Des systèmes et des procédures existent qui permettent ce type d'accès à des données confidentielles sans que cela implique un site unique.

Mais si ces conditions sont réunies et qu'on admet que le fichier est indirectement¹⁶ nominatif, il n'y a sans doute pas lieu d'appauvrir les données fournies : la distinction actuelle entre les

¹⁵ Un échantillon au 1/10 a été suggéré dès 2006 dans un rapport des Professeurs Bégaud et Costagliola.

¹⁶ Bien sûr, puisque les traitements autorisés visent exclusivement à tirer des conclusions générales à partir de données individuelles, il n'y a pas lieu de mentionner dans les données traitées les données directement identifiantes (nom, NIR, adresse etc.).

personnes autorisées ou non à croiser les quatre données sensibles (dates de soins, mois de naissance, commune de résidence et date de décès) n'a a priori plus de raison d'être.

Ce système à trois niveaux devrait permettre de présenter une offre aussi vaste que possible, tout en préservant la confidentialité des données.

Annexe 10

Contributions des membres de la commission

Contribution à la commission open data en santé transmise par Chantal Cases (INED), Marcel Goldberg (INSERM/UVSQ), Yann Bourgueil (IRDES), Jean Paul Moatti (INSERM/AVIESAN), Grégoire Rey (INSERM/AVIESAN), Pierre-Yves Geoffard (CNRS-PSE Ecole d'économie de Paris)

La nécessaire séparation de la gouvernance et de l'expertise

Le projet de rapport de la Commission Open data propose, dans la partie consacrée au dispositif de régulation des accès au SNIIRAM-PMSI concernant les accès ponctuels à des données porteuses de risque de réidentification liés à des projets, un modèle d'organisation des accès comprenant deux structures séparées : une structure de gouvernance chargée d'élaborer la doctrine et de veiller à sa déclinaison opérationnelle, et une instance technique chargée de l'évaluation des demandes en liens avec des projets spécifiques.

La première, dont le rôle serait de fixer les orientations et les conditions de mise en œuvre des accès aux données de santé doit être composée de l'ensemble des parties prenantes qui ont une légitimité à faire valoir leurs intérêts : professionnels et établissements de santé, organismes gestionnaires des bases de données, chercheurs, industriels, représentants des patients, organismes d'assurance maladie complémentaire, État.

La seconde serait une instance scientifique et technique, composé d'experts des différents domaines disciplinaires concernés, chargée d'instruire les demandes d'accès en la plaçant sur le seul champ de l'expertise au regard d'un référentiel d'analyse élaboré dans le cadre d'orientations établies par l'instance de gouvernance.

En tant que chercheurs concernés par l'utilisation du SNIIRAM-PMSI ou d'autres bases de données individuelles en santé, nous soutenons fortement cette proposition de séparation formelle entre gouvernance et expertise, qui correspond aux pratiques usuelles en matière d'évaluation scientifique adoptées de façon universelle. A notre sens, ces pratiques sont également adaptées aux demandes relevant de ces accès ponctuels qui relèveraient d'autres opérateurs que des chercheurs. Nous soutenons également la mise en place de procédures simplifiées quand il s'agit de méthodologies de références, pour des données à faibles risques de ré-identification.

Cette séparation est d'autant plus justifiée lorsqu'il s'agit de travailler sur des données de santé. En effet, les données de santé à caractère personnel sont des données sensibles, et à ce titre leur traitement est par principe interdit. Seule une finalité d'intérêt collectif justifie les dérogations prévues par la loi 78-17 du 6 janvier 1978 modifiée fixant les règles applicables en matière de traitement des données individuelles. Il est donc indispensable de s'assurer que les projets s'appuyant sur l'accès à ce type de données offrent *a priori* des garanties suffisantes de bénéfice collectif, tant au niveau des objectifs de connaissance que de pertinence des traitements au regard de la finalité poursuivie, des compétences scientifiques du demandeur et de sa capacité à respecter un référentiel de sécurité et de préservation de la confidentialité des accès que de l'engagement à rendre publics les résultats obtenus. Notamment, la qualité méthodologique du projet doit être bonne, car une investigation mal faite peut conduire à des décisions erronées qui peuvent avoir un impact profond sur la santé des personnes. La qualité des projets a donc une dimension éthique qui doit être prise en compte, et un projet qui ne correspondrait pas à ces critères ne saurait justifier l'accès à des données de santé à caractère personnel, et deviendrait contraire aux règles de l'éthique scientifique et du contrat implicite passé entre la communauté

scientifique, les sujets sur lesquels porte l'investigation et le reste de la société : cette dernière accepte que des investigations de plus en plus poussées soient menées sur ses membres à la condition qu'un mécanisme de contrôle existe permettant de veiller à ce que les résultats des études puissent bénéficier à l'ensemble de la collectivité.

Nous recommandons également d'envisager que les projets de recherche ayant fait l'objet d'une évaluation positives par des comités scientifiques indépendants, notamment dans le cadre d'appels à projets, ou s'inscrivant clairement dans le programme de recherches de structures labellisées par les Alliances publiques de Recherche et ayant fait l'objet d'une évaluation indépendante par le Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (qui succède à l'AERES) puissent faire l'objet d'une procédure simplifiée qui tienne compte de ces évaluations antérieures. Dans tous ces cas, le rôle de l'instance d'expertise scientifique et technique ne doit pas être de doublonner l'évaluation proprement scientifique de la qualité du projet, mais plutôt de vérifier que les données dont l'accès est demandé correspondent bien aux objectifs proposés et que les conditions de sécurité et les procédures fixées seront bien respectées.

Dans ce contexte, il est absolument indispensable que la phase d'évaluation scientifique soit tout à fait indépendante de la gouvernance générale des accès aux données individuelles du SNIIRAM-PMSI. En effet s'il est justifié que l'instance de gouvernance soit composée des parties prenantes, car celles-ci sont légitimes à faire valoir leurs intérêts, ceux-ci ne peuvent entrer en ligne de compte dans l'évaluation scientifique et technique.

Contribution de l'ITMO santé publique



[Contribution ITMO santé publique.pdf](#)

Contribution de l'UNPS



[Contribution UNPS.pdf](#)

Contribution du Syntec-numérique



[Syntec Numérique – Position paper.pdf](#)



[Syntec Numérique – Position paper – Annexes.pdf](#)

Contribution de la Fédération Hospitalière de France

La fédération Hospitalière de France (FHF) partage les objectifs annoncés d'ouverture des données de santé.

Elle considère que c'est un facteur majeur d'amélioration de la qualité et d'efficacité économique. Cependant de nombreux freins subsistent en particulier concernant les bases de données de l'assurance maladie et dans une moindre mesure celle de l'Atih.

Afin de garantir un accès large dans un cadre sécurisé, la FHF propose que l'ensemble des données de l'assurance maladie et de l'agence technique de l'information hospitalière soit géré par une institution indépendante qui aurait en charge l'ensemble des nomenclatures, classification et échelle de coûts pour la partie scientifique.

Cet organisme regrouperait les bases de données du SNIIR-AM et du PMSI.

Cette organisation serait garante d'une véritable ouverture de données dans un cadre sécurisé.

La mise en œuvre pourrait être réalisée à moyen constant par redéploiement de moyens techniques du SNIIR – AM et de l'Atih.

Contribution conjointe du LEEM et du SNITEM



[Contribution conjointe du LEEM et du SNITEM.pdf](#)

Contribution du SNITEM



[Contribution du SNITEM.pdf](#)

Contribution de la FEHAP



[Contribution FEHAP-Open Data en sante](#)

Contribution de la FHP



[Contribution de la FHP.pdf](#)

Contribution de l'IDS

Propositions d'amendements au pré rapport examiné par la commission open data en santé le 30 avril 2014

Après 6 réunions de la commission open data en santé, et après avoir écouté et lu les préoccupations exprimées par les différentes parties prenantes, un compromis est facilement accessible si ces dernières le souhaitent.

C'est la raison pour laquelle les propositions ci-dessous visent à avoir un caractère consensuel en permettant de bâtir une gouvernance des données de santé qui :

- est simple, transparente, cohérente, efficace et soucieuse des finances publiques
- permet l'expression des aspirations et des besoins en matière de données de santé de l'ensemble des parties prenantes.

1) L'importance de la place de la société civile au sein de la gouvernance des données de santé

L'ouverture des données de santé est un enjeu de société. La présence active de la société civile est donc essentielle non seulement pour animer le débat de société mais aussi pour constituer un des piliers d'une réelle démocratie sanitaire. La société civile doit en outre avoir accès aux données de santé car elle a démontré sa capacité à utiliser ce capital immatériel que constituent les bases de données pour améliorer le fonctionnement du système de santé, sans que cela porte atteinte à la vie privée des personnes. La gouvernance des données de santé – c'est-à-dire la supervision du dispositif d'ouverture des données et les autorisations d'accès – doit permettre l'atteinte de ces deux objectifs complémentaires.

2) la supervision du dispositif d'ouverture des données et les autorisations d'accès

Les amendements portent sur la deuxième option du pré-rapport examiné le 30/4 par la commission open data en santé.

En effet, comme cela a été dit par différents acteurs, la supervision du dispositif d'ouverture des données de santé et les autorisations d'accès, pour l'open data appelé « régulé » puisque concernant des données considérées comme pouvant être indirectement ré-identifiantes, doivent être rassemblées au sein d'une même entité juridique, l'IDS. C'est le scénario le plus simple, transparent, cohérent, efficace et soucieux des finances publiques. C'est aussi celui qui répond le mieux aux

aspirations de la société civile décrites ci-dessus, car il lui permet d'agir à la fois sur l'expression des besoins et les accès accordés. Enfin, sur un plan institutionnel, s'il y avait deux organismes distincts et indépendants, l'un chargé de la supervision du dispositif d'ouverture des données, et l'autre des autorisations d'accès à ces données, cela comporterait des risques évidents de conflits, et donc de blocages.

Pour autant, cette deuxième option décrite dans le pré-rapport examiné le 30/4 par la commission open data en santé doit être amendée pour répondre aux aspirations exprimées le 30/4 par toutes les parties prenantes.

2.1 Les premiers aménagements portent sur les critères d'évaluation des autorisations d'accès qu'il est proposé d'amender de la manière suivante (page 53 du pré-rapport examiné le 30/4):

- Le bénéfice attendu en matière de santé publique, de protection de la santé, de protection sociale et de gestion des services de santé, de médecine préventive, de diagnostics et de soins, et de connaissance scientifique (y compris dans les domaines économique et social) des traitements envisagés ;
- La pertinence des traitements au regard de la finalité poursuivie et le besoin effectif d'accéder aux données ;
- Les références données par le demandeur quant à ses compétences et à sa capacité à respecter un référentiel de sécurité et de préservation de la confidentialité des accès ;
- La transparence du demandeur sur les méthodes employées et les résultats obtenus.

Ces modifications sont apportées notamment afin d'être en phase avec l'évolution de la réglementation européenne, telle qu'elle résulte du projet de règlement européen, et éviter des contentieux futurs

2.2 Les autres aménagements portent sur les critères sur lesquels le comité technique rend son avis et la portée de son avis.

2.2.1 Les critères sur lesquels le comité technique rendrait son avis seraient :

- La pertinence des traitements au regard de la finalité poursuivie et le besoin effectif d'accéder aux données ;
- Les références données par le demandeur quant à ses compétences et sa capacité à respecter un référentiel de sécurité et de préservation de la confidentialité des accès ;

La composition du comité technique doit être conforme aux critères qu'il examine.

Les instances délibérantes de l'IDS (bureau et assemblée générale) examineront les deux autres critères, leur composition étant adaptée à cet examen, puisqu'elles rassemblent les acteurs de la santé, notamment la société civile :

- Le bénéfice attendu **en matière de santé publique, de protection de la santé, de protection sociale et de gestion des services de santé, de médecine préventive, de diagnostics et de soins**, et de connaissance scientifique (y compris dans les domaines économique et social) des traitements envisagés ;
- La transparence du demandeur sur les méthodes employées et les résultats obtenus.

2.2.2 La portée de l'avis rendu par le comité technique

L'avis du comité technique, sur les critères qu'il serait chargé d'examiner, serait rendu de manière totalement indépendante (y compris par le fait que chacun de ses membres établirait une déclaration de conflits d'intérêts) et transmis aux instances délibérantes de l'IDS. Ces dernières examinerait ensuite les critères qui relèvent de leur champ de compétence. Elles approuveraient ou non les demandes d'accès avant l'autorisation de la CNIL. Elles ne pourraient s'opposer à l'avis du comité technique. Elles pourraient soit faire leurs observations, soit les signaler dans le cadre de l'approbation d'accès qu'elles donneraient ou non au demandeur.

Les motifs d'approbation ou de refus d'accès seraient rendus publics.

Différents paragraphes du pré-rapport examiné le 30/4 par la commission open data en santé doivent donc être amendés pour tenir compte de cette proposition.

- 3) **Sur d'autres points, le pré rapport examiné le 30/4 par la commission open data en santé doit être amendé comme suit :**

Page 5 Introduction

(....)

Les membres de la commission ont partagé des attentes communes et fortes. Ce sont ces attentes que ce document retrace ; les préconisations, si elles ne recueillent pas toujours un accord unanime, visent à répondre à ces attentes ; c'est à l'aune de leurs résultats que leur mise en œuvre devra être évaluée.

Ce paragraphe doit être mis au point après les débats de la dernière réunion de la commission open data en santé

Page 23

(....)

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)....

Page 26

➔ 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, CCMSA et RSI ; CNSA ; INVS ; HAS ; ANSM ; ARS sur leur champ de compétence régional et sous l'autorité d'un médecin salarié de l'agence**).

Page 27

➔ 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 sont ouverts à **30** organismes.

3.5.2 L'expression des parties prenantes (page 52)

(...)

Les membres de la Commission insistent sur le fait **qu'une instance permettant l'expression des acteurs ne trouve** sa justification qu'à la condition qu'elle soit dotée d'une véritable capacité à **pouvoir agir à la fois** sur les orientations et **sur** la mise en œuvre des accès aux données de santé. **Pour ce qui concerne l'expression des besoins des acteurs**, 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. **Pour ce qui concerne les autorisations d'accès, afin de pouvoir agir réellement sur elles**, l'instance serait en charge de les approuver après avis indépendant d'un comité technique (cf. **selon les modalités décrites ci-dessus**)

3.5.3.2 Les accès liés à des projets (page 54)

(...)

Par ailleurs, la Commission est favorable à la demande émise de longue date **par les acteurs de la santé d'utiliser toutes les possibilités offertes par la loi du 6 janvier 1978 pour l'emploi du NIR** afin de 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**.

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 d'un décret du 9 novembre 2012. 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 circonscrit ». L'acte réglementaire unique ne peut donc porter sur une finalité définie de manière trop générale.

3.5.4 Un niveau opérationnel de mise à disposition des données (page 55)

La mise en place d'une organisation facilitant les accès est essentielle sauf à créer un droit virtuel. **L'ATIH et la CNAMTS responsables** d'organiser ces accès, doivent 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 **au sein de l'ATIH et de la CNAMTS** 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.

Ces prestations doivent être servies par un service dédié et identifiable **au sein de l'ATIH et de la CNAMTS** afin notamment de permettre à l'instance de gouvernance (décrise en 3.5.2) d'évaluer la qualité du service rendu et d'en publier les résultats.

Bien entendu, 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.

Contribution de l'UNOCAM

1. Open data en santé : un enjeu pour l'efficience de notre système de santé

L'UNOCAM soutient pleinement la démarche initiée en faveur de l'ouverture des données publiques de santé, y compris celles concernant les établissements et professionnels de santé, en raison des enjeux majeurs qui y sont associés (et exposés dans la partie 2 du rapport) :

- Renforcer la démocratie sanitaire
- Renforcer l'autonomisation du patient
- Accroître l'efficacité de l'action publique
- Favoriser la recherche, l'innovation et le développement industriel.

Pour les organismes complémentaires d'assurance maladie, l'open data contribuera à améliorer l'efficience globale de notre système de santé :

- l'accès aux données permet de construire un partage des connaissances entre acteurs du système de santé (AMO, AMC et offreurs de soins) et représentants de ces 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 efficience du notre système de santé.
- 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'efficience 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'efficience 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.

Pour les organismes complémentaires d'assurance maladie, l'amélioration de l'information du citoyen sur les professionnels et établissements de santé, les parcours de soins, les médicaments et les traitements, etc., est indispensable. Si le préalable consiste à travailler sur les besoins des citoyens, la pertinence et l'actualisation des données rendues publiques, la pédagogie nécessaire à leur diffusion, il ne doit pas constituer un frein à l'ouverture de ces données et leur réutilisation. 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. Accès aux données : les acteurs à pied d'égalité

Pour l'UNOCAM, si l'ouverture des données publiques (agrégées ou individuelles) à tous les acteurs, publics comme privés, est souhaitable, cette ouverture doit se faire dans le cadre de la protection de la vie privée. L'UNOCAM est favorable au renforcement de l'effectivité des sanctions, en cas de manquement à ce principe fondamental.

Dès lors que les données sont parfaitement anonymes, leur ouverture et leur réutilisation doivent être organisées à court terme. L'UNOCAM soutient les propositions faites en ce sens en ateliers.

Pour les données présentant un risque potentiel de réidentification, l'UNOCAM soutient 3 types de mesures complémentaires :

- premièrement, la poursuite des travaux relatifs à l'évaluation du risque, de façon à mesurer si le risque est réel et, le cas échéant, à caractériser son niveau ;
- deuxièmement, une action sur ces données, destinée à supprimer le risque de réidentification et permettre un premier niveau d'ouverture : production de tableaux statistiques, exclusion des données sensibles, micro-agrégation, floutage, renforcement des critères d'anonymat par le relèvement du niveau minimum de personnes présentant des caractéristiques identiques... ;

- troisièmement, la régulation de l'accès aux données exhaustives (le risque de réidentification ne devant pas entraîner la fermeture des accès alors que les besoins d'utilisation de données fines et complètes, par différents acteurs du système, sont réels). Le pré-rapport stipule qu' hormis les services de l'Etat (pour lesquels la décision relève de la compétence du Ministre de la santé), la décision d'autoriser tel ou tel organisme, de façon permanente ou de manière ponctuelle sur un projet, est fondée sur l'analyse de ses missions/de son projet, et sur les garanties apportées en termes de respect de la confidentialité, d'organisation des accès et le respect d'un référentiel de sécurité. Si l'UNOCAM adhère à ce principe général, elle souhaite qu'il soit assorti de règles claires et non discriminantes :
 - o Les organismes publics ou privés doivent être traités à égalité dès lors qu'il s'agit de demandes pouvant intéresser la santé publique ou la gestion du risque santé. Les conditions d'accès doivent donc, a priori, être identiques, sans discrimination selon le statut du demandeur.
 - o Pour l'examen des missions/projets, si un critère de bénéfice collectif devait être retenu, il ne pourrait être réduit à la seule amélioration de la connaissance scientifique ou de santé publique, mais être étendu à l'amélioration de la connaissance économique et sociale. La pluralité des critères concourant au bénéfice collectif doit être reconnue.
 - o S'agissant des garanties à apporter, un cahier des charges devra être établi à cet effet.

3. Une gouvernance simple et garante des principes d'ouverture des données de santé

Pour l'UNOCAM, la gouvernance de l'open data en santé doit reposer sur 2 principes :

- o La simplification des démarches et la lisibilité des process, comme élément de promotion de l'open data ;
- o La pluralité des points de vue, comme garant du respect des principes posés pour l'accès aux données de santé (cf. ci-dessus) ;

C'est pourquoi l'UNOCAM défend l'idée de renforcer le rôle de l'Institut des données de santé (IDS) dans la gouvernance de l'accès aux données de santé :

- La composition de l'IDS serait adaptée de façon à ce que l'ensemble des parties prenantes y soient représentées : Etat, opérateurs publics, caisses d'assurance maladie, organismes d'assurance maladie complémentaires, professionnels et établissements de santé, chercheurs, industriels, représentants des patients et des usagers, etc. ;

- L'IDS serait l'instance chargée de l'instruction et de l'autorisation des accès permanents et ponctuels dans le cadre de l'open data « régulé ». Dans un souci d'efficacité, l'IDS serait structuré autour de deux instances : un comité technique (chargé de l'expertise des dossiers, sur la base des principes et critères définis, transmise à la CNIL) et un conseil de surveillance (lieu d'échange et de partage de la connaissance entre acteurs et garant du respect des procédures d'octroi des accès aux données de santé). Cette organisation offre l'avantage d'avoir une structure unique et un interlocuteur unique.

En conclusion, l'UNOCAM soutient pleinement la démarche initiée en faveur de l'ouverture des données publiques de santé. Les travaux de la Commission posent de bons principes basés sur une réelle volonté de partage et de transparence, dans un cadre sécurisé mais aussi simplifié. Comme de nombreux acteurs, l'UNOCAM appelle à la concrétisation très prochaine de ces grands principes et au lancement des travaux qui en découlent. Elle considère que cette démarche contribuera à améliorer l'efficience globale de notre système de santé.

En conséquence, l'UNOCAM restera attentive à **deux questions majeures** associées à la mise en œuvre effective des intentions inscrites dans le projet de rapport de la commission : la gouvernance et le modèle économique retenu.

Sur **la question de la gouvernance**, l'UNOCAM considère qu'il est indispensable d'y associer de façon effective les composantes de la société civile dans toute leur diversité, dans une instance indépendante. Sans nier le rôle central de l'Etat et de la CNIL en tant que garants de la protection de la vie privée, la gouvernance doit être réellement partagée afin de reconnaître la pluralité des enjeux associés à l'ouverture des données de santé et d'évaluer, collectivement, les bénéfices et les risques associés à la mise à disposition des données, en définissant, le cas échéant, les règles à respecter pour les accès à des données avec risque de réidentification. C'est pourquoi l'UNOCAM défend l'idée de renforcer le rôle de l'Institut des données de santé (IDS) dans la gouvernance des accès.

Sur **la question du modèle économique**, l'UNOCAM estime qu'il est urgent de traiter des moyens et sources de financement alloués à l'ouverture des données de santé. De ce modèle dépend en effet l'effectivité de la mise en œuvre des principes posés par la commission, tant en termes d'accès immédiat à l'information et de réutilisation de celle-ci, que d'efficacité dans le process de mise à disposition des données aux acteurs intéressés (chercheurs, acteurs du système de santé, professionnels, usagers...).

CONTRIBUTION DE CHRISTIAN BAUDELOT ET CHRISTIAN SAOUT

Objet : Réflexions suite aux échanges lors de la réunion du 2 avril 2014.

Le pré rapport proposé aux membres de la commission open data en santé le 2 avril 2014 traduit des évolutions pour une plus grande ouverture des données de santé, quel que soit le statut juridique des acteurs. Le point de vue de la société civile, s'il n'est pas encore suffisamment pris en compte, semble avoir été pour partie retracé.

C'est ainsi qu'ont été bien distingués :

- l'Open Data en santé stricto sensu, c'est-à-dire les données qui devraient être accessibles directement, y compris par le grand public, car complètement anonymisées,
- l'Open Data que l'on peut appeler « régulé », qui devrait concerner uniquement des données anonymisées considérées comme pouvant éventuellement présenter un risque de ré-identification des personnes. Leur mise à disposition devrait faire l'objet d'une instruction individuelle. Il ne s'agit pas d'interdire leur utilisation, quel que soit le statut juridique du demandeur, mais de les mettre à disposition sous certaines conditions (voir plus bas).

Par ailleurs, tout au long des travaux de la commission, a aussi été mise en évidence l'existence de nouvelles technologies permettant d'éviter des risques éventuels de ré-identification des personnes en raison d'un éventuel acte malveillant répréhensible juridiquement. La micro-agrégation permet par exemple, à partir de données anonymisées considérées comme potentiellement ré-identifiantes, d'agréger des variables (mois et année de naissance en lieu et place du jour, mois et année de naissance par exemple) pour les rendre ainsi complètement anonymisées sans pour autant les appauvrir afin de répondre aux besoins. Ce travail doit pouvoir être engagé sans tarder afin de pouvoir mettre à disposition de tous les acteurs, y compris le grand public, dans le cadre de l'Open Data stricto sensu, un grand nombre de données anonymisées.

*

Des échanges qu'ont pu avoir postérieurement à la réunion du 2 avril dernier de nombreux membres, représentant la société civile, il ressort qu'il est nécessaire d'opérer encore des avancées.

A cet égard, le rapport de la commission devrait indiquer aussi clairement, en ce qui concerne l'Open Data « régulé » :

- Que les organismes publics et privés seront traités à égalité dès lors qu'il s'agit de demandes pouvant intéresser la santé publique ou la gestion du risque maladie.

- Que les demandes d'accès directs permanents ou ponctuels, ou d'extractions de données, relevant de l'Open Data « régulé », feront l'objet d'une instruction individuelle sur la base d'un cahier des charges (garanties apportées en termes de respect de la confidentialité, régulation des personnes habilitées, traçabilité des accès et des requêtes etc.).
- Qu'une instance de régulation élaborera le cahier des charges que la CNIL approuve.
- Qu'une instance de régulation approuvera les accès directs permanents et ponctuels, ou les extractions de données, relevant de l'Open Data « régulé », sauf dans deux types de cas :
 - les organismes en ayant une manifeste et impérieuse nécessité pour l'exercice de leurs missions (les services de l'Etat et ses opérateurs directs). La décision relèvera de la compétence directe du Ministre en charge de la santé,
 - les organismes de toute nature ayant besoin pour répondre à des demandes faites et validées par une autorité publique (HAS, ANSM, ...), ces demandes n'ayant pas lieu d'être réinstruites.
- Que l'instance de régulation comprendra un organe technique sur lequel elle s'appuiera en tant que de besoin pour élaborer le cahier des charges et instruire les demandes.
- Que tous les accès directs permanents et ponctuels, ou les extractions de données, relevant de l'open data « régulé », feront l'objet d'une autorisation de la CNIL, après l'approbation de l'instance de régulation (ou du Ministre en charge de la santé dans les cas qui relèvent de lui).

*

En ce qui concerne la gouvernance du dispositif, de nombreuses voix considèrent que pour mettre en œuvre de manière opérationnelle l'Open Data « régulé » la société civile doit y être pleinement associée, dans une gouvernance simple et centrée autour de l'Institut des données de santé. A cet égard, les avancées, qu'il faut reconnaître au terme des échanges du 2 avril dernier, resteront un vœu pieu si la gouvernance des accès reste dans un entre soi administratif ou dans une instance nouvelle dénuée de pouvoir et de compétences.

Ainsi, dès lors que les membres de la commission Open Data ont souligné tout au long des travaux la pertinence de l'action de l'Institut des données de santé depuis 2007 pour plus de connaissance, de partage, de transparence, de déontologie à travers la charte qu'il a élaborée, et d'équité entre les acteurs de la santé, c'est bien autour de cet institut qu'il faut consolider ce qui a déjà fait ses preuves et qui marche. L'Institut des données de santé ayant acquis à ce titre, expérience et légitimité.

Dès lors, conformément à la loi de 2004, l'Institut des données de santé doit être l'instance de régulation qui à la fois :

- instruit et autorise les demandes d'accès permanents et ponctuels, ou les extractions de données, relevant de l'Open Data « régulé »,
- et supervise le dispositif décrit ci-dessus et ses évolutions.

Ces deux missions ne sauraient être disjointes pour les confier à deux organes différents comme l'envisage le pré-rapport remis le 2 avril. Une telle gouvernance bicéphale est inefficace, source de confusions et opaque. A rebours, elles doivent être liées pour plus d'efficacité et de simplicité, et éviter des conflits futurs. Le pré rapport du 2 avril 2014 évite une source de conflits en repositionnant, conformément à la loi, les missions du COPIIR SNIIRAM. Il faut éviter de recréer une autre source de conflits potentiels. Un organe unique est aussi la solution la plus économique et la plus claire pour tous les acteurs.

Dans un tel schéma « unifié », la gouvernance et les moyens de l'Institut des données de santé peuvent bien entendu être revus afin notamment de distinguer en son sein :

- d'une part, une instance composée, comme aujourd'hui, des principaux acteurs, qui instruit et autorise les accès permanents et ponctuels dans le cadre de l'Open Data « régulé », avec en tant que de besoin l'avis d'un comité technique
- d'autre part, une sorte de Conseil de Surveillance, à créer, qui rassemblerait un panel le plus large possible d'acteurs et qui superviserait le dispositif d'accès aux données de santé.

En revanche, la mise à disposition des données relevant de l'Open Data « régulé » doit rester de la compétence des gestionnaires de ces bases, notamment de la CNAMTS et de l'ATIH qui ont toute la légitimité et les compétences requises.

Enfin, la mission du COPIIR SNIIRAM doit être recentrée sur la gestion et le renseignement de ce système d'information conformément à la loi de financement de la sécurité sociale pour 1999.

Contribution de l'UNPS

OPEN DATA ET DONNEES MEDICALES INDIVIDUELLES

Dans le cadre de l'Open Data, dans une volonté de faire vivre la démocratie sanitaire, se pose la problématique d'une publication de données de santé liées à l'activité personnelle des praticiens. Si sur le principe il est difficile de s'opposer à l'amélioration des pratiques médicales et donc à la publication de ces données, il existe toutefois un danger important dans ces publications tant au niveau initial par celui qui corrige les données médicales, que par l'utilisation qui peut en être fait par des tiers sans possibilité pour le praticien d'avoir accès ni aux sources, ni à l'utilisation ultérieure, avec un risque réel de publication ou d'utilisation erronée.

Il faut souligner que déjà l'interprétation des données du PMSI par la presse, conduit à l'élaboration de palmarès dont les critères sont parfois loin d'être pertinents, conduisant des établissements à être positionné d'une année à l'autre d'une manière radicalement différente en bien ou en mal sans possibilité de réel regard sous la pertinence et l'application des critères utilisés pour juger la qualité médicale de l'établissement. Une telle application au niveau individuel du médecin aurait des effets délétères inacceptables.

Il faut opposer la méthodologie de certification des établissements de la Haute Autorité de Santé, à celle de l'Assurance Maladie pour les médecins. Dans le cadre de la certification des établissements, les cadres administratifs et médicaux de l'établissement sont confrontés pendant cinq jours aux experts-visiteurs, puis un pré-rapport est élaboré par la H.A.S auquel l'établissement peut répondre point par point évitant ainsi toutes mauvaises interprétations, omissions ou malentendus. C'est une démarche remarquable qui permet alors une publication objective des données de qualité de l'établissement tant localement que sur les sites de l'HAS. Cette méthodologie a incontestablement conduit à améliorer l'organisation administrative et médicale des établissements.

En ce qui concerne l'Assurance Maladie, il n'en n'est pas de même puisque l'agrégation et l'interprétation des données se fait exclusivement au sein des services de l'Assurance Maladie, régulièrement lors des négociations conventionnelles nous sommes amenés à redresser des présentations qui nous sont faites et qui ne correspondent pas d'une manière exacte à la réalité de l'organisation au quotidien. Ceci pour de multiples raisons et en particulier la difficulté d'identifier par exemple la qualification du médecin fournissant les données. A titre d'exemple dans le domaine vasculaire, les actes peuvent être faits par des médecins spécialistes cardiologues, par des médecins angiologues, parfois par des médecins spécialistes en médecine générale, mais également des chirurgiens vasculaires ou des radiologues. Il est alors difficile pour l'Assurance Maladie de préciser avec pertinence qui est responsable de la pratique de tel ou tel acte. Les conséquences sont minimes puisqu'il s'agit de données collectives et qu'il n'y a aucune publication vis-à-vis de l'extérieur ou simplement parfois des présentations générales dans le cadre de conférences de presse.

Une étape complémentaire a été faite avec la mise en ligne des honoraires sur le site Ameli-direct. Cette mise en ligne se fait le plus souvent à l'insu du praticien, le principe n'est pas attaquant, par contre sa pertinence est parfois très loin de la réalité. D'une part seul un certain nombre d'actes

figurent sur le site de l'Assurance Maladie, d'autre part, notamment en cas de changement d'organisation tarifaire, les erreurs sont nombreuses sans compter les bugs informatiques ou les tarifs ont été multipliés par cent y compris en terme de remboursement, pour une difficulté de maîtrise informatique du monde Apple. Nous sommes très loin du recueil et de l'utilisation conjointe établissement HAS.

L'autre élément à prendre en compte est l'utilisation par des tiers des données de l'Assurance Maladie déjà compliquée lorsqu'il s'agit de publier les tarifs. Cette situation pourrait devenir totalement inacceptable si les données de rémunération sur objectif de santé publique était prise en compte dans la volonté de l'Open Data. En effet la ROSP est basée sur un certain nombre de critères élaborés dans le cadre conventionnel mais qui doit prendre en compte la situation du praticien et de sa patientèle. Ces critères sont forcément différents en fonction de la patientèle avec des pertinences variées selon que le praticien exerce dans le 16^{ème} arrondissement de Paris ou en Seine-Saint-Denis. Des comparaisons hâtives entraîneraient forcément un préjudice individuel pour le praticien et ne contribuerait pas à la démocratie sanitaire. Nous sommes très loin dans le cadre de la ROSP, de la démarche de la H.A.S et surtout le praticien éprouve de grandes difficultés à faire corriger les données qui lui sont attribuées liées à la lourdeur administrative et technologique du recueil de l'Assurance Maladie.

Le dernier écueil à la publication de données de type ROSP est leur utilisation ultérieure par des tiers. Face aux difficultés des praticiens vis-à-vis de l'Assurance Maladie pour obtenir un correctif des données qui lui sont attribuées, leur utilisation par des tiers rendrait impossible toute mise à jour puisque l'Assurance Maladie elle-même ne serait plus maître de ceux qui utilisent les données qu'elle publie d'une manière erronée.

C'est donc avec réticence que nous voyons certaines volontés mises en place dans ce rapport, si les intentions de transparence ne sont pas contestables, le mode de réalisation reste particulièrement difficile l'exemple récent du dossier médical personnel, de son échec, nous laisse extrêmement interrogatifs sur la réalisation d'une publication des données médicales des praticiens.