

# L'ouverture des données de santé à la recherche en France

Bibliographie thématique

Juin 2023

Centre de documentation de l'Irdes

Véronique SUHARD

**Synthèses & Bibliographies**

Reproduction sur d'autres sites interdite mais lien vers le document accepté  
[www.irdes.fr/documentation/syntheses-et-dossiers-bibliographiques.html](http://www.irdes.fr/documentation/syntheses-et-dossiers-bibliographiques.html)

Sans être exhaustive, cette bibliographie essaie de présenter les principaux documents de référence sur l'ouverture des données de santé, selon les thèmes mentionnés dans le sommaire. Les références portent sur la période 1990-2021.

Les recherches ont été réalisées sur la base documentaire de l'Irdes, sur la banque de données en santé publique (BDSP) ainsi que sur Cairn.

Les références bibliographiques sont classées par date, puis par ordre alphabétique d'auteurs et/ou de titres.

Certaines publications sont disponibles gratuitement en ligne (voir mention à la fin de la notice).

Les publications payantes peuvent être consultées, sur rendez-vous, au [centre de documentation de l'Irdes](#) ou bien commandées auprès des éditeurs concernés. Des copies d'articles peuvent aussi être obtenues auprès des bibliothèques universitaires ([Sudoc](#)), de la [BIU Santé](#) ou de la [British Library](#).

Il ne sera, en revanche, **délivré aucune photocopie par courrier** des références signalées dans cette bibliographie.

## Sommaire

<b>Introduction .....</b>	<b>4</b>
<b>L'ouverture des données : un choix de politique publique .....</b>	<b>4</b>
<b>La diffusion des données de santé : de la création du SNIIRAM à la mise en place du Health Data Hub .....</b>	<b>10</b>
<b>FOCUS : QUELQUES REPERES HISTORIQUES .....</b>	<b>10</b>
<b>EN FRANCE.....</b>	<b>11</b>
<b>A L'ETRANGER.....</b>	<b>25</b>
<b>L'utilisation des bases de données de santé par la recherche: les aspects méthodologiques .....</b>	<b>32</b>
<b>La protection des données personnelles dans le cadre de la recherche .....</b>	<b>39</b>
<b>Quelques liens utiles .....</b>	<b>50</b>

## Introduction

L'ouverture des données désigne l'effort que font les institutions, notamment administratives, qui partagent les données dont elles disposent. L'accès aux données vise d'une part à permettre aux citoyens de mieux contrôler l'administration, d'autre part d'exploiter ces données, ce qui implique que ce droit d'accès s'accompagne d'un droit à la réutilisation des données. La loi n° 2016-1321 pour une République numérique a été définitivement adoptée le 28 septembre 2016 et est entrée en vigueur le 9 octobre 2016. Elle vise à favoriser l'ouverture et la circulation des données et du savoir, à garantir un environnement numérique ouvert et respectueux de la vie privée des internautes et à faciliter l'accès des citoyens au numérique.

Dans le domaine de la santé, l'élaboration d'un système d'information recueillant des données de santé répond à trois types de besoins : le premier concerne la gestion du risque maladie qui est dévolue à l'Assurance maladie. Le second répond à celui du professionnel de santé dans le cadre de sa pratique et de son évaluation. Enfin, le troisième type de besoins est celui de la connaissance en santé publique et de la recherche.

L'accélération de l'informatisation a permis la constitution de nombreuses bases de données permettant de répondre à ces trois types de besoins. Les données collectés pour des fins administratives peuvent servir à des fins de recherche et d'évaluation. Le cas de la base SNIIRAM est significatif à cet égard. Cette dernière, créée par l'assurance maladie, sous le pilotage de l'État, pour ses finalités propres de gestion du risque, est progressivement devenue un outil essentiel pour la recherche, la connaissance et l'évaluation.

La première partie de cette bibliographie place dans le contexte plus large de l'ouverture des données numériques et des enjeux qui lui sont liées. La seconde partie concerne l'ouverture des données de santé à la recherche (hors secteur industriel), notamment celles contenues dans les bases médico-administratives, ces dernières constituant le socle du Health Data Hub. Certains aspects méthodologiques de l'exploitation de ces bases par les chercheurs sont ensuite explorés. Enfin la dernière partie couvre les aspects juridiques de l'ouverture des données de santé à la recherche

## L'ouverture des données : un choix de politique publique

Deroche, C. (2023). "Données de santé : une réforme encore en cours de chargement. Rapport d'information". Sénat, P.

<https://www.senat.fr/notice-rapport/2022/r22-873-notice.html>

La loi santé de 2019 relative à l'organisation et à la transformation du système de santé a porté plusieurs dispositions structurantes dans le champ des données de santé en :

élargissant considérablement le périmètre des données versées au système national des données de santé (SNDS) ; en substituant à l'institut national des données de santé (INDS) une plateforme des données de santé (PDS, ou Health Data Hub) chargée de favoriser et promouvoir leur exploitation ; en créant l'espace numérique de santé (ENS), devenu Mon espace santé (MES), portail personnel ayant vocation à réunir l'ensemble des données du patient. À l'issue de ses travaux d'évaluation, la mission d'information constate que les réformes ambitieuses portées en 2019, qui conservent leur pleine pertinence, peinent toutefois à trouver une mise en œuvre effective quatre ans après. Le présent rapport formule quinze recommandations afin de renforcer le pilotage politique de la stratégie en matière de données de santé, de répondre urgemment à des enjeux organisationnels et techniques et d'améliorer la connaissance et l'appropriation de ces dispositifs par les patients comme les acteurs du système de santé.

Baranes, E., Sauron, J. L., Hoag, M., et al. (2021). "Le règne des données. Dossier." Cahiers Français(419): 15-95.

Ce dossier porte sur les données numériques, leur production, leur utilisation et leur contrôle. Les auteurs analysent les principaux aspects des multiples enjeux : économiques, politiques, démocratiques... En effet, L'accumulation massive de ces données, le Big data, fait naître de nouveaux géants économiques mondiaux qui défient notre rapport aux libertés individuelles et à la souveraineté nationale. Mais qui produit ces données ? Qui les utilise ? Comment peut-on réguler leur production et leur collecte ? Quelles sont les spécificités de ces nouvelles entreprises de données ? Où en est-on dans l'ouverture des données publiques ? Comment l'État protège ses données sensibles ? Et enfin, les données numériques ainsi que l'intelligence artificielle vont-elles bouleverser l'humanité ?

Borgman, C. L. et Matoussowsky, C.(2020). "Qu'est-ce que le travail scientifique des données ?", Marseille : OpenEdition Press

<https://books.openedition.org/oep/14692>

Puisant ses analyses et ses exemples dans des champs scientifiques variés, cet ouvrage (dont l'original est paru en 2015 chez MIT Press) offre une étude inédite des utilisations des données au sein des infrastructures de la connaissance – utilisations qui varient largement d'une discipline à l'autre. Bien que le big data ait régulièrement les honneurs de la presse des deux côtés de l'Atlantique, Christine L. Borgman met en évidence qu'il vaut mieux disposer des bonnes données qu'en avoir beaucoup. Elle montre également que les little data peuvent s'avérer aussi précieuses que les big data, et, que, dans bien des cas, il n'y a aucune donnée, parce que les informations pertinentes n'existent pas, sont introuvables ou sont indisponibles... Au travers d'études de cas pratiques issus d'horizons divers, Christine L. Borgman met aussi en lumière que les données n'ont ni valeur ni signification isolément : elles s'inscrivent au sein d'une infrastructure de la connaissance, c'est-à-dire d'un écosystème de personnes, de pratiques, de technologies, d'institutions, d'objets matériels et de relations. Pour l'autrice, gérer les données et les exploiter sur le long terme requiert ainsi des investissements massifs dans ces infrastructures de la connaissance. L'avenir de la recherche, dans un monde en réseau, en dépend (4<sup>e</sup> de couv.)

Warsmann, J. L. et Latombe, P. (2021). "Rapport d'Information sur le thème "Bâtir et promouvoir une souveraineté numérique nationale et européenne ». 3 tomes". Paris Assemblée nationale.

[https://www.assemblee-nationale.fr/dyn/15/dossiers/batir\\_promouvoir\\_souverainete\\_numerique\\_mi](https://www.assemblee-nationale.fr/dyn/15/dossiers/batir_promouvoir_souverainete_numerique_mi)

Bothorel, E. (2020). "Pour une politique publique de la donnée". Paris Premier ministre.  
<https://www.vie-publique.fr/rapport/277879-pour-une-politique-publique-de-la-donnee>

La France a besoin de plus d'ouverture, sous toutes ses formes, ouverture des données publiques (open data), mais aussi partage et accès sécurisé aux données sensibles. La France s'est placée à l'avant-garde européenne de la politique de la donnée et des codes sources depuis 2013, mais l'avance acquise est fragile, alors que le potentiel de connaissance et d'innovation dans ce domaine est immense. Ce rapport propose des réformes ambitieuses pour participer aux transformations en cours au niveau européen, et aussi un grand nombre de mesures qui sont des actions réalisables sous un an, pouvant avoir des effets importants pour l'avenir.

Conseil national du numérique (2020). "Confiance, innovation, solidarité : Pour une vision française du numérique en santé". Paris CNNum.  
<https://cnnumerique.fr/files/uploads/2020/ra-sante-cnnum-web.pdf>

Ce rapport fait suite à la saisine de Cédric O en juillet 2019 sur la mise en œuvre d'une identité numérique publique, des dispositifs FranceConnect et de la carte nationale d'identité électronique. Le rapport du Conseil National du Numérique (CNNum) décrit précisément, via 35 recommandations actionnables, les conditions d'une confiance dans les outils de l'identité numérique et l'importance de l'intérêt du citoyen, afin d'enrichir les prochains travaux du Gouvernement sur les identités numériques.

(2019). "Tous acteurs des données. Appréhender les données pour mieux les valoriser", Paris : Renaissance numérique  
<https://www.renaissancenumerique.org/publications/tous-acteurs-des-donnees-apprehender-les-donnees-pour-mieux-les-valoriser>

La mise en données du monde est une réalité qui n'est pas toujours bien maîtrisée, y compris par ceux qui doivent lui en donner un cadre. Or, face à son accélération, son appropriation par tous - organismes publics, entreprises, société civile - devient un enjeu crucial pour le devenir de notre économie et de notre société. À l'heure où l'Europe entame la deuxième mandature du Marché unique numérique (Digital Single Market), avec un certain nombre de débats sur les données à son agenda, le think tank Renaissance Numérique et l'organisation professionnelle Syntec Numérique ont décidé de réunir leurs expertises pour offrir des clés de compréhension dans ce nouveau contexte.

Boul, M. (2018). "Réflexions sur la notion de donnée publique." Revue française d'administration publique **167**(3): 471-478.  
<https://www.cairn.info/revue-francaise-d-administration-publique-2018-3-page-471.htm>

Dans une société et une économie numérisées, les données sont au centre de problématiques majeures. Données publiques et données personnelles se répondent avec des encadrements juridiques croisés : la loi pour une République numérique du 7 octobre 2016 et le règlement général pour la protection des données « RGPD ». Pourtant la notion de donnée publique est difficilement saisie par le droit. Elle ne fait l'objet d'aucune définition. Information du service public, elle est concurrencée par de nouvelles notions. Bien au service du public, elle ne répond pas totalement aux définitions des biens publics et des facilités essentielles. Ainsi, ni le discours du droit ni le discours sur le droit n'ont réussi à appréhender cette notion.

Cluzel-Métayer, L. (2018). "La construction d'un service public de la donnée." Revue française d'administration publique **167**(3): 491-500.

<https://www.cairn.info/revue-francaise-d-administration-publique-2018-3-page-491.htm>

Créé par la loi « pour une République numérique », le service public de la donnée a été conçu comme l'élément structurant de ce que l'on peut appeler « l'infrastructure des données », ce service public est pour l'heure restreint à la mise à disposition des « données de référence ». Pour lui permettre de prendre de l'ampleur, le législateur l'a conçu comme une mission d'intérêt général évolutive, dont la gouvernance est publique, mais malgré tout ouverte et collaborative. Son régime juridique est à la fois exigeant et protecteur de cette infrastructure en construction.

Verdier, H. (2018). "La donnée comme infrastructure essentielle". Paris Dila.

[https://www.etalab.gouv.fr/wp-content/uploads/2018/04/RapportAGD\\_2016-2017\\_web.pdf](https://www.etalab.gouv.fr/wp-content/uploads/2018/04/RapportAGD_2016-2017_web.pdf)

La donnée est devenue un outil indispensable à la transformation de l'action publique et plus largement de l'économie. Pour en assurer la meilleure exploitation possible, il est nécessaire d'en garantir la qualité et la circulation. Au même titre que les réseaux de transport, d'énergie ou de télécommunications, une nouvelle forme d'infrastructure devient ainsi essentielle dans notre société moderne : celle de la donnée. Dans son deuxième rapport, remis au Premier ministre, l'administrateur général des données prône la construction de cette infrastructure, en s'appuyant notamment sur une analyse comparée des approches de plusieurs pays européens. Il y propose des pistes d'amélioration afin que l'État joue pleinement son rôle dans ce nouveau paysage des données.

Villani, C. (2018). "Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne". Paris Conseil National du numérique.

<https://www.vie-publique.fr/rapport/37225-donner-un-sens-lintelligence-artificielle-pour-une-strategie-nation>

Ce document identifie quatre secteurs prioritaires où la France doit particulièrement concentrer son effort de développement de l'intelligence artificielle : la défense, les transports, l'environnement et la santé. En matière de santé, l'intelligence artificielle ouvre des perspectives très prometteuses pour améliorer la qualité des soins au bénéfice du patient et réduire leur coût – à travers une prise en charge plus personnalisée et prédictive – mais également leur sécurité – grâce à un appui renforcé à la décision médicale et une meilleure traçabilité. Elle peut également contribuer à améliorer l'accès aux soins des citoyens, grâce à des dispositifs de prédiagnostic médical ou d'aide à l'orientation dans le parcours de soin. Près de vingt ans après la création de la base nationale des données médico-administratives (SNIIRAM), la France doit à nouveau faire figure de pionnière en investissant massivement dans les capacités de recherche et d'innovation en matière d'IA appliquée à la santé. Ces capacités reposent à la fois sur la mise en place de systèmes adaptés aux usages liés à l'IA visant à plateformiser le domaine, des procédures d'accès aux données plus fluides et un cadre pour expérimenter « en conditions réelles » les solutions émergentes. L'appropriation de l'IA en santé dépend également de notre capacité à accompagner les patients et professionnels de santé dans la transformation des pratiques médicales, ainsi qu'à structurer et animer en continu un débat national sur les usages éthiques de l'IA en santé

Bozio, A. et Geoffard, P. Y. (2017). "L'accès des chercheurs aux données administratives. Etat des lieux et propositions d'actions". Paris La Documentation française.

<http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/174000314.pdf>

Ce rapport, commandé par la secrétaire d'Etat chargée du numérique et de l'innovation, dresse un bilan, en forme de point d'étape, du mouvement d'ouverture des données administratives aux chercheurs. Le droit sur l'accès aux données administratives à des fins de recherche est complexe, mouvant, mais surtout très mal connu, des administrations productrices, comme des chercheurs. Le rapport recommande la réalisation d'un vade-mecum des dispositifs juridiques en vigueur. Le rapport fait le constat d'une absence large d'obstacles juridiques - avec quelques exceptions - à l'accès aux données par les chercheurs mais les difficultés pratiques restent notables. Le rapport préconise la création d'une instance de concertation impliquant les producteurs de données, les chercheurs, et les administrations compétentes pour la protection et l'accès aux données publiques (Archives, Cnil, administrateur des données, Cada). Le rapport rappelle également que l'accès aux données administratives a un coût (coût de support pour les instances, coût d'accès via des dispositifs sécurisé et coût de préparation de données anonymisées ou dé-identifiées). Le rapport étudie plusieurs pistes de modèle économique, tout en soulignant l'importance à limiter au maximum ces coûts pour éviter tout gaspillage de ressources limitées.

Cormont, A. et Goeta, S. (2017). "Ouvrir les données." *Statistique Et Société* 5(3)

<http://statistique-et-societe.fr/issue/view/72>

(2016) République numérique : qu'a changé la loi du 7 octobre 2016 ?. Paris ; vie publique

<https://www.vie-publique.fr/eclairage/20301-loi-republique-numerique-7-octobre-2016-loi-lemaire-quels-changements>

Le premier volet de la loi pour une République numérique (dite "loi Lemaire") vise à favoriser la "circulation des données et du savoir" à travers l'ouverture des données publiques et d'intérêt général, la création d'un service public de la donnée et le libre accès aux écrits de la recherche publique.

Bouchoux, C. (2014). "L'accès aux documents administratifs et aux données publiques. 2 tomes". Paris Sénat.

<http://www.senat.fr/rap/r13-589-1/r13-589-11.pdf>

<http://www.senat.fr/rap/r13-589-2/r13-589-21.pdf>

En 1978, le Parlement reconnaissait à toute personne le droit d'obtenir communication des documents de l'administration (loi Cada). Il ouvrait ainsi aux citoyens, en rupture avec la confidentialité caractéristique de la culture administrative française, la possibilité d'analyser et de comprendre les décisions publiques et de les contester. À l'heure du numérique, l'accès aux informations produites et recueillies par l'administration ainsi que la possibilité de les réutiliser prennent une dimension entièrement nouvelle, dont les enjeux sont à la fois stratégiques, démocratiques et économiques : la possibilité de réutiliser les données publiques ouvre en effet des perspectives encore largement inexploitées en matière de contrôle de l'action publique, d'amélioration de son efficacité et de sa qualité ou de développement de nouveaux services. La mission commune d'information a donc souhaité, d'une part, s'assurer de l'effectivité du droit formulé il y a 35 ans et, d'autre part, évaluer la

pertinence et l'efficacité des politiques engagées depuis quelques années par les pouvoirs publics en matière de diffusion de l'information publique et, plus récemment, d'ouverture des données publiques (open data) (tiré de la synthèse).

Baudot, P. Y., Marrel, G., Nonjon, M., et al. (2015). "Open et big data." *Informations Sociales* (191) <https://www.cairn.info/revue-informations-sociales-2015-5.htm>

Le mouvement d'ouverture des données administratives, appelé « open data » ou « données ouvertes », participe à la réforme de l'État. La notion de « big data » (« données massives »), souvent associée au terme d'open data, relève d'une toute autre démarche. Celle-ci vise à exploiter sous un angle nouveau des données créées à l'origine pour une finalité déterminée. Le big data désigne également le travail d'enrichissement des données par leur croisement, par exemple en appariant différentes sources statistiques pour lier l'observation de comportements d'usage et des données sur le profil socio-économique des personnes. Ce numéro de la revue Informations sociales cherche tout d'abord à comprendre les raisons de l'ouverture des données mais aussi à évaluer l'impact de ce mouvement sur les acteurs de l'action publique. Le soutien politique à cette démarche est motivé par la relégitimation des pratiques de gouvernement que permettrait l'ouverture des données, celles-ci rendant plus transparente l'action des pouvoirs publics. Les opérations d'informatisation nécessaires à la production de données de qualité impliquent une transformation des modalités d'organisation et des périmètres professionnels des administrations comme de leur contrôle. La spécificité du secteur social en matière d'ouverture des données publiques et d'usage des données massives peut ainsi être interrogée tant sur le plan juridique qu'économique.

Conseil d'Etat (2014). "Le numérique et les droits fondamentaux : Etude annuelle 2014 du Conseil d'Etat". Paris La Documentation française. <https://www.vie-publique.fr/rapport/34281-etude-annuelle-2014-du-conseil-detat-le-numerique-et-les-droits-fonda>

Le numérique, parce qu'il conduit à la mise en données et à la mise en réseau du monde, pose problème aux droits fondamentaux : il met en question leur contenu et leur régime. S'il renforce la capacité des individus à jouir de certains droits, comme la liberté d'expression, la liberté d'entreprendre, il en fragilise d'autres, comme le droit à la vie privée ou le droit à la sécurité. L'étude annuelle du Conseil d'État intervient alors qu'un triple basculement se manifeste dans les innovations techniques, dans l'économie et dans l'appréhension du numérique par la société. Face à ces bouleversements, l'étude s'attache à repenser la protection des droits fondamentaux et à répondre aux questions en débat : la neutralité d'internet, sa gouvernance, le « droit à l'oubli », la propriété des données, leur exploitation et agrégation en Big Data, le rôle inédit des grandes « plateformes ». L'étude présente 50 propositions de l'étude dont l'objectif est de mettre le numérique au service des droits individuels et de l'intérêt général.

Gorce, G. et Pillet, F. (2014). "Rapport d'information sur l'open data et la protection de la vie privée". Paris Sénat. <http://www.senat.fr/rap/r13-469/r13-4691.pdf>

La France s'est résolument engagée sur la voie de l'ouverture et du partage des données publiques, plus connue sous le nom d'open data. Deux idées animent cette politique.

Comptables de leur gestion auprès des citoyens, les administrations leur ouvrent leurs fichiers. Elles leur donnent ainsi le moyen de mieux les contrôler. Par ailleurs, à l'ère du numérique, où l'information est source de richesse, elles leur offrent l'opportunité d'exploiter le formidable gisement que constituent ces données. En créant une mission d'information chargée d'étudier l'open data et la protection de la vie privée de nos concitoyens, la commission des lois a souhaité poursuivre sa réflexion sur les nouveaux usages numériques et la façon dont ils peuvent se concilier avec les principes fondamentaux que le législateur a posés dès la fin des années 1970. L'open data soulève à cet égard une question spécifique : en principe, il exclut toute diffusion de données à caractère personnel, mais bien souvent, les données détenues par les administrations ont été élaborées à partir d'informations individuelles, qui peuvent être retrouvées grâce aux formidables capacités de traitement que permet l'informatique moderne. L'impératif de protection de la vie privée est-il en mesure de toujours prévaloir ? Comment s'en assurer ? À l'issue de leurs travaux, les rapporteurs de la mission d'information, les sénateurs Gaëtan Gorce et François Pillet jugent aujourd'hui nécessaire de faire d'une exigence fondamentale -la protection de la vie privée de nos concitoyens- une opportunité pour donner une nouvelle impulsion au déploiement de l'open data (résumé d'éditeur).

## La diffusion des données de santé : de la création du SNIIRAM à la mise en place du Health Data Hub

### FOCUS : QUELQUES REPERES HISTORIQUES

---

1977 : création du SNIR (Système National Inter-Régimes) par l'Assurance maladie. Il recense les activités des professionnels et des établissements au niveau national

1992 : Début du projet de Médicalisation des Systèmes d'Information (PMSI) qui décrit de façon synthétique et standardisée l'activité médicale des établissements de santé. Il repose sur l'enregistrement de données médico-administratives normalisées dans un recueil standard d'information.

1999 : création du Sniiram (Système national d'information inter-régimes de l'Assurance maladie) par la loi de financement de la sécurité sociale pour 1999

2002 : Arrêté du 11 avril 2002 relatif à la mise en œuvre du système national d'information inter-régimes de l'assurance maladie qui approuve le protocole de modalité de gestion et de renseignement du SNIIRAM

2004 : Migration du SNIR et première utilisation du SNIIRAM par la CNAM

2004 : Loi relative à l'assurance maladie prévoit la création d'un institut de données de santé (IDS)

2005 : Création de l'Echantillon généraliste des bénéficiaires (EGB) et mise en place de magasins de données agrégées

2010 : chainage effectif PSMI/SNIIRAM

2016 : [La loi n° 2016-1321 pour une République numérique](#)

2016 : L'[article 193](#) de la loi de modernisation de notre système de santé instaure le Système National des Données de Santé (SNDS)

2017 : L'IDS devient INDS un Groupement d'Intérêt Public et est défini comme le point d'entrée unique pour les demandes d'accès aux bases de données de santé, notamment au SNDS.

2018/03 : Parution du rapport Villani sur le développement de l'intelligence artificielle qui propose la création d'une structure centralisatrice des données de santé avec un élargissement des accès aux données

2018/10 : Mission de préfiguration du Health Data Hub

2019 : [La Loi du 24 juillet 2019 relative à l'organisation et la transformation du système de santé](#) crée Health Data Hub, plateforme nationale des données de santé

2021 : parution du décret d'application décret n° 2021-848 du 29 juin 2021 relatif au traitement de données à caractère personnel dénommé « système national des données de santé qui consacre le Health data Hub comme responsable de traitement du SNDS au même titre que l'assurance maladie

---

## EN FRANCE

Health data Hub (2023). "Health data Hub. Rapport annuel 2022 et feuille de route 2023". Paris Health data hub.

<https://health-data-hub.fr/actualites/le-health-data-hub-publie-son-rapport-annuel-2022>

(2023). "Données de santé en France : état des lieux et enjeux des bases de données, rôle de l'Assurance Maladie" Paris : Cnam

<https://assurance-maladie.ameli.fr/presse/2023-04-20-dp-donnees-de-sante>

Ce document fait le point sur l'organisation des grandes bases de données de santé en France, leurs usages et leurs bénéfices, les différents acteurs y ayant accès selon différentes procédures.

Vie Publique (2023). " Données numériques de santé : entre enjeux médicaux, technologiques et juridiques. Paris : Vie Publique

<https://www.vie-publique.fr/eclairage/289281-donnees-numeriques-de-sante-quels-enjeux-pour-quel-progres-medical>

Les données numériques en matière de santé ont connu récemment un développement exponentiel. Protection des données personnelles, exploitation de ces données par les professionnels de santé ou par des systèmes d'intelligence artificielle, développement du dossier médical partagé et de l'espace numérique de santé... les enjeux sont multiples.

Doutreligne, M., et al. (2022). "Entrepôts de données de santé hospitaliers en France . Quel potentiel pour la Haute Autorité de santé ?". <https://hal.science/hal-04013447>

Les données de vie réelle suscitent un intérêt croissant des agences sanitaires à travers le monde, que ce soit pour étudier l'usage, l'efficacité et la sécurité des produits de santé, suivre et améliorer la qualité des soins, réaliser des études épidémiologiques ou faciliter la veille sanitaire. Parmi les différentes sources de données, les entrepôts de données de santé hospitaliers (EDSH) connaissent actuellement un développement rapide sur le territoire français. Dans la perspective de mobiliser ces données dans le cadre de ses missions, la Haute Autorité de santé (HAS) a souhaité mieux comprendre cette dynamique et le potentiel de ces données. Elle a initié en novembre 2021 un travail de recherche visant à dresser un état des lieux des EDSH en France. Ce rapport pose d'abord le cadre et des définitions en s'appuyant sur la littérature. Il détaille ensuite la méthodologie de recherche, fondée sur des entretiens menés auprès d'acteurs impliqués dans les EDSH de 17 CHU et 5 autres établissements hospitaliers. Le résultat de ces entretiens est structuré par thématiques : historique, gouvernance, données intégrées, usages couverts, transparence, architecture technique et qualité de la donnée. Ce rapport discute ensuite les points d'attention identifiés pour le bon développement des EDSH et des usages secondaires des données. Il ébauche enfin deux cas d'usages pertinents pour la HAS.

Floyrac, A., et al. (2022). "Cartographie interactive de l'utilisation du Système national des données de santé: une base de données française unique à forte valeur ajoutée pour la recherche et l'innovation en santé." *Revue d'Epidémiologie et de Santé Publique* **70**: S38. <https://www.sciencedirect.com/science/article/pii/S0398762022000165>

**Introduction** La base principale du Système national des données de santé (SNDS) réunit les données issues des remboursements de l'Assurance maladie, des prises en charge hospitalières et des causes médicales de décès. Afin de caractériser son utilisation pour la recherche et l'innovation en santé, une cartographie interactive a été réalisée permettant d'appréhender le cadre de son utilisation et d'identifier les acteurs de l'écosystème SNDS. **Méthodes** Une recherche bibliographique via PubMed a été réalisée afin d'identifier les articles scientifiques publiés entre janvier 2007 et octobre 2021 sur des études utilisant les données du SNDS. La liste des auteurs, leurs affiliations, les mots clés et le nombre de citations ont été collectés. Une analyse descriptive a été réalisée afin d'évaluer les tendances temporelles et géographiques de l'utilisation des données du SNDS. **Résultats** Les résultats de la cartographie sont présentés sous la forme d'un dashboard interactif. Au 29 octobre 2021, 954 publications ont été identifiées. L'analyse descriptive a montré une augmentation importante du nombre de publications par an qui est passé en moyenne de 35/an entre 2010 et 2016 à 105/an entre janvier 2016 et octobre 2021. La discipline médicale la plus représentée est la pharmacoépidémiologie et les deux domaines les plus étudiés sont les pathologies liées à la grossesse et la sclérose en plaques. En se basant sur les affiliations du premier et des six derniers auteurs, les équipes de recherche qui ont publiées le plus d'études sont principalement localisées en région parisienne, Bordeaux, Toulouse et Marseille. Les possibles interactions entre les acteurs de l'écosystème ont été illustrées dans un graphique interactif. **Discussion/Conclusion** Cette cartographie interactive a permis d'identifier les études réalisées à partir des données du SNDS et a mis en évidence une

hétérogénéité des domaines dans lesquels elles sont utilisées. Elle sera, ainsi que l'équation de recherche bibliographique, mise en open source sur le site internet du Health Data Hub avec une mise à jour annuelle.

Henrard, J.-C. (2022). "Les données de santé et leur utilisation." *Santé Publique* **34**(3): 333-334.  
<https://www.cairn.info/revue-sante-publique-2022-3-page-333.htm>

Ce dossier illustre différents exemples d'utilisation de données de santé au service des politiques de santé, à l'interface de la recherche et de données opérationnelles et les problématiques juridiques soulevées par la transformation numérique appliquée aux données de santé.

Tonon, C. (2022). "La gouvernance des données de santé en France et en Europe: dynamiques et obstacles." *Annales des mines*(18): 52-58.  
<https://www.annales.org/enjeux-numeriques/2022/resumes/juin/08-en-resum-FR-AN-juin-2022.html#08FR>

La gouvernance nationale et européenne des données de santé en construction repose sur un triple mouvement de centralisation et d'interopérabilité des bases, d'élargissement des accès à un nombre croissant d'acteurs et d'extension des finalités justifiant leur exploitation. Ce mouvement pose à la fois la question du déploiement et de la sécurisation de l'infrastructure technique sous-jacente, mais aussi des modalités juridiques d'accès et de traitement permettant de maximiser l'utilisation des données de santé tout en assurant le niveau de protection adéquat imposé par leur niveau particulier de sensibilité. Dans cet article, nous revenons sur les caractéristiques juridiques et politiques de la construction des modèles français et européen de gouvernance des données de santé, dont la plus récente formalisation devrait intervenir courant 2022 avec l'acte législatif sur l'espace européen des données de santé.

Galissaire, J. (2021). "Données de santé : Passer le cap citoyen !", Paris : Renaissance numérique  
<https://www.renaissancenumerique.org/publications/donnees-de-sante-passer-le-cap-citoyen>

Renaissance Numérique appelle à passer un cap dans l'exploitation des données au service du système de santé et à mettre les citoyens au cœur de cette transformation. Le think tank formule onze recommandations autour de trois grands axes : gouvernance du système de santé, partage des données de santé et redéfinition de la gouvernance des données de santé. Elles sont l'aboutissement d'une réflexion lancée au printemps 2020, qui a réuni une dizaine de membres du think tank et a donné lieu à des échanges avec trente-neuf personnalités clés de l'écosystème.

Health data Hub (2021). "Guide pédagogique Health Data Hub : Note d'information aux personnes". Paris HDH.  
<https://health-data-hub.fr/actualites/benchmark-international>

L'objectif de ce document est d'aider les porteurs de projet à fournir l'information aux personnes concernées par un projet de recherche impliquant un traitement de données de santé. Ce guide contient notamment un arbre de décision qui permettra d'identifier la situation juridique dans laquelle le chercheur se trouve et ses obligations, les éléments nécessaires pour réaliser l'information des personnes et présente quelques cas particuliers si le chercheur n'est pas en mesure d'informer.

Health data Hub (2021). "Health data Hub. Rapport annuel 2020 et feuille de route 2021". Paris Health data hub.

(2020). "Health data Hub. Plan stratégique 2019-2022". Paris Health data hub.

<https://www.health-data-hub.fr/sites/default/files/2020-11/Feuille%20de%20route%20strat%C3%A9gique%20HDH-compressed.pdf>

Ce document présente les grandes orientations du Health Data Hub pour la période 2019-2022.

Goldberg, M. et Zins, M. (2021). "Le Health Data Hub (fin)." *Med Sci (Paris)* **37**(3): 277-281.

<https://doi.org/10.1051/medsci/2021017>

Goldberg, M. et Zins, M. (2021). "Le Health Data Hub (suite)." *Med Sci (Paris)* **37**(3): 271-276.

<https://doi.org/10.1051/medsci/2021016>

Zins, M., Cuggia, M. et Goldberg, M. (2021). "Les données de santé en France." *Med Sci (Paris)* **37**(2): 179-184.

<https://doi.org/10.1051/medsci/2021001>

Bernelin, M., et al. (2020). "Données massives, big data et santé publique." *Actualite Et Dossier En Sante Publique*(112): 13-58.

<https://www.hcsp.fr/explore.cgi/adsp?clef=1173>

La quantité de données de santé concernant les personnes (poids, remboursements de soins, habitudes de vie, etc.) s'est accrue avec leur recueil systématique et à grande échelle. Parallèlement, il est devenu possible de traiter ces données massives et de livrer rapidement des informations autrefois difficiles d'accès, d'opérer des rapprochements inattendus et d'offrir des modèles prédictifs inédits. Cela sera-t-il révolutionnaire ou non pour la santé publique ?

Elbaum, M., Aquille, M., Minault, B., et al. (2020). "Les cohortes pour les études et la recherche en santé". Paris Igas.

<http://www.igas.gouv.fr/spip.php?article788>

A la demande des ministres en charge de la santé et de la recherche, l'Igas et l'IGéSR ont procédé à une évaluation de l'ensemble des 14 cohortes épidémiologiques soutenues dans le cadre du « programme des investissements d'avenir » (PIA) lancé en 2010. Les questionnements sur leur place dans la recherche s'inscrivent dans un environnement évolutif qui voit se développer d'autres instruments d'observation sur la base de données individuelles 'longitudinales' : bases de données médico-administratives, registres de morbidité ou de patients, enquêtes et panels statistiques. Les inspections générales ont donc adopté une vision large des enjeux d'observation et d'articulation des sources existant dans ce domaine. Elles proposent des modalités renouvelées de consolidation, d'accompagnement et de pilotage des cohortes existantes, et esquissent une stratégie nationale d'utilisation des données longitudinales pour la recherche en santé.

Longuet, G. et Villani, C. (2019). "Rapport sur l'intelligence artificielle et les données de santé". Paris Sénat, Paris Assemblée Nationale.

<http://www.senat.fr/notice-rapport/2018/r18-401-notice.html>

L'audition publique organisée par l'Office le 21 février 2019, sous forme de deux tables rondes, sur l'intelligence artificielle et les données de santé a montré combien la mise en œuvre de la stratégie nationale pour l'intelligence artificielle dans le domaine de la santé passe par une organisation adaptée de la collecte des données et par la régulation de leurs modalités d'accès et d'utilisation. Production, partage et protection des données comptent parmi les enjeux essentiels de leur bonne utilisation à des fins diagnostiques et thérapeutiques, dans une relation de confiance avec les individus, sans laquelle rien ne sera possible. L'Office présente des propositions pour une organisation efficace et contrôlée de la collecte et de l'utilisation des données de santé, dans la perspective de la discussion du projet de loi relatif à l'organisation et à la transformation du système de santé en cours d'examen par le Parlement.

Cugia, P., Polton, D. et Wainrib, G. (2018). "Health data Hub : mission de préfiguration". Paris Ministère chargé de la santé.

<https://solidarites-sante.gouv.fr/ministere/documentation-et-publications-officielles/rapports/sante/article/rapport-health-data-hub-mission-de-prefiguration>

Ce rapport rassemble les conclusions de la mission de préfiguration du « Health Data Hub », pilotée par trois experts, Dominique Polton, présidente de l'Institut national des données en santé (INDS), Marc Cuggia, professeur d'informatique médicale et praticien hospitalier au CHU de Rennes et Gilles Wainrib, président fondateur de la start-up Owkin. Il fait suite aux engagements du président de la République qui avait annoncé au moment de la remise du rapport Villani, que la santé serait un des secteurs prioritaires pour le développement de l'intelligence artificielle, avec deux actions majeures, la création d'un « Health Data Hub » et l'élargissement du système national de données de santé. Le rapport propose une feuille de route pour la mise en œuvre opérationnelle de cette plateforme d'exploitation des données de santé, ainsi que des recommandations, notamment sur les aspects organisationnels et réglementaires. La ministre de la Santé a confirmé la mise en place de ce Hub des données de santé dans les tous premiers mois de 2019 et a confié au directeur de la Drees, Jean-Marc Aubert, cette mission. Concrètement, le « Health Data Hub », prendra la forme d'un guichet unique, sécurisé dans l'objectif de permettre le partage des données de santé dans le respect du droit des patients et en assurant la transparence avec la société civile. Il mutualisera des ressources technologiques et humaines et constituera un outil de promotion de l'innovation pour faire de la France un leader des données de santé.

Pon, D. et Coury, A. (2018). "Stratégie de transformation du système de santé : Accélérer le virage numérique. Rapport final". Paris Ministère chargé de la santé.

[https://solidarites-sante.gouv.fr/IMG/pdf/masante2022\\_rapport\\_virage\\_numerique.pdf](https://solidarites-sante.gouv.fr/IMG/pdf/masante2022_rapport_virage_numerique.pdf)

Engagés début mars 2018, les travaux sur le chantier « Numérique » de la Stratégie de transformation du Système de santé (STSS), sur le thème « Accélérer le virage numérique », avaient pour objectif d'aboutir pour fin juin 2018 à la production d'une feuille de route opérationnelle couvrant la période 2018-2022 et ayant pour but de renforcer le virage numérique dans le système de santé. Cette note présente, de façon synthétique, les principaux éléments de diagnostic ainsi que des propositions clés élaborées suite aux différentes consultations.

Morange, P. (2017). "Rapport d'information sur les données médicales personnelles inter-régimes détenues par l'assurance maladie, versées au SNIIRAM, puis au Système national des données de santé (SNDS)". Paris Assemblée nationale.

<http://www.assemblee-nationale.fr/14/pdf/rap-info/i4533.pdf>

Babusiaux, C. (2016). "L'ouverture des bases de données publiques : le point et les enjeux de santé." *I2D - Information, données & documents* 53(3): 48-50.

<https://www.cairn.info/revue-i2d-information-donnees-et-documents-2016-3-page-48.htm>

Dans l'univers de la santé, où se multiplient les données numériques, les bases de données anonymisées gérées par des entités publiques sont un enjeu majeur. Pour qu'elles soient pleinement utiles, leur accès doit être suffisamment ouvert. En 2016, la loi de modernisation du système de santé prévoit des évolutions positives mais risque de conduire à des retours en arrière et crée un système administratif trop complexe. La voie d'avenir est la mise en place d'un véritable système de régulation.

Burnel, P. (2016). "L'action publique pour favoriser le développement vertueux du numérique en santé." *I2D - Information, données & documents* 53(3): 51-52.

<https://www.cairn.info/revue-i2d-information-donnees-et-documents-2016-3-page-51.htm>

Dans la santé comme dans d'autres secteurs, le numérique est porteur d'un fort potentiel de création de nouveaux services et d'émergence d'opportunités, sources de gains d'efficacité. L'enjeu pour la puissance publique est d'en favoriser la diffusion et l'usage en mobilisant la palette des instruments d'intervention à sa disposition.

Babusiaux, C. p. (2015). "Rapport au Parlement 2015. Ouverture, qualité, partage : des avancées". Charenton-Le-Pont Institut des données de santé.

Ce rapport, que l'Institut des données de santé remet comme chaque année au Parlement après adoption par son Assemblée Générale conformément à la loi du 13 août 2004, traduit des progrès importants dans trois domaines d'actions : L'ouverture des données pour la recherche s'est accélérée. L'IDS en a approuvé 161 depuis 2009, dont 83 depuis début 2014. L'utilisation des données par près de 200 chercheurs devrait permettre à la recherche de progresser dans des domaines majeurs de l'épidémiologie (cancer, diabète, asthme, etc.), des effets des médicaments et des interactions médicamenteuses, et de l'évaluation médico-économique (dispositifs médicaux, inégalités territoriales etc.). Le service d'aide à la décision est passé à 39 tableaux de bord couvrant les champs ambulatoires et hospitaliers et permettant le suivi d'indicateurs essentiels. Ce service, aujourd'hui à la disposition des membres de l'IDS et d'autres organismes œuvrant dans le domaine de la santé, favorise la connaissance et le partage d'informations essentielles entre les grands acteurs de la santé et de la protection sociale. La qualité et la cohérence des bases de données publiques ont été à nouveau améliorées. L'IDS a réalisé 20 rapports depuis 2011 et a proposé 154 améliorations afin que ces bases répondent mieux aux besoins des acteurs et de la recherche. Dans le même temps, l'IDS s'est assuré du respect de l'anonymat des personnes, du secret médical, ainsi que de l'éthique et de la déontologie auxquels veille le Comité d'Experts présidé par Didier Sicard, dont l'audition annuelle devant l'Assemblée Générale est retracée dans le présent rapport.(d'après l'éditorial).

Cour des comptes (2016). "Les données personnelles de santé gérées par l'assurance maladie. Une utilisation à développer, une sécurité à renforcer". Paris Cour des Comptes.

<https://www.ccomptes.fr/fr/publications/les-donnees-personnelles-de-sante-gerees-par-lassurance-maladie>

La Cour des comptes rend public, le 3 mai 2016, un rapport sur les données personnelles de santé gérées par l'assurance maladie, demandé par la commission des affaires sociales et la mission d'évaluation et de contrôle des lois de financement de la sécurité sociale de l'Assemblée nationale, en application de l'article LO. 132-3-1 du code des juridictions financières. Le système national d'information interrégimes de l'assurance maladie (SNIIRAM) constitue une base de données médico-administratives sans équivalent en Europe. Mise en service depuis 2004, elle a été progressivement enrichie et structurée, mais son potentiel, très important, demeure à parfaire. La qualité de la base peut être améliorée en réduisant les fragilités de codage et les remontées d'informations incomplètes. Les efforts doivent être poursuivis afin de réduire les risques pouvant affecter la confidentialité et la sécurité des données. Par ailleurs, les utilisations du SNIIRAM restent très limitées par rapport aux enjeux sanitaires et financiers. La Caisse nationale d'assurance maladie devrait notamment davantage y avoir recours, en particulier pour lutter contre les abus et la fraude des professionnels de santé. Devant être intégré au nouveau système national des données de santé institué en janvier 2016, le Sniiram demeure pour longtemps le cœur du système des données de santé. Pour encourager leur utilisation à des fins d'intérêt général, sécuriser l'existant, améliorer la gouvernance et assurer la fluidité des accès selon une approche à la fois rigoureuse et ouverte est prioritaire. La Cour formule 13 recommandations.

(2014). "Rapport de la Commission Open data en santé". Paris Ministère chargé de la Santé.

<https://drees.solidarites-sante.gouv.fr/publications/rapports/rapport-de-la-commission-open-data-en-sante>

La Commission " 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. Son rapport, remis le 9 juillet 2014 à la ministre de la Santé, retrace les travaux et discussions de la Commission. Adopté consensuellement par l'ensemble des membres de la commission, qui partagent des attentes communes et fortes, il dresse tout d'abord un panorama de l'existant : 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. Dans une seconde partie, il évoque les enjeux pour l'avenir. Enfin, il identifie les actions à mener : données à ouvrir en open data, orientations en matière de données réidentifiantes, données relatives aux professionnels et aux établissements.

Babusiaux, C.(2014). "Rapport au Parlement 2014. Progrès, enjeux, solutions". Charenton-Le-Pont Institut des données de santé.

Le rapport 2014 de l'Institut des données de santé traduit des progrès dans l'ouverture des données de santé. Ils ont permis des avancées significatives, par exemple : La création d'une base pour mieux connaître le parcours de soins des personnes atteintes d'un cancer afin d'optimiser leur traitement, la création d'une base permettant de mesurer l'impact des habitudes alimentaires des personnes sur leur santé afin de développer la prévention, le développement de l'expertise en matière de sécurité sanitaire, notamment celle de l'Agence Nationale de Sécurité du Médicament et des produits de santé, qui a constitué une équipe pour développer l'analyse des données auxquelles elle a désormais accès, et qui a fait émerger des relais dans des organismes de recherches afin d'accroître ses possibilités

d'analyse au service de la sécurité des patients, la possibilité pour des acteurs de la société civile d'utiliser les données pour réaliser des études, en particulier celles du Collectif Inter associatif Sur la Santé portant sur l'évolution du reste à charge des patients en cas de consultation d'un professionnel de santé ou de séjour hospitalier, la possibilité pour certaines professions de santé d'avoir accès aux données leur permettant d'éclairer leur dialogue avec l'assurance maladie. Compte tenu de l'actualité, ce rapport traite également des enjeux liés au nécessaire élargissement de l'ouverture de ces données dans l'intérêt des patients, et analyse les solutions proposées par les différents acteurs. Il vise ainsi à éclairer les débats, comme c'est aussi le rôle de l'Institut (d'après l'éditorial).

Babusiaux, C. p. (2013). "Rapport au Parlement 2013. Des avancées et des propositions raisonnées". Charenton-Le-Pont Institut des données de santé.

Ce rapport 2013 de l'Institut des données de santé, traduit, comme l'indique son titre, des avancées raisonnées dans tous les domaines confiés par le législateur à l'Institut : La création de nouvelles bases pour couvrir de nouveaux besoins, comme la cohorte « Cancer » mise en place par l'INCa afin de mieux évaluer les plans cancer, la mise en commun de données pour accroître les possibilités d'analyse : celles de l'assurance maladie obligatoire avec celles en provenance des établissements d'hébergement pour personnes âgées dépendantes ; celles des assureurs maladie obligatoires et complémentaires, après un premier test réussi pour en vérifier la faisabilité, afin de mieux connaître la dépense qui reste finalement à la charge des personnes, les travaux sur la qualité et la cohérence des bases, par exemple l'analyse du contenu, du périmètre, des règles de gestion et de la profondeur de l'historique des données exhaustives individuelles anonymisées de l'assurance maladie obligatoire, indispensables à des fins de santé publique et de veille sanitaire, l'extraction de données ou l'accès en routine des organismes de recherche aux systèmes d'information publics. Par ailleurs, et étendant son action à de nouveaux domaines, l'Institut a réalisé un ensemble de tableaux de bord constituant un Service d'Aide à la Décision permettant de suivre les indicateurs souhaités par les acteurs de la santé et de la protection sociale, en répondant précisément à leurs besoins. Ce rapport formule aussi des propositions raisonnées visant à franchir des pas supplémentaires pour répondre de manière adaptée et proportionnée aux besoins de chacun des acteurs de la santé et de la protection sociale vers une plus grande ouverture des bases pour améliorer leur connaissance. (tiré de l'éditorial).

Bras, P. L. et Loth, A. (2013). "Rapport sur la gouvernance et l'utilisation des données de santé". Paris Ministère chargé de la Santé.

<https://www.vie-publique.fr/rapport/33536-la-gouvernance-et-lutilisation-des-donnees-de-sante>

Par lettre du 16 avril 2013, la ministre des Affaires sociales et de la santé a saisi l'auteur de ce rapport, Pierre-Louis Bras d'une mission sur la gouvernance et l'utilisation des données de santé. Le sujet est complexe. Il concerne surtout la plus importante des bases de données publiques de santé dans notre pays, voire dans le monde, le Système national d'information inter-régime de l'assurance maladie (SNIIRAM). Issues des feuilles de soins et des résumés de sortie hospitaliers, dont on a retiré tous les éléments directement identifiants, les données du SNIIRAM décrivent l'offre et la consommation de soins ainsi que l'état de santé des quelques 65 millions d'habitants de notre pays dans la période récente. Le présent rapport propose les voies et les moyens pour mettre en place un dispositif d'accès et d'utilisation des bases de données médico-administratives, adapté aux besoins de santé publique et de sécurité sanitaire, dans des conditions fiables et sécurisées, respectant notamment le strict anonymat des patients. Le rapport s'inscrit dans le cadre plus général de la réflexion menée sur l'ouverture de l'accès aux données de santé. Il complète ainsi le rapport sur la pharmaco-

surveillance remis le 15 septembre 2013 à la ministre par les Professeurs Bégaud et Costagliola - qui recommandait la création d'une structure regroupant et analysant l'ensemble des données de santé sur l'utilisation des médicaments et produits de santé. Il s'articule par ailleurs avec la mission « Open data » sur la politique d'ouverture des données publiques.

(2012). "Pour une meilleure utilisation des bases de données administratives et médico-administratives nationales pour la santé publique et la recherche". Paris HCSP.  
[http://www.hcsp.fr/docspdf/avisrapports/hcspr20120309\\_bddadministration.pdf](http://www.hcsp.fr/docspdf/avisrapports/hcspr20120309_bddadministration.pdf)

La France dispose de bases de données médico-sociales et économiques nationales centralisées, constituées et gérées par des organismes publics, couvrant de façon exhaustive et permanente l'ensemble de la population dans divers domaines stratégiques pour la santé publique et la recherche : recours aux soins, hospitalisation, handicaps, prestations et situation professionnelle, sociale et économique. De plus, un identifiant individuel unique (le NIR : Numéro d'identification au répertoire) est actuellement utilisé par pratiquement toutes les bases de données nationales. Malgré certaines limites en termes de couverture, de qualité et de validité des données, ces bases de données, concernant plus de 60 millions de personnes, constituent un patrimoine considérable, vraisemblablement sans équivalent au monde. Cependant, l'utilisation à des fins de recherche et de surveillance de ces bases de données nationales se heurte actuellement à des obstacles divers, dont les plus importants sont de nature juridique et opérationnelle. En réponse à une saisine de la Direction générale de la santé (DGS), ce document fait le point sur les principales bases existantes, expose les difficultés pour leur utilisation et présente de façon détaillée une série de propositions pour les surmonter.

Babusiaux, C.(2012). "Rapport au Parlement 2012. Progrès et projets". Paris Institut des données de santé.

Le rapport 2012 de l'IDS (Institut des données de santé) traduit de nouveaux progrès dans tous les domaines confiés par le législateur à l'Institut : un accroissement du nombre d'acteurs de la santé et de la protection sociale ayant accès aux données anonymisées des systèmes d'information ambulatoire et hospitalier pour améliorer leur connaissance ; la création de nouvelles bases pour couvrir de nouveaux besoins, comme le système national des données mutualistes, maintenant opérationnel et comportant cinq millions de personnes ; la mise en commun de données pour accroître les possibilités d'analyse, comme le test réussi mené avec les assureurs maladie obligatoires et complémentaires afin de mieux caractériser le reste à charge final des assurés ; la finalisation de 29 des 30 tableaux de bord souhaités par les partenaires de l'Institut, constituant le service d'aide à la décision ; la poursuite des travaux sur la qualité et la cohérence des bases, par exemple l'accroissement de la profondeur de l'historique des données ambulatoires et hospitalières, fondamental pour la santé publique ; l'accès désormais en routine des organismes de recherche aux systèmes d'information publics ; des travaux de fond pour réunir les conditions de mise en commun de données tout en assurant l'anonymat des personnes. Ces progrès ont conduit ses membres à souhaiter la reconduction de l'institut des données de santé (tiré de l'éditorial).

Babusiaux, C. (2011). "Rapport 2011 : progrès et consolidation". Charenton Institut des données de santé.

Ce rapport présente l'activité 2011 de l'Institut des données de santé (IDS) et porte le sous-titre "Progrès et consolidation". L'instance installée en 2007 a pour mission d'assurer la cohérence et de veiller à la qualité des systèmes d'information utilisés pour la gestion du risque maladie, et de faciliter la mise à disposition des données issues des systèmes d'information de ses membres (État, assurance maladie obligatoire et complémentaire, CISS, établissements publics et privé, CNSA et l'UNPS). Christian Babusiaux, son président, note que l'année 2011 a été marquée par des étapes importantes dans plusieurs domaines. Pour améliorer la connaissance que peuvent avoir les acteurs de la santé et de la protection sociale, un service d'aide à la décision a été ouvert. Il compte une quinzaine de tableaux de bord permettant de suivre des problèmes clefs comme le reste à charge pour les patients ou les transferts financiers de l'hôpital vers la ville. La nouvelle base de données, le Système National d'Informations Hospitalières (SNATIH), a été largement ouverte depuis février 2011 à tous les acteurs, donnant accès aux données essentielles sur le financement et l'activité de l'ensemble des établissements de soins. Un nombre croissant d'organismes de recherches accèdent au SNIIRAM et peuvent désormais obtenir des extractions de données de cette base. Autre fait marquant : la signature d'une charte d'engagements entre les assureurs maladie obligatoires et complémentaires pour le partage et la mise en commun de données issues de leur système d'information. La loi sur la sécurité sanitaire des produits de santé du 29 décembre 2011 prévoit, dans son article 22, la création d'un GIP Etudes de santé. Il permettra à l'ANSM (qui remplace l'Afssaps), à la HAS et à l'InVS d'avoir accès aux données du système national d'information interrégimes de l'assurance maladie, notamment en vue de réaliser des études de vigilance et d'épidémiologie impliquant des produits de santé.

Babusiaux, C.(2010). "Rapport 2010 : enjeux et étapes". Charenton Institut des données de santé.

La mission donnée par la loi à l'institut des données de santé est d'inciter au développement de bases de données, à leur cohérence et d'en favoriser l'accès et le partage. Cet objectif n'est pas seulement technique. Il est important pour le développement de la connaissance en matière de santé et de protection sociale, l'efficacité du système de soins, la prévention des risques sanitaires. Ces derniers mois ont permis des progrès. Au plan stratégique, l'orientation a été prise par la Caisse nationale de solidarité pour l'autonomie, avec les pouvoirs publics, de constituer une base de données sur les aides aux personnes âgées dépendantes, qui permettra à terme des rapprochements avec leur prise en charge par l'assurance maladie ; une charte d'engagements a été adoptée entre assureurs maladie obligatoires et complémentaires pour un échange partenarial des données ; dans cette perspective, plusieurs organismes et fédérations d'assurance santé ont progressé vers la mise en place de bases de données. Au plan technique, des tableaux de bord répondant aux besoins exprimés par les différents partenaires ont commencé à être développés, la CNAMTS étant engagée à les mettre en œuvre ; l'IDS a désormais accès au système national d'information sur hospitalisation ; les travaux entre l'Irdes et l'Ids sur un premier échantillon chaînant des données des assureurs maladie complémentaires et obligatoires se déroulent comme prévu ; la CNIL a donné son accord de principe à un allongement de la durée de conservation des données du SNIIRAM1. Pour la recherche, l'accès des organismes publics aux données du SNIIRAM est entré en phase de routine ; un « livre blanc » sur les évolutions nécessaires pour faciliter encore l'accès et l'utilisation des bases a été publié par le Comité d'experts de l'institut.

(2009). "Les systèmes d'information pour la santé publique". Paris HCSP.

[http://www.hcsp.fr/docspdf/avisrapports/hcspr20091111\\_sisp.pdf](http://www.hcsp.fr/docspdf/avisrapports/hcspr20091111_sisp.pdf)

Dès son installation en mars 2007, le Haut Conseil de la Santé Publique, a fait le point sur le suivi des objectifs de la loi de santé publique (LSP) et a constaté que certains indicateurs restaient manquants. Un passage en revue systématique des objectifs n'ayant pu donner lieu à un suivi en raison de l'insuffisance des données disponibles a été réalisé. À la suite de ce rapport, une réflexion globale sur les besoins, l'organisation et l'articulation des systèmes d'information nationaux et régionaux dans le cadre de l'évaluation des objectifs de la loi de santé publique 2004 et des plans de santé publique a eu lieu. Le HCSP a établi que les sources de données étaient nombreuses mais difficilement accessibles et insuffisamment coordonnées et il a formulé une série de recommandations spécifiques et de court terme et ainsi que des propositions plus générales pour améliorer la mesure des indicateurs et l'estimation de l'atteinte des objectifs de la future loi de santé publique. À court terme, le HCSP préconise particulièrement que soit engagée une réflexion juridique sur l'utilisation du numéro d'identification NIR de sorte à faciliter l'appariement de diverses sources ainsi que d'initier ou développer certains systèmes d'information relevant d'organismes publics dans des domaines cruciaux et actuellement non couverts, notamment en médecine ambulatoire. Parmi ses recommandations générales le HCSP préconise de mieux coordonner les différentes actions de pilotage et de suivi des systèmes d'information pour la santé publique, d'inciter les organismes producteurs et utilisateurs de données à dégager les ressources nécessaires à une meilleure utilisation des données, de développer des dispositifs d'observation longitudinale en s'appuyant sur des cohortes existantes ou à initier et de mettre en place un dispositif de surveillance systématique et permanente.

(2009). "Rapport au parlement 2008". Charenton le Pont IDS.

L'institut des données de santé a été mis en place au printemps 2007. Le rapport retrace les premières réalisations : ouverture aux membres de l'IDS et à leurs propres membres de l'accès au système d'information des régimes obligatoires ; travaux avec la FNMF et la CNSA pour développer des systèmes analogues chez les mutuelles et sur la dépendance ; amélioration des statistiques mises à la disposition de tous les acteurs ; collaborations engagées avec l'Irdes et la DREES pour mieux connaître la protection complémentaire, avec l'ATIH et la DGFIP pour développer la connaissance de l'activité hospitalière. Dans le même temps, des garanties ont été mises en place : un comité d'experts, veillant à la déontologie et à la qualité des méthodologies, une charte de déontologie.

(2009). "Rapport au Parlement 2009 - Les progrès". Charenton le Pont IDS.

Depuis la création de l'institut des données de santé au printemps 2007 et ses premières réalisations en 2007-2008, de nouveaux progrès ont été accomplis. Des progrès prometteurs, enregistrés malgré la lourdeur et la complexité des procédures à suivre et la sensibilité des sujets. Dans la même logique que celle qui avait conduit les régimes obligatoires d'assurance maladie à constituer une base de données (SNIIRAM), les assureurs maladie complémentaires sont désormais engagés dans une voie analogue et à permettre l'utilisation de leurs bases au service d'une meilleure connaissance du système de santé. Des organismes de recherche, de plus en plus nombreux, ont pu, à la suite d'avis positifs émis par l'Irdes, accéder aux données du SNIIRAM pour travailler sur des thèmes de santé publique. Les garanties mises en place dès l'origine de l'institut, charte de déontologie et Comité d'experts chargé de veiller à son respect, ont pleinement fonctionné et permis ces progrès en veillant au respect de la confidentialité des données et des libertés individuelles.

Jolly, D., Lancry, P. J. et Thelot, B. (1997). "La médecine à l'épreuve de la société d'information". Paris Médecine Sciences Flammarion.

Cet ouvrage établit un constat de toute première importance sur les nécessités et les contraintes de l'information médicale, et son cadre juridique actuel : comment protéger la vie privée et la confidentialité des données médicales ? quelles sont les règles de transmission des informations médicales ? quelles sont les marges de manœuvre, encore offertes à la négociation ? comment résoudre les questions en suspens, comme les informations souhaitées par les différents acteurs, le choix des nomenclatures et l'utilisation par le patient de son droit d'opposition. Dans sa partie finale, ce livre formule un certain nombre de recommandations aux décideurs, afin que l'ensemble du cadre juridique garantissant les règles d'exploitation des données médicales puisse évoluer, en accord avec les principaux acteurs, et contribuer à une meilleure organisation du système de santé français. (Résumé d'auteur).

Combes, S. et al., e. (2019/07-08). "Le Health Data Hub, un an après." Revue Hospitalière De France(589).

À la suite du rapport Villani, présenté le 29 mars 2018 au Collège de France, la création d'un Health Data Hub fut annoncée par le président de la République, Emmanuel Macron. Cette plateforme vise à enrichir et à valoriser le Système national de données de santé (SNDS) en incluant l'ensemble de celles dont la collecte est soutenue par la solidarité nationale. Créé par la loi de modernisation du système de santé, le SNDS est actuellement composé des données de l'assurance maladie, des données de facturation hospitalière, des causes médicales de décès, des données médico-sociales des personnes handicapées et d'un échantillon de données de remboursement des organismes complémentaires. Dans un cadre qui garantit le respect des droits fondamentaux de chacun, cette structure contribuera à lever les freins à l'usage des données de santé. La dynamique engagée renforce nos capacités à innover et contribuera à faire de la France un leader dans l'intelligence artificielle en santé.

Devillier, N. (2017). "Chapitre 6. Les dispositions de la loi de modernisation de notre système de santé relatives aux données de santé." Journal International de Bioéthique **28**(3): 57-61.  
<https://www.cairn.info/revue-journal-international-de-bioethique-2017-3-page-57.htm>

La loi sur la modernisation de notre système de santé adoptée en 2016 étend le champ du secret médical en créant le dossier médical partagé (DMP) du patient. Tous les professionnels de santé intervenant dans le parcours de soin ont accès aux informations qui y sont renseignées. Le patient y accède par un site Internet dédié. Les décrets d'application de ce dispositif fixent les modalités de ce partage et encadrent la création de ce traitement automatisé de données à caractère personnel. Enfin, la loi crée un accès ouvert et sécurisé aux données de santé dans l'intérêt de la collectivité. Cet open data en santé est placé sous la gouvernance de l'Institut national des données de santé.

Rial-Sebbag, E. (2017). "Chapitre 4. La gouvernance des Big data utilisées en santé, un enjeu national et international." Journal International de Bioéthique **28**(3): 39-50.  
<https://www.cairn.info/revue-journal-international-de-bioethique-2017-3-page-39.htm>

L'utilisation des données de santé est de plus en plus considérée comme un enjeu central pour la recherche mais également pour le soin. La génération de ces données est une valeur ajoutée pour la conduite d'études à grande échelle, elle est même considérée comme une (r)évolution dans la méthodologie de la recherche ou encore la médecine personnalisée. Plusieurs facteurs ont influencé l'accélération de l'utilisation des données de santé (progrès de la génétique, de la technologie, diversification des sources) conduisant à questionner les principes juridiques posés pour la protection des données de santé tant en droit français

qu'en droit européen. En effet, premièrement, la production de masse (Big Data) de données dans le champ de la santé influe sur la quantité et la qualité des données venant dès lors reconfigurer les outils de protection de la vie privée en insistant sur le risque informationnel. Deuxièmement, l'utilisation de ces données repose quant à elle sur des principes fondamentaux existants tout en soulevant de nouveaux challenges pour leur gouvernance.

Caillol, H. (2015). "Ouverture des données de santé : l'expérience de l'Assurance maladie." Informations Sociales(191): 60-67.

L'Assurance maladie met à disposition des données ouvertes sur les professionnels de santé et les assurés, qui présentent un fort potentiel. L'objectif est d'élargir le périmètre des données centralisées et celui des destinataires tout en garantissant la protection des données individuelles. Réalisée progressivement, l'ouverture s'est appuyée sur un accompagnement des utilisateurs qui a évolué vers la coopération et le partenariat. (R.A.).

Dimeglio, C., Delpierre, C., Savy, N., et al. (2015). "Big data et santé publique : plus que jamais, les enjeux de la connaissance." Actualité Et Dossier En Santé Publique(93): 5-7.

De nombreuses bases de données existent dans le domaine de la santé. Quelles données intégrer ? Comment s'assurer de leur fiabilité ? Pour quelle utilité avec quels risques ? (introd.).

Gissot, C. et Polton, D. (2014). "Les bases de données de l'assurance-maladie : un potentiel pour l'amélioration du système de santé et pour la recherche." Statistique Et Société 2(2): 19-25.  
[http://publications-sfds.fr/index.php/stat\\_soc/article/view/317/297](http://publications-sfds.fr/index.php/stat_soc/article/view/317/297)

Les bases de données de l'assurance-maladie sont rassemblées dans un système national qui ouvre des possibilités remarquables en santé publique. Ce système a en effet des utilisations multiples : il permet d'examiner la dépense de soins selon des critères médicaux, d'analyser des parcours de soins, de connaître l'efficacité de traitements en vie réelle, et peut contribuer à surveiller la sécurité des médicaments. Ce système est de plus en plus utilisé, notamment par les chercheurs, et la Caisse nationale d'assurance-maladie s'est organisée pour faciliter ces utilisations. Un débat a été lancé récemment sur l'intérêt et les enjeux d'une ouverture plus large des données, et les modalités d'accès pourraient être amenées à évoluer (tiré de l'intro).

Thabalard, J. C. et Fieschi, M. (2014). "Données de santé : données sensibles." Statistique Et Société 2(2)  
[http://publications-sfds.fr/index.php/stat\\_soc/article/view/317/297](http://publications-sfds.fr/index.php/stat_soc/article/view/317/297)

Ce numéro s'intéresse à l'accès aux données de santé. Ces dernières, parmi lesquelles beaucoup émanent d'un système de sécurité sociale ancien et très organisé, se démultiplient de façon impressionnante – elles semblent assurément « big » aujourd'hui - et qu'en outre elles deviennent de plus en plus aisément accessibles. Mais quelle est la nature exacte de ces données ? Qui y a accès et qui ne devrait pas avoir accès ? Pour quoi faire ? Telles sont les questions auxquels nos auteurs se sont attachés à fournir des éléments de réponse.

Babusiaux, C., Sicard, D. et al.. (2010). "Éditorial - L'institut des données de santé." Rev Epidemiol Sante Publique 58(2): 85-88.

[Cet éditio présente l'Institut des Données de Santé (I.D.S.), sa composition, ses missions et son apport aux organismes réalisant des études et recherches en proposant un accès aux données du système d'information inter-régimes de l'assurance-maladie (Sniiram).

Tuppin, P., De Roquefeuille, L., Weill, A., et al. (2010). "Occupational note - French national health insurance information system and the permanent beneficiaries sample." Rev Epidemiol Sante Publique **58**(4): 286-290.

En France, au début des années 2000, le législateur a souhaité que les régimes d'assurance-maladie développent un système d'information inter-régimes (SNIIR-AM) afin de mieux connaître et évaluer le recours aux soins et les dépenses de santé des assurés. En 2009, il contenait des données du régime général d'assurance-maladie qui prend en charge 86% de la population française, soit environ 53 millions de personnes. Ces données ne sont disponibles que sur une période de deux ans au-delà de l'année en cours. En outre, un échantillon généraliste de bénéficiaires (EGB) a été créé à partir du SNIIR-AM. Il est permanent, représentatif de la population protégée et les bénéficiaires et leur consommation de soins seront suivis pendant 20 ans depuis 2004. Il contient des informations anonymes sur les caractéristiques sociodémographiques et médicales des bénéficiaires et les prestations qu'ils ont perçues. Il résulte d'un sondage au 1/97e sur une clef du numéro d'identification des bénéficiaires, qu'ils aient ou non perçu des remboursements de soins dans l'année. Il regroupait en 2009 près de 500 000 bénéficiaires du régime des travailleurs salariés hors fonctionnaires et étudiants (77% de la population résidant en France). L'EGB permet la réalisation d'études longitudinales en reconstituant les parcours et recours aux soins des patients, en ville et à l'hôpital, de calculer les dépenses individuelles, d'étudier certaines pathologies relativement fréquentes caractérisées par un remboursement à 100% pour certaines maladies chroniques et le remboursement de médicaments traceurs. Le SNIIR-AM doit, à terme, couvrir les assurés de l'ensemble des régimes d'assurance-maladie de France. (R.A.).

Polton, D. (2009). "Les bases de données de l'assurance maladie, un potentiel pour la recherche en économie de la santé : éditorial." Lettre Du Collège (La) **20**(1)  
[http://www.ces-asso.org/docs/Let\\_CES\\_1-2009.pdf](http://www.ces-asso.org/docs/Let_CES_1-2009.pdf)

Pene, P., Gremy, F. et Dubois, G. (2008). "Les données individuelles de santé." Bulletin de l'Académie Nationale de Médecine **192**(2): 437-447.

Les données individuelles de santé constituent une base essentielle de l'avenir de l'épidémiologie et permettent le suivi des soins donnés aux patients. Leur utilisation est indispensable aux épidémiologistes pour une meilleure connaissance de l'état de santé. Sans ces informations il est difficile de définir une politique de santé. La commission s'est intéressée aux modalités suivantes : le Système National d'Information Inter Régime de l'Assurance Maladie (SNIR-AM) 16 régimes mais ne permet pas le suivi des patients. Le projet Constance propose d'utiliser les centres d'examen de santé pour la recherche épidémiologique. Les Observatoire régionaux de la santé ont pour mission une aide à la décision. Ils sont d'excellents bureaux d'études multidisciplinaires. L'Institut national des données de santé (INDS) est un groupement d'intérêt public qui va permettre de regrouper les données, de définir leurs modalités d'accès et de partage. L'Ordre national des pharmaciens a développé avec succès un dossier pharmaceutique qui comporte l'ensemble des médicaments délivrés au cours des quatre derniers mois. Le Dossier Médical Personnel (DMP) a pour but le recueil par les médecins de données individuelles de santé des patients.

Sa généralisation prévue en 2007 ne pourra intervenir avant plusieurs années car la priorité a été donnée à la dimension informatique alors que son but premier est la qualité des soins. L'Académie nationale de médecine propose les recommandations suivantes : soutenir la mise en place de grandes cohortes en population générale ; former des épidémiologistes, médecins et non médecins ; renforcer l'action des ORS tout en conservant leur indépendance ; former les étudiants en médecine à l'utilisation des données individuelles de santé ; considérer que le partage des données individuelles de santé peut améliorer la qualité des soins et diminuer la redondance d'examen ; le masquage partiel des données par les malades pouvant nuire aux soins, les patients auront à prendre la responsabilité de leurs décisions ; encourager l'expérience du DMPE car le suivi de l'enfant est essentiel et mérite d'être pris en compte jusqu'à seize ans ; utiliser l'expérience acquise par les pharmaciens pour le Dossier Pharmaceutique et l'intégrer au DMP ; orienter le DMP vers un projet plus sobre. L'Académie nationale de médecine s'inquiète cependant du retard pris par le DMP qui est un projet important pour tous, mais également pour la politique de santé et la recherche médicale de notre pays. Une période d'expérimentation permettant une évaluation devrait constituer un préalable essentiel.

Chevreur, K., Le Fur, P. et Sermet, C. (2004). "Systèmes d'information sur les pratiques médicales de ville : enseignements à partir de quelques expériences étrangères." Questions D'économie De La Santé (Irdes)(85)

<https://www.irdes.fr/Publications/Qes/Qes85.pdf>

Certains pays ont déjà mis en place des systèmes d'information publics décrivant les situations cliniques auxquelles font face les médecins et les attitudes diagnostique et thérapeutique qu'ils adoptent en réponse. Leur analyse permet de connaître les conditions de leur mise en œuvre et de souligner leurs nombreux points communs. Ces similitudes suggèrent qu'elles sont sans doute nécessaires au bon fonctionnement des systèmes en place et doivent servir de base à la réflexion française. Huit systèmes informatisés ont été identifiés, situés dans cinq pays : le Royaume Uni, l'Australie, la Nouvelle-Zélande, les Pays-Bas et l'Italie.

(2002). "Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé." Actualité Et Dossier En Santé Publique(39)

Chaque demande de traitement automatisé de données nominatives ayant pour fin la recherche dans le domaine de la santé doit être soumise à l'avis du Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé (CCTIR), préalablement à la saisine de la CNIL. Présentation de ce comité.

## A L'ETRANGER

Gruny, P. et Haribbey, L. (2023). "Rapport d'information sur la proposition de règlement du Parlement européen et du Conseil relatif à l'espace européen des données de santé - COM(2022) 197 final". Sénat, P.

<https://www.senat.fr/rap/r22-848/r22-8481.pdf>

La Commission européenne a présenté le 3 mai 2022 la proposition de règlement relatif à l'espace européen des données de santé – COM(2022) 197 final. Celle-ci a pour objectif de déterminer dans quelles conditions les données de santé électroniques pourront être traitées, d'une part, à des fins d'utilisation primaire, c'est-à-dire dans le but de fournir des services de santé à la personne physique concernée, et d'autre part, à des fins d'utilisation secondaire dont les finalités, parmi lesquelles la recherche, sont précisées à l'article 34 de la

proposition de règlement. La commission des affaires européennes du Sénat a examiné cette proposition de règlement, sur le rapport de Mmes Pascale Gruny et Laurence Harribey, et a conclu au dépôt de la proposition de résolution n° 849 (2022-2023) du 5 juillet 2023. Après avoir mis en évidence l'intérêt d'un espace européen des données de santé, Mmes Pascale Gruny et Laurence Harribey ont formulé des propositions pour, d'une part, garantir la cohérence de la proposition de règlement avec le règlement général sur la protection des données (RGPD) et d'autre part, assurer la circulation et la sécurité des

March, S., et al. (2023). "[Research Data Center on Health - Vision for Further Development from the Research Perspective]." Gesundheitswesen (Bundesverband der Ärzte des Öffentlichen Gesundheitsdienstes (Germany)) **85**(S 02): S145-S153.  
<https://doi.org/10.1055/a-1999-7436>

The German research data center for health will provide claims data of statutory health insurances. The data center was set up at the medical regulatory body BfArM pursuant to the German data transparency regulation (DaTraV). The data provided by the center will cover about 90% of the German population, supporting research on healthcare issues, including questions of care supply, demand and the (mis-)match of both. These data support the development of recommendations for evidence-based healthcare. The legal framework for the center (including §§ 303a-f of Book V of the Social Security Code and two subsequent ordinances) leaves a considerable degree of freedom when it comes to organisational and procedural aspects of the center's operation. The present paper addresses these degrees of freedom. From the point of view of researchers, ten statements show the potential of the data center and provide ideas for its further and sustainable development.

Sénat. (2023). "Note sur les données de santé. Allemagne – Israël – Suède". Paris Sénat  
<https://www.senat.fr/lc/lc324/lc324.html>

Cette note constitue une étude sur le cadre juridique de collecte et d'utilisation des données de santé, ainsi que sur leur gouvernance d'ensemble dans les trois pays suivants : Allemagne, Israël et Suède. Pour chaque pays considéré, les points suivants sont examinés : Principaux acteurs et gouvernance d'ensemble; recueil des données de santé : périmètre et modalités; utilisation primaire et secondaire des données de santé; cadre juridique de la protection des données de santé; accès des patients et des professionnels de santé aux données de santé (utilisation primaire); accès aux données de santé pour la recherche, l'innovation et l'élaboration des politiques publiques (utilisation secondaire); développement de l'intelligence artificielle en santé (utilisation secondaire); altruisme des données.

Andrew, D. M., et al. (2022). "Faciliter l'accès aux données administratives sur la vaccination anti-SRAS-CoV-2 aux fins de la recherche." Canadian Medical Association Journal **194**(33): E1147.  
<http://www.cmaj.ca/content/194/33/E1147.abstract>

Moorthie, S., et al. (2022). "Rapid systematic review to identify key barriers to access, linkage, and use of local authority administrative data for population health research, practice, and policy in the United Kingdom." BMC Public Health **22**(1): 1263.

**BACKGROUND:** Improving data access, sharing, and linkage across local authorities and other agencies can contribute to improvements in population health. Whilst progress is being made to achieve linkage and integration of health and social care data, issues still exist in creating such a system. As part of wider work to create the Cambridge Child Health Informatics and Linked Data (Cam-CHILD) database, we wanted to examine barriers to the access, linkage, and use of local authority data. **METHODS:** A systematic literature search was conducted of scientific databases and the grey literature. Any publications reporting original research related to barriers or enablers of data linkage of or with local authority data in the United Kingdom were included. Barriers relating to the following issues were extracted from each paper: funding, fragmentation, legal and ethical frameworks, cultural issues, geographical boundaries, technical capability, capacity, data quality, security, and patient and public trust. **RESULTS:** Twenty eight articles were identified for inclusion in this review. Issues relating to technical capacity and data quality were cited most often. This was followed by those relating to legal and ethical frameworks. Issue relating to public and patient trust were cited the least, however, there is considerable overlap between this topic and issues relating to legal and ethical frameworks. **CONCLUSIONS:** This rapid review is the first step to an in-depth exploration of the barriers to data access, linkage and use; a better understanding of which can aid in creating and implementing effective solutions. These barriers are not novel although they pose specific challenges in the context of local authority data.

(2021). "Les pratiques de gouvernance et d'accès aux données de santé dans le monde : restitution du benchmark 2020". Paris HDH.

<https://health-data-hub.fr/actualites/benchmark-international>

Le Health Data Hub a réalisé fin 2020 un benchmark international de 17 structures dont les stratégies et le fonctionnement sont comparables aux siennes dans 11 pays. Ce comparatif, élaboré à partir de recherches documentaires, approfondies par des entretiens avec les responsables des plateformes et bases de données des différents pays, permet d'obtenir un aperçu des pratiques internationales relatives à l'accès et à l'hébergement des données de santé, aux relations avec les responsables de données et les citoyens, aux modèles économiques, ou encore aux règles de valorisation scientifique.

Gordon, E. (2020). "Administrative Data Research UK." *Patterns (N Y)* **1**(1): 100010.

ADR UK is helping to transform the way researchers access the UK's wealth of administrative data, enabling government policy to be informed by the best evidence available. Emma shares her insights into the ADR UK approach to making this happen, explaining why building trust is central to the ADR UK mission.

Grath-Lone, L. M., et al. (2022). "What makes administrative data "research-ready"? A systematic review and thematic analysis of published literature." *Int J Popul Data Sci* **7**(1): 1718.

**INTRODUCTION:** Administrative data are a valuable research resource, but are under-utilised in the UK due to governance, technical and other barriers (e.g., the time and effort taken to gain secure data access). In recent years, there has been considerable government investment in making administrative data "research-ready", but there is no definition of what this term means. A common understanding of what constitutes research-ready administrative data is needed to establish clear principles and frameworks for their development and the realisation of their full research potential. **OBJECTIVE:** To define the characteristics of research-ready administrative data based on a systematic review and

synthesis of existing literature. METHODS: On 29(th) June 2021, we systematically searched seven electronic databases for (1) peer-reviewed literature (2) related to research-ready administrative data (3) written in the English language. Following supplementary searches and snowball screening, we conducted a thematic analysis of the identified relevant literature. RESULTS: Overall, we screened 2,375 records and identified 38 relevant studies published between 2012 and 2021. Most related to administrative data from the UK and US and particularly to health data. The term research-ready was used inconsistently in the literature and there was some conflation with the concept of data being ready for statistical analysis. From the thematic analysis, we identified five defining characteristics of research-ready administrative data: (a) accessible, (b) broad, (c) curated, (d) documented and (e) enhanced for research purposes. CONCLUSIONS: Our proposed characteristics of research-ready administrative data could act as a starting point to help data owners and researchers develop common principles and standards. In the more immediate term, the proposed characteristics are a useful framework for cataloguing existing research-ready administrative databases and relevant resources that can support their development.

Youens, D., et al. (2020). "Using general practice clinical information system data for research: the case in Australia." *Int J Popul Data Sci* 5(1): 1099.

General practice is often a patient's first point of contact with the health system and the gateway to specialist services. In Australia, different aspects of the health system are managed by the Commonwealth Government and individual state / territory governments. Although there is a long history of research using administrative data in Australia, this split in the management and funding of services has hindered whole-system research. Additionally, the administrative data typically available for research are often collected for reimbursement purposes and lack clinical information. General practices collect a range of patient information including diagnoses, medications prescribed, results of pathology tests ordered and so on. Practices are increasingly using clinical information systems and data extraction tools to make use of this information. This paper describes approaches used on several research projects to access clinical, as opposed to administrative, general practice data which to date has seen little use as a resource for research. This information was accessed in three ways. The first was by working directly with practices to access clinical and management data to support research. The second involved accessing general practice data through collaboration with Primary Health Networks, recently established in Australia to increase the efficiency and effectiveness of health services for patients. The third was via NPS MedicineWise's MedicineInsight program, which collects data from consenting practices across Australia and makes these data available to researchers. We describe each approach including data access requirements and the advantages and challenges of each method. All approaches provide the opportunity to better understand data previously unavailable for research in Australia. The challenge of linking general practice data to other sources, currently being explored for general practice data, is discussed. Finally, we describe some general practice data collections used for research internationally and how these compare to collections available in Australia.

(2019). "Health in the 21st Century. Putting Data to Work for Stronger Health Systems", Paris : OCDE [https://www.oecd-ilibrary.org/social-issues-migration-health/health-in-the-21st-century\\_e3b23f8e-en](https://www.oecd-ilibrary.org/social-issues-migration-health/health-in-the-21st-century_e3b23f8e-en)

This report explores how data and digital technology can help achieve policy objectives and drive positive transformation in the health sector while managing new risks such as privacy, equity and implementation costs. It examines the following topics: improving service delivery

models; empowering people to take an active role in their health and their care; improving public health; managing biomedical technologies; enabling better collaboration across borders; and improving health system governance and stewardship. It also examines how health workforces should be equipped to make the most of digital technology. The report contains findings from surveys of OECD countries and shares a range of examples that illustrate the potential benefits as well as challenges of the digital transformation in the health sector. Findings and recommendations are relevant for policymakers, health care providers, payers, industry as well as patients, citizens and civil society.

Mourby, M. J., et al. (2019). "Health Data Linkage for UK Public Interest Research: Key Obstacles and Solutions." *Int J Popul Data Sci* **4**(1): 1093.

**INTRODUCTION:** Analysis of linked health data can generate important, even life-saving, insights into population health. Yet obstacles both legal and organisational in nature can impede this work. **APPROACH:** We focus on three UK infrastructures set up to link and share data for research: the Administrative Data Research Network, NHS Digital, and the Secure Anonymised Information Linkage Databank. Bringing an interdisciplinary perspective, we identify key issues underpinning their challenges and successes in linking health data for research. **RESULTS:** We identify examples of uncertainty surrounding legal powers to share and link data, and around data protection obligations, as well as systemic delays and historic public backlash. These issues require updated official guidance on the relevant law, approaches to linkage which are planned for impact and ongoing utility, greater transparency between data providers and researchers, and engagement with the patient population which is both high-profile and carefully considered. **CONCLUSIONS:** Health data linkage for research presents varied challenges, to which there can be no single solution. Our recommendations would require action from a number of data providers and regulators to be meaningfully advanced. This illustrates the scale and complexity of the challenge of health data linkage, in the UK and beyond: a challenge which our case studies suggest no single organisation can combat alone. Planned programmes of linkage are critical because they allow time for organisations to address these challenges without adversely affecting the feasibility of individual research projects.

Rial-Sebbag, E. (2017). "Chapitre 4. La gouvernance des Big data utilisées en santé, un enjeu national et international." *Journal International de Bioéthique* **28**(3): 39-50.

<https://www.cairn.info/revue-journal-international-de-bioethique-2017-3-page-39.htm>

L'utilisation des données de santé est de plus en plus considérée comme un enjeu central pour la recherche mais également pour le soin. La génération de ces données est une valeur ajoutée pour la conduite d'études à grande échelle, elle est même considérée comme une (r)évolution dans la méthodologie de la recherche ou encore la médecine personnalisée. Plusieurs facteurs ont influencé l'accélération de l'utilisation des données de santé (progrès de la génétique, de la technologie, diversification des sources) conduisant à re-questionner les principes juridiques posés pour la protection des données de santé tant en droit français qu'en droit européen. En effet, premièrement, la production de masse (Big Data) de données dans le champ de la santé influe sur la quantité et la qualité des données venant dès lors reconfigurer les outils de protection de la vie privée en insistant sur le risque informationnel. Deuxièmement, l'utilisation de ces données repose quant à elle sur des principes fondamentaux existants tout en soulevant de nouveaux challenges pour leur gouvernance.

Guerrazzi, C. (2019). "An International Perspective on Health Information Exchange: Adoption in OECD Countries With Different Health Care System Configurations." Medical Care Research and Review 0(0): 1077558719858245.

<https://pubmed.ncbi.nlm.nih.gov/31216934/>

The sharing of information among various care providers is becoming an essential feature of health care systems, and many countries are now adopting policies to foster health information exchange, defined as the electronic transfer of data or information among health care organizations involved in the delivery of care. Given the increasing adoption of this type of policy in several Organization for Economic Cooperation and Development countries, it is important to compare experiences from different countries, because policy adoption in one country can be explained more comprehensively and coherently through comparison with similar policies adopted in other nations. To make a more meaningful cross-country comparison, this article identifies a taxonomy of health systems, and it analyzes institutional and resource-based factors related to health information exchange adoption and how they differ in three main types of health systems: the National Health Service, social health insurance, and private health insurance.

Smith, M., et al. (2019). "The Concept Dictionary and Glossary at MCHP: Tools and Techniques to Support a Population Research Data Repository." Int J Popul Data Sci 4(1): 1124.

The Manitoba Centre for Health Policy's Concept Dictionary and Glossary, and the Data Repository they document, broaden the analytic possibilities associated with administrative data. The aim of the Repository is to describe and explain patterns of health care and illness, while the Concept Dictionary and Glossary create consistency in documenting research methodologies. The Concept Dictionary alone contains detailed operational definitions and programming code for measures used in MCHP research that are reusable in future projects. Making these tools available on the internet allows reaching a heterogeneous audience of academic and government health service partners, epidemiologists, planners, programmers, clinicians, and students extending around the globe. They aid in the retention of corporate knowledge, facilitate researcher/analyst communication, and enhance the Centre's knowledge translation activities. Such documentation has saved countless hours for programmers, analysts and researchers who frequently need to tread paths previously taken by others.

Paris, V., et al. (2017). "New Health Technologies : Managing Access, Value and Sustainability". Paris OCDE

This report discusses the need for an integrated and cyclical approach to managing health technology in order to mitigate clinical and financial risks, and ensure acceptable value for money. The analysis considers how health systems and policy makers should adapt in terms of development, assessment and uptake of health technologies. The first chapter provides an examination of adoption and impact of medical technology in the past and how health systems are preparing for continuation of such trends in the future. Subsequent chapters examine the need to balance innovation, value, and access for pharmaceuticals and medical devices, respectively, followed by a consideration of their combined promise in the area of precision medicine. The final chapter examines how health systems can make better use of health data and digital technologies. The report focuses on opportunities linked to new and emerging technologies as well as current challenges faced by policy makers, and suggests a new governance framework to address these challenges.

(2015). "Health Data Governance: Privacy, Monitoring and Research". Paris OCDE.  
[http://www.oecd-ilibrary.org/social-issues-migration-health/health-data-governance\\_9789264244566-en](http://www.oecd-ilibrary.org/social-issues-migration-health/health-data-governance_9789264244566-en)

All countries are investing in health data, however; there are significant cross-country differences in data availability and use. Some countries stand out for their innovative practices enabling privacy-protective respectful data use; while others are falling behind with insufficient data and restrictions that limit access to and use of data, even by government itself. Countries that develop a data governance framework that enables privacy-protective data use will not only have the information needed to promote quality, efficiency and performance in their health systems, they will become a more attractive centre for medical research. After examining the current situation in OECD countries, a multi-disciplinary advisory panel of experts identified eight key data governance mechanisms to maximise benefits to patients and to societies from the collection, linkage and analysis of health data and to, at the same time, minimise risks to the privacy of patients and to the security of health data. These mechanisms include coordinated developing of high-value, privacy-protective health information systems; legislation that permits privacy-protective data use; open and transparent public communication ; accreditation or certification of health data processors; transparent and fair project approval processes; data de-identification and data security practices that meet legal requirements and public expectations without compromising data utility; and a process to continually assess and renew the data governance framework as new data and new risks emerge (résumé des éditeurs).

## L'utilisation des bases de données de santé par la recherche: les aspects méthodologiques

Benda, L., et al. (2023). "Health Data Hub et interopérabilité du Système national des données de santé." *Revue d'Epidémiologie et de Santé Publique* **71**: 101457.

<https://www.sciencedirect.com/science/article/pii/S0398762023000354>

**Introduction** La France dispose d'une des bases de données médico-administratives les plus complètes et homogènes au monde: le Système national des données de santé (SNDS). Cette base de données, initialement destinée à la gestion financière de l'Assurance maladie, a longtemps été sous-exploitée pour la recherche en santé en raison de sa complexité. Pour faciliter sa réutilisation, une piste de travail est sa standardisation via l'utilisation de modèles de données communs. **Méthodes** Depuis 2020, le Health Data Hub (HDH) transforme le SNDS vers le modèle OMOP-CDM (« Observational Medical Outcomes Partnership - Common Data Model »). Cette transformation permet de créer un modèle relationnel centré sur une table "patient" et de facilement reconstruire les parcours de soins. A partir d'un extrait du SNDS couvrant la période 2019-2020 pour une population hospitalisée pour COVID (SNDS Fast-Track), un alignement des schémas de données et des terminologies ont été réalisés. Les scripts de transformation sont développés en Python et les validations sont effectuées via les logiciels du consortium OHDSI. Des travaux d'harmonisation ont été réalisés avec des partenaires institutionnels (AP-HP, BPE). **Résultats** Ce travail a permis de passer d'une base de données de plus de 180 tables à moins de 20 tables. Les alignements de terminologies ont été réalisés par des internes en médecine sur plusieurs milliers de codes de différentes nomenclatures (CCAM, NABM, CSARR, etc.) vers SNOMED-CT. L'ensemble est disponible via la documentation collaborative du SNDS. Des travaux sont menés pour élargir le périmètre temporel (2015-2021). **Discussion/Conclusion** La standardisation des bases de données de santé assure leur normalisation et interopérabilité, rendant leur exploitation croisée au niveau national et international plus efficace. L'utilisation de modèles de données communs accélère le partage de données, de documentation et de programmes. Plusieurs initiatives européennes sont actuellement en cours telles que l'action conjointe TEHDAS et le pilote EHDS2 mené par le HDH. **Mots-clés** Interopérabilité ; SNDS ; OMOP-CDM ; Open source ; ETL **Déclaration de liens d'intérêts** Les auteurs déclarent ne pas avoir de liens d'intérêts.

Curmin, R., et al. (2022). "Algorithmes visant à identifier les patients asthmatiques dans les bases de données médico-administratives : une revue systématique de la littérature." *Revue d'Epidémiologie et de Santé Publique* **70**: S27-S28.

<https://doi.org/10.1016/j.respe.2022.01.111>

**Introduction** Les bases de données médico-administratives (BDMA) présentent de nombreux avantages pour la réalisation d'études sur l'asthme. Cependant, l'identification des patients asthmatiques nécessite l'utilisation d'algorithmes de repérage de cas basés sur les données de remboursement. L'objectif de cette étude était de décrire les algorithmes publiés permettant d'identifier les patients asthmatiques dans les BDMA. **Méthodes** Une revue de la littérature internationale des études publiées entre le 01/01/2016 et le 31/12/2020 dans Pubmed® a été réalisée. Les études identifiées par une équation de recherche comportant des termes MeSH ou du texte libre étaient sélectionnées si elles portaient sur l'asthme, étaient menées sur des BDMA, et si une méthode d'identification des patients asthmatiques était présentée. Les caractéristiques des études, les algorithmes utilisés, et leurs données de

validation étaient ensuite extraits à l'aide d'une grille standardisée assortie d'un guide de remplissage. A chaque étape, un échantillon était analysé par deux chercheurs et les discordances résolues par consensus. Résultats Trois cent trois études ont été sélectionnées totalisant 466 algorithmes dont 269 (57,7 %) algorithmes pour identifier les patients asthmatiques, 40 (8,6 %) reflétant les exacerbations d'asthme et 92 (19,7 %) pour caractériser la sévérité ou le contrôle de l'asthme. Ces études étaient majoritairement menées aux États-Unis (n=91, 30,0 %), à Taiwan (n=69, 22,8 %) ou au Canada (n=51, 16,8 %). La plupart des algorithmes identifiant les patients asthmatiques utilisaient uniquement des codes diagnostiques et leur origine – consultations externes ou en ville, hospitalisations (n=155, 57,6 %) ou les combinaient avec les délivrances de médicaments (n=60, 22,3 %). L'algorithme le plus utilisé pour identifier les patients asthmatiques était :  $\geq 1$  hospitalisation pour asthme et/ou  $\geq 2$  visites en soins ambulatoires pour asthme sur deux années. Seulement 73 (15,7 %) des algorithmes utilisés étaient rapportés comme validés. Discussion/Conclusion Différents algorithmes de repérage des cas d'asthme sont publiés dans la littérature internationale ; le choix d'un algorithme doit reposer sur sa pertinence par rapport au contexte (type d'asthme, objectif de l'étude, type de BDMA) et son éventuelle validation.

Karimi, M., et al. (2022). "Chaîner une base de données hospitalière avec les données de mortalité Insee." *Revue d'Epidémiologie et de Santé Publique* **70**: S69.

<https://doi.org/10.1016/j.respe.2022.03.089>

Introduction Appariement deux bases de données se définit par un ensemble de processus permettant de retrouver les individus d'une base dans une autre base. Dans le cadre des données hospitalières, retrouver le statut vital des patients est un enjeu très important, qui permettra à la fois de mener les études épidémiologiques sur ce type de données et à la fois d'avoir un intérêt logistique pour les hôpitaux. La mise en ligne des données du statut vital des personnes décédées en France donne la possibilité de retrouver le statut vital. L'objectif de ce travail est donc de proposer un algorithme d'appariement. Méthodes Nous avons utilisé les données du statut vital mise en ligne par l'Insee, accompagnées des variables identifiantes (à savoir, le nom de naissance, les prénoms, la date de naissance, le lieu de naissance et le sexe) et la base hospitalière Gustave Roussy, contenant les mêmes variables, dans le respect de la réglementation en vigueur sur la protection des données individuelles. L'algorithme proposé consiste en plusieurs étapes: i) Le nettoyage et la standardisation des deux bases, ii) Le blocage sur une ou plusieurs variables, iii) La comparaison de tous les couples possibles au sein du blocage en comparant judicieusement les variables d'appariement citées précédemment par un découpage des dates en jour, mois, année, des codes communes en département et commune et par la transformation phonétique des noms et prénoms et, iv) La classification des couples en deux classes appariés et non-appariés en utilisant un indice de similarité déduit des découpages des variables précédentes contraint à un seuil préalablement fixé. Le programme de l'algorithme proposé a été implémenté en logiciel SAS ([https://github.com/Oncostat/INSEE\\_DC](https://github.com/Oncostat/INSEE_DC)). L'algorithme a été validé sur l'ensemble des patients pour lesquelles le statut vital était disponible dans la base hospitalière de Gustave Roussy (collecté par le biais de GR, de la famille ou autre). Résultats Pour l'ensemble des nouveaux patients de Gustave Roussy entre 2015 et 2019 (99 017 patients), nous avons réussi à retrouver 8955 patients dans la base de l'Insee et donc mettre à jour les dates de décès dans la base de GR. Si on regarde les patients pour lesquels la date de décès est disponible dans la base GR, en moyenne nous avons un taux de 91 % d'appariement. Par exemple, parmi les 1671 nouveaux patients de 2016 et décédés depuis, nous avons apparié 1579 décès (taux de 94,5 %). Sachant que la base de l'Insee ne contient que les décès survenus en France, ce taux atteint à 98,6 % si l'on se concentre uniquement

sur les personnes nées en France. Conclusion Les processus proposés dans ce travail ont permis de démontrer qu'il est possible de retrouver le statut vital de façon fortement automatisée en utilisant un algorithme d'appariement inspiré des méthodes d'appariement indirectes. En effet ces algorithmes prennent en compte des multiples sources d'erreurs dans les bases hospitalières et pourraient être généralisés pour l'ensemble des établissements hospitaliers et notamment les registres de pathologie en France. Mots clés Appariement; Bases de données hospitalière; Données de mortalité Déclaration de liens d'intérêts Les auteurs n'ont pas précisé leurs éventuels liens d'intérêts.

CNIL (2020). "Guide pratique sur les modalités de circulation du NIR pour la recherche en santé aux fins d'appariement de données avec le SNDS". Paris CNIL.

<https://www.cnil.fr/fr/appariement-de-donnees-snds-numero-de-securite-sociale-nir-guide-pratique>

La CNIL publie un guide pratique afin d'aider les chercheurs désirant travailler avec les données du Système national des données de santé (SNDS) à mettre en œuvre un circuit d'appariement conforme aux exigences de sécurité.

Dubost, C. L. et Leduc, A. (2020). "Un appariement des données socio-économiques de l'échantillon démographique permanent au Système national des données de santé. Paris Drees

<https://drees.solidarites-sante.gouv.fr/publications/les-dossiers-de-la-drees/ledp-sante-un-appariement-des-donnees-socio-economiques-de>

Si le système national des données de santé apporte une grande quantité d'informations sur la consommation et les parcours de soins des bénéficiaires de l'Assurance maladie, son contenu est très limité pour caractériser la situation socio-économique de ces personnes. Dans l'optique d'évaluer la Stratégie nationale de santé 2018- 2022, et en particulier son objectif de réduire les inégalités sociales de santé, la DREES a donc piloté un projet visant à enrichir le SNDS de ce type de données en appariant l'échantillon démographique permanent de l'Insee au SNDS. La base de données résultant de cet appariement, appelée l'EDP-Santé, doit permettre de réaliser des études longitudinales sur les parcours de soins entre 2008 et 2022, à partir d'un échantillon d'un peu plus de 3 millions de personnes, qu'il est possible de croiser avec des données issues des fichiers fiscaux, du recensement (exhaustif, puis par enquête), du panel d'actifs tous salariés de l'Insee, du fichier électoral et de l'État civil. Ce document présente les contours de ce projet d'appariement, sa finalité et sa place dans le paysage des données de santé (partie 1), ses enjeux juridiques et les étapes de sa conception (partie 2), et enfin sa mise en œuvre ainsi que quelques premiers résultats méthodologiques sur sa qualité et ses perspectives d'utilisation (partie 3).

Safon, M. O. (2019). "Le repérage des maladies et du handicap dans les bases médico-administratives", Paris : Irdes

<https://www.irdes.fr/documentation/syntheses/le-reperage-des-maladies-et-du-handicap-dans-les-bases-medico-administratives.pdf>

Cette bibliographie a pour objectif de recenser de la littérature scientifique sur le repérage des maladies et du handicap dans les bases médico-administratives.

Gissot, C. (2018). "Les données médico-administratives de l'Assurance maladie : les utilisations par la Cnam". In: [Santé et intelligence artificielle.]. Paris : CNRS éditions: 75-84.

(2017). Le réseau REDSIAM pour améliorer l'utilisation du SNIIRAM-SNDS, le système national d'information inter-régimes de l'assurance-maladie. Rev Epidemiol Santé Publique **65**: suppl. 5

<https://www.em-consulte.com/revue/RESPE/65/S4/table-des-matieres/>

Bergeat, M. (2016). "La gestion de la confidentialité pour les données individuelles". Paris INSEE.

Ce document de travail effectue un recensement non exhaustif des méthodes de protection des fichiers de données individuelles, en se focalisant sur les techniques implémentées dans les logiciels de protection des données individuelles les plus utilisés en Europe,  $\mu$ -Argus et le package R sdcMicro. Différentes méthodes existant dans la littérature sont présentées en suivant les 3 étapes principales utilisées lors de l'anonymisation des données individuelles : mesure du risque de réidentification, techniques de protection pour réduire le risque de réidentification, mesure de la perte d'information suite à l'anonymisation des données. Quelques exemples d'expériences méthodologiques et de choix de diffusion opérés en France et en Europe sont ensuite discutés. Ce document de travail s'appuie essentiellement sur (Hundepool et al., 2012).

Ferdynus, C. et Hulard, L. (2016). "Optimisation de la constitution de cohortes issues de bases de données médico-administratives : mise à disposition d'un algorithme pour l'intégration et la normalisation des données adapté au Système national d'information inter-régimes de l'assurance maladie (SNIIRAM)." *Revue d'Epidémiologie et de Santé Publique* **64**(4): 263-269.

<http://www.em-consulte.com/article/1080254/alerteM>

Goldberg, M., Carton, M., Gourmelen, J., et al. (2016). "L'ouverture du Système national d'information inter-régimes de l'assurance maladie (SNIIRAM) : des opportunités et des difficultés. L'expérience des cohortes Gazel et Constances." *Rev Epidemiol Sante Publique* **64**(4): 313-320.

Le Système national d'information inter-régimes de l'assurance maladie (SNIIRAM) comporte des données d'hospitalisation et de consommation de soins concernant plus de 60 millions de personnes. Bien qu'elles n'informent pas sur de nombreuses données personnelles et environnementales, les données du SNIIRAM ont un intérêt majeur pour l'épidémiologie, la surveillance et la santé publique. Une de ses utilisations particulièrement intéressante est l'appariement avec des données d'enquêtes recueillant des informations directement auprès des personnes. D'accès encore relativement restreint, son ouverture va être largement facilitée par les prochaines évolutions législatives, réglementaires et organisationnelles. Son utilisation nécessite cependant un important travail technique et méthodologique, de contrôle et de validation, du fait qu'il s'agit d'une base de données gigantesque et d'une très grande complexité, qui concerne trois aspects principaux : la volumétrie, l'architecture des données et l'interprétation de celles-ci. Des outils sont en cours de développement afin d'utiliser le SNIIRAM pour enrichir les données des cohortes Gazel et Constances : documentation interactive regroupant les informations sur le SNIIRAM, logiciel d'aide au recettage des données du SNIIRAM, méthodes de calcul à partir des données brutes de variables indicatrices permettant de noter la présence d'un événement (acte, consommation, diagnostic, etc.) pour chaque participant des cohortes, et de variables synthétiques permettant de produire un ensemble de variables sur un domaine spécifique (médicaments en soins de ville, actes, diagnostics au cours d'un séjour hospitalier.). Par ailleurs, le réseau REDSIAM mis en place récemment a pour objectif de développer, évaluer et mettre à disposition des algorithmes pour identifier des pathologies dans le SNIIRAM. Si l'on veut que le potentiel de la base de données exceptionnelle que constitue le SNIIRAM soit utilisé à sa juste valeur, il est indispensable de développer les outils permettant de faciliter son utilisation.

Berlorgey, N., Le Gléau, J.-P. et Goldberg, M. (2015). "Deux débats sur les données." Statistique Et Societe 3(1)  
<http://statistique-et-societe.fr/issue/view/50>

Loth, A., Le Gléau, J.-P. et Blum, D. (2015). "Données de santé : anonymat et risque de ré-identification." Dossiers Solidarité Et Santé (Drees)(64)  
<https://www.casd.eu/wp/wp-content/uploads/dss64-2.pdf>

Ce dossier présente les résultats du travail préparatoire mené par la DREES pour élaborer l'article 47 du projet de loi de modernisation de notre système de santé voté en avril 2015. Quels sont les risques de ré-identification pour des bases de données en apparence anonymes ? Comment encadrer leur accès et permettre l'utilisation des données pour le bénéfice de tous, sans mettre en danger le droit à la protection de la vie privée ? Sont aussi évoquées les avancées du projet de loi pour la simplification des procédures, notamment le numéro national d'identification, les appariements et le rôle d'un tiers de confiance. Le débat juridique sur la définition de l'anonymat est également abordé : existe-t-il des critères absolus ou doit-on s'en tenir aux moyens susceptibles d'être raisonnablement mis en œuvre pour identifier une personne ? La formulation de la loi française est plus exigeante que celle de la directive européenne. Est-ce bien raisonnable ?

(2014). "Rapport de la Commission Open data en santé". Paris Ministère chargé de la Santé.  
<https://drees.solidarites-sante.gouv.fr/publications/rapports/rapport-de-la-commission-open-data-en-sante>

La Commission " 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. Son rapport, remis le 9 juillet 2014 à la ministre de la Santé, retrace les travaux et discussions de la Commission. Adopté consensuellement par l'ensemble des membres de la commission, qui partagent des attentes communes et fortes, il dresse tout d'abord un panorama de l'existant : 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. Dans une seconde partie, il évoque les enjeux pour l'avenir. Enfin, il identifie les actions à mener : données à ouvrir en open data, orientations en matière de données réidentifiantes, données relatives aux professionnels et aux établissements.

Goldberg, M. (2012). "Pour une meilleure utilisation des bases de données administratives et médico-administratives nationales pour la santé publique et la recherche". Paris HCSP.  
[http://www.hcsp.fr/docspdf/avisrapports/hcspr20120309\\_bddadministration.pdf](http://www.hcsp.fr/docspdf/avisrapports/hcspr20120309_bddadministration.pdf)

La France dispose de bases de données médico-sociales et économiques nationales centralisées, constituées et gérées par des organismes publics, couvrant de façon exhaustive et permanente l'ensemble de la population dans divers domaines stratégiques pour la santé publique et la recherche : recours aux soins, hospitalisation, handicaps, prestations et situation professionnelle, sociale et économique. De plus, un identifiant individuel unique (le NIR : Numéro d'identification au répertoire) est actuellement utilisé par pratiquement toutes les bases de données nationales. Malgré certaines limites en termes de couverture, de qualité et de validité des données, ces bases de données, concernant plus de 60 millions de personnes, constituent un patrimoine considérable, vraisemblablement sans équivalent au monde. Cependant, l'utilisation à des fins de recherche et de surveillance de ces bases de données nationales se heurte actuellement à des obstacles divers, dont les plus importants

sont de nature juridique et opérationnelle. En réponse à une saisine de la Direction générale de la santé (DGS), ce document fait le point sur les principales bases existantes, expose les difficultés pour leur utilisation et présente de façon détaillée une série de propositions pour les surmonter.

Le Gléau, J. P. et Royer, J. F. (2011). "Le centre d'accès sécurisé aux données de la statistique publique française : un nouvel outil pour les chercheurs." *Courrier Des Statistiques*(130): 1-5.  
<https://www.comite-du-secret.fr/wp-content/uploads/2017/12/Le-CASD-J-P-Le-Gl%C3%A9au-J-F-Royer.pdf>

En 2010, l'Insee a mis en place un nouveau dispositif d'accès aux fichiers individuels de la statistique publique, essentiellement à destination des chercheurs : le CASD, « centre d'accès sécurisé distant ». Rendu possible par l'évolution récente de la loi de 1951 sur le secret en matière de statistiques, il permet d'accéder à des données individuelles selon une procédure très réglementée. Cet article en donne une présentation générale.

Chevreur, K., Le, Fur P., Renaud, T., et al. (2006). "Faisabilité d'un système d'information public sur la médecine de ville". Paris IRDES.  
<http://www.irdes.fr/Publications/Rapports2006/rap1648.pdf>

Cette étude s'inscrit dans le cadre d'une réflexion sur le développement d'un système d'information public permanent sur la médecine libérale en France. En effet, il n'existe actuellement aucun système public permettant de connaître les motifs de recours de la population aux médecins libéraux et de les lier avec les prescriptions. En l'absence de cette information, un volet entier des pratiques professionnelles échappe à l'évaluation, alors même que les différents acteurs du système de santé s'accordent pour dire toute l'importance qu'il y aurait à le faire. Dans cette étude, nous faisons le point sur les différents recueils informatisés existant en France début 2005, qu'ils proviennent de sources administratives, de sociétés privées ou de sociétés savantes. Nous recensons et décrivons également les différents systèmes mis en place à l'étranger et accessibles aux pouvoirs publics. Enfin, en fonction des niveaux d'informations souhaités, nous développons différentes propositions allant de la création d'un nouvel observatoire des pratiques médicales à l'utilisation de bases de données existantes.

Goldberg, M. (2006). "Les bases de données d'origine administrative peuvent-elles être utiles pour l'épidémiologie ?" *Rev Epidemiol Sante Publique* **54**: 297-303.

[BDSP. Notice produite par ORSLR 33JR0xnY. Diffusion soumise à autorisation]. Comme le montre l'expérience de certains pays, l'utilisation de bases de données d'origine administrative peut grandement faciliter les travaux des épidémiologistes pour de nombreuses procédures, voire améliorer la qualité des études dans certaines situations. Il reste cependant, en France, de nombreux problèmes d'ordre financier, organisationnel et technique à résoudre pour leur utilisation optimale.

Weill, A. et Fender, P. (2005). "Épidémiologie, santé publique et bases de données médico-tarifaires." *Revue Médicale De L'assurance Maladie*(2): 163-168.

[BDSP. Notice produite par CNAMTS p3AsQR0x. Diffusion soumise à autorisation]. L'ensemble des bases de données médico-tarifaires comprend trois catégories : le PMSI, les bases de remboursement des soins de l'assurance maladie et celles constituées par les services médicaux des caisses d'assurance maladie. Les pouvoirs publics et l'assurance

maladie ont décidé de créer un ensemble de bases de données anonymes appelé le Système national d'information inter-régimes de l'assurance maladie (SNIRAM). Ces bases ont été créées avec une finalité de régulation et de suivi des dépenses de santé, de compréhension du système de soins et d'allocation des ressources. Elles sont également des outils puissants d'épidémiologie descriptive, de pharmaco-épidémiologie et de veille sanitaire.

## La protection des données personnelles dans le cadre de la recherche

(2023). "Plateformes de santé : enjeux d'éthique". Paris CCNE

<https://www.ccne-ethique.fr/fr/publications/avis-143-du-ccne-et-5-du-cnpen-plateformes-de-donnees-de-sante-enjeux-dethique>

Le comité consultatif national d'éthique et le comité national pilote d'éthique du numérique font vingt et une recommandations dans un avis sur les enjeux éthiques des plateformes de données de santé. Parmi eux, la notion du consentement des patients qui doit concilier respect de la vie privée et participation à l'intérêt commun de santé publique.

Peugeot, V. (2023). "La protection des données à caractère personnel à l'épreuve de la crise sanitaire." *Annales des mines*: 134-138.

[https://www.annales.org/enjeux-numeriques/2023/en\\_21\\_03\\_23.html](https://www.annales.org/enjeux-numeriques/2023/en_21_03_23.html)

La pandémie liée à la Covid-19 a mis à l'épreuve la capacité du droit des données à caractère personnel à accompagner une situation d'urgence sanitaire sans équivalent. L'expérience a montré, d'une part, la robustesse et la souplesse de ces textes, et, d'autre part, la réactivité de l'institution en charge de veiller à leur application. Dans des délais extrêmement courts, le gouvernement a pu être accompagné au plus près par la CNIL dans la mise en place de dispositifs inédits, veillant à ce que les dérogations consenties soient minimales, transitoires et contrôlées, pendant que les chercheurs obtenaient dans des délais record des autorisations pour traiter les données de santé.

Lesaulnier, F. (2022). "Valorisation de la recherche en santé humaine et protection des données à l'ère du numérique." *Annales des Mines-Réalités industrielles*(3): 30-36.

<https://www.cairn.info/revue-realites-industrielles-2022-3-page-30.htm>

La recherche en santé humaine connaît une révolution numérique du fait des masses considérables de données disponibles, qui sont collectées dans des environnements multiples, et de la possibilité d'en extraire des connaissances et des corrélations grâce aux technologies qui augmentent les capacités de stockage et de traitement. À l'heure des grands défis de la Science ouverte, le RGPD conduit à un recentrage de l'organisation de la protection des données sur les organismes qui les traitent. Le respect de la réglementation doit être intégré dans une démarche globale de gouvernance des données et requiert l'implication des personnes à l'origine des données. Nous présenterons ici quelques pistes de réflexion et d'action développées par les acteurs de la recherche pour une utilisation éthique et responsable des données personnelles de santé

Hulier-Ammar, E., et al. (2022). "Recherche sur données : aspects juridiques et éthiques à travers l'expérience de l'hôpital Foch." *Médecine & Droit* **2022**(172): 8-14.

<https://www.sciencedirect.com/science/article/pii/S1246739121000622>

Résumé Les données de santé font l'objet de convoitise par de nombreux acteurs, bien entendu par les industriels de la santé, notamment du médicament et du dispositif médical, par les hôpitaux et les instituts de recherche, mais aussi par tout type d'entreprises marchandes qui souhaiteraient en retirer des bénéfices conséquents. Face à ce phénomène, les instances de l'Union Européenne, dans le cadre du règlement européen de 2016 (RGPD) remplaçant la directive européenne sur les données de santé de 1995, ont renforcé la protection des données de santé qui sont des données particulièrement sensibles pour tout

un chacun et, en conséquence, la réalisation des études sur les données de santé. Les recherches n'impliquant pas la personne humaine, improprement appelées « Études sur données », sont très encadrées par le RGPD et, pour la France, par la Loi Informatique et Libertés (LIL) qui a été modifiée en conséquence. Cet article expose les étapes à réaliser pour la mise en place d'études sur données et les droits des participants/patients dans le cadre de ces recherches. Une vigilance éthique demanderait à ce que ces recherches soient examinées par un comité ad hoc. Health data is the object of covetousness by many actors, of course by the healthcare industry, in particular the drug and medical device industry, by hospitals and research institutes, but also by all types of merchant companies which would like to derive substantial benefits from it. Faced with this phenomenon, the European Commission, within the framework of the replacement of the European Directive on health data of 1995, by the General Data Protection Regulation of 2016 (GDPR) has strengthened the protection of health data which is particularly sensitive data for everyone and, consequently, carrying out studies on health data. Researches that does not involve human beings, improperly called "Data studies", are closely regulated by the GDPR and, for France, by the Data Protection Act, which has been amended accordingly. This article describes the steps to be taken for the implementation of data studies and what are the rights of participants/patients in the context of these researchs. Ethical vigilance would require that such research be reviewed by an ad hoc committee.

Lombrail, P., et al. (2022). "Note d'étape sur le Health Data Hub, les entrepôts de données de santé et les questions éthiques posées par la collecte et le traitement de données de santé dites « massives »". *sl Hal*

<https://www.hal.inserm.fr/inserm-03533863>

Le CEI a mis en place un groupe de travail en octobre 2020 du fait des interrogations soulevées par la décision de confier l'hébergement des données du Système national des Données de Santé (SNDS) rassemblées par le Health Data Hub (HDH ou PDS pour Plateforme des données de santé) à la société Microsoft à travers son « cloud » Azure. Le groupe a été amené rapidement à élargir sa réflexion à un ensemble plus vaste de questions éthiques soulevées par la collecte et le traitement de données dites « massives » pouvant s'apparenter de près ou de loin à des données de santé.

(2021). "The protection of personal data in health information systems - principles and processes for public health". Copenhagen OMS.

<https://www.euro.who.int/en/data-and-evidence/evidence-informed-policy-making/publications/2021/the-protection-of-personal-data-in-health-information-systems-principles-and-processes-for-public-health-2021>

In recent years, countries across Europe have implemented either new or considerably stricter data protection and cybersecurity laws. These laws continue to have a substantive impact on health information systems (HIS) and most public health activities in a wider sense. This document aims to explore the conceptual implications and to give some guidance on how specific decisions that are unavoidable to balance the rights and interests at stake should be taken. With a few easy-to-implement steps, any organization in public health can increase its level of data protection compliance significantly. As data protection is based on principles that have evolved over time, section 2 gives a short historical overview, followed by a deep dive into the legal principles behind data protection. Section 3 covers the practical implications of these principles and addresses the rights of data subjects, as these are at the heart of the regulatory framework. Section 4 examines the elements that need to be balanced against these rights – in particular, the right to health and to public health in

general. Section 5 looks again at the secondary use of data for public health purposes, and at how the balancing of the interests at stake works in this context. Finally, section 6 gives an overview of the steps to be taken to make this happen, such as empowerment and oversight mechanisms. This guidance document is part of the WHO Regional Office for Europe's work on supporting Member States in strengthening their health information systems. Helping countries to produce solid health intelligence and institutionalized mechanisms for evidence-informed policy-making has traditionally been an important focus of WHO's work and continues to be so under the European Programme of Work 2020–2025.

(2021). "Rgpd : guide pour la recherche SHS". Paris : CNRS

[https://inshs.cnrs.fr/sites/institut\\_inshs/files/pdf/guide-rgpd\\_2.pdf](https://inshs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd_2.pdf)

Bévière-Boyer, B. (2021). "La protection des données de santé mises à disposition par le Health Data Hub pour les recherches sur la Covid-19." *Journal du Droit de la Santé et de l'Assurance - Maladie (JDSAM)* 29(2): 37-48.

<https://www.cairn.info/revue-journal-du-droit-de-la-sante-et-de-l-assurance-maladie-2021-2-page-37.htm>

La Plateforme nationale de données de santé, appelée aussi « Health Data Hub » fut instituée par l'arrêté du 29 novembre 2019 et lancée par le biais d'un plan stratégique 2019-2022 en janvier de la même année. Cette nouvelle structure a pour finalité « de réunir, organiser et mettre à disposition les données du système national des données de santé mentionné à l'article L.1461-1 et de promouvoir l'innovation dans l'utilisation des données de santé ». Pas encore opérationnelle lors de l'apparition de la pandémie Covid-19, elle le fut de manière assez précipitée par l'arrêté du 21 avril 2020. Cette plateforme, ayant le statut juridique de groupement d'intérêt public, devint ainsi un objectif prioritaire de santé publique lors de la pandémie de Covid-19..

Bévière-Boyer, B. (2021). "La gestion des données de santé par le Health Data Hub : le recours à la société Microsoft, entre risques et précautions." *Droit, Santé et Société* 3(3): 42-48.

<https://www.cairn.info/revue-droit-sante-et-societe-2021-3-page-42.htm>

Alors même que la nouvelle plateforme des données de santé française, appelée aussi Health Data Hub est dans la tourmente, puisque critiquée pour avoir choisi la société Microsoft, notamment au regard de l'arrêt Schrems de la Cour de Justice de l'Union européenne du 16 juillet 2020, elle reste soutenue par le Conseil d'État lors de son ordonnance du 13 octobre 2020. L'étude de cette affaire permet de mieux appréhender les risques, ainsi que les précautions à prendre pour la gestion des données de santé au niveau national et communautaire.

Cayol, A. (2021). "La protection des données personnelles de santé en France et en Europe par le Règlement Général sur la Protection des Données (RGPD)." *Droit, Santé et Société* 3(3): 49-54.

<https://www.cairn.info/revue-droit-sante-et-societe-2021-3-page-49.htm>

Le développement du numérique dans le domaine de la santé a profondément modifié la question de la protection des informations relatives aux patients en permettant de conserver des masses de données importantes, d'y avoir accès plus facilement et de les transmettre plus largement. L'accès à toutes ces données présente un intérêt majeur pour les tiers dits « intéressés ». Afin de préserver les droits et libertés fondamentaux, le traitement des données à caractère personnel a rapidement été encadré en France, puis en Union européenne. Du fait de leur lien étroit avec la personne concernée, les données de santé font

l'objet, au sein de l'Union européenne, d'une protection renforcée par rapport aux autres données à caractère personnel. Le RGPD offre une avancée notable sur ce point : définissant pour la première fois expressément les données de santé, il en retient en effet une conception extensive. Se saisissant de la marge de manœuvre laissée aux États membres par le RGPD, la France a, en outre, prévu des exigences supplémentaires pour leur traitement dans un certain nombre de cas.

de Fallois, M. (2021). "Les traitements de données de santé à des fins de recherche liés à la COVID-19 : quelle régulation par la CNIL ?" *Journal du Droit de la Santé et de l'Assurance - Maladie (JDSAM)* **29**(2): 26-32.

<https://www.cairn.info/revue-journal-du-droit-de-la-sante-et-de-l-assurance-maladie-2021-2-page-26.htm>

Les traitements de données à caractère personnel mis en œuvre dans le cadre de recherches, d'études ou d'évaluations dans le domaine de la santé relèvent de la loi n° 78-17 du 6 janvier 1978 modifiée (dite « loi informatique et libertés ») depuis l'introduction de dispositions particulières en son sein par l'une des lois de bioéthique de 1994. Ces dispositions, qui ont été modifiées à plusieurs reprises, notamment suite à l'entrée en vigueur du Règlement général sur la protection des données figurent désormais aux articles 64 et suivants de la loi « informatique et libertés » au sein d'une section spécifique aux traitements de données à caractère personnel dans le domaine de la santé. En sus de l'application du RGPD et de la loi « informatique et libertés », ces traitements sont également soumis à certaines dispositions spécifiques contenues dans le code de la santé publique et le code civil.

Lesaulnier, F. (2021). "La protection des données personnelles au cœur de la recherche Covid-19 à l'Inserm." *Journal De Droit De La Santé et de L' Assurance Maladie*(29): 33-36.

<https://institutdroitsante.fr/publications/publications-ids/revues-livres/journal-de-la-sante-et-de-l-assurance-maladie-jdsam/jdsam-n29-juillet-2021/>

Robin, A. (2021). "Chapitre 3. Open data et santé : quelles modalités pour la diffusion et l'exploitation des données de santé ?" *Journal international de bioéthique et d'éthique des sciences* **32**(2): 33-44.

<https://www.cairn.info/revue-journal-international-de-bioethique-et-d-ethique-des-sciences-2021-2-page-33.htm>

La diffusion et l'exploitation des données de santé interrogent en premier lieu la production des données de santé et, en particulier, la production des données de recherche en santé. En outre, parler de l'ouverture des données oblige aussi à comprendre comment les données de santé sont produites au plan scientifique et surtout comment la diffusion de cette production a été encadrée par le législateur. Enfin, il convient de rappeler que les données de santé, y compris les données de recherche en santé, sont le plus souvent des données personnelles, auxquelles s'applique un dispositif légal de protection des données qui est aujourd'hui régi en Europe par le Règlement général sur la protection des données.

Soupizet, J.-F. (2021). "Données de santé françaises : une alerte. France et Europe face aux géants américains et chinois." *Futuribles* **440**(1): 81-88.

<https://www.cairn.info/revue-futuribles-2021-1-page-81.htm>

En 2019 a été créé, en France, le Health Data Hub (HDH), chargé de regrouper toutes les données de santé françaises. L'hébergement de ces données a été confié six mois plus tard à Microsoft, une multinationale soumise à une législation américaine peu compatible avec les

règles européennes visant à leur sauvegarde et à l'indépendance de l'Union européenne. Jean-François Soupizet, prenant appui sur cette décision et ses conséquences potentielles, montre en quoi elle constitue un précédent problématique, en raison du défi plus large que constitue la maîtrise des données européennes dans un contexte dominé par les États-Unis et par la Chine. Ainsi nous alerte-t-il sur un défi majeur face auquel l'Union européenne et ses États membres ont pris un retard important par rapport à leurs concurrents, et sur les dangers qui peuvent en résulter au regard de la protection des libertés et de l'indépendance des Européens. H.J.

(2020). "Commission nationale de l'informatique et des libertés. Rapport d'activité 2019". Paris CNIL. [https://www.cnil.fr/sites/default/files/atoms/files/cnil-40e\\_rapport\\_annuel\\_2019.pdf](https://www.cnil.fr/sites/default/files/atoms/files/cnil-40e_rapport_annuel_2019.pdf)

Le constat est indéniable : un an après l'entrée en application du RGPD, l'année 2019 démontre que le RGPD est au cœur des préoccupations des Français et des Européens. Cette année a été marquée par un nombre toujours plus élevé de plaintes adressées à la CNIL et par la coopération avec ses homologues européens devenue une réalité quotidienne. L'année 2019 en chiffres, c'est notamment plus de 14 000 plaintes, soit plus de 27 % d'augmentation par rapport à l'année précédente, dont 20 % de plaintes transfrontalières, 145 000 appels, 8 millions de visites sur notre site web et 17 300 requêtes électroniques sur Besoin d'aide, soit 12 % d'augmentation sur la même période. En parallèle, près de 65 000 organismes ont désormais déclaré un délégué à la protection des données (DPO) (tiré du texte).

Ragot, N. (2020). "Le cycle de vie des données à caractère personnel relatives à la santé. Étude du cadre légal applicable à la recherche universitaire". Bordeaux, LEH Édition, 2020, Mémoires numériques de la BNDS.

Dans le cadre de la recherche en milieu universitaire, des données relatives à l'état physique, mental et social d'individus sont collectées et exploitées à l'occasion de projets menés par des chercheurs issus de formations relevant notamment des domaines médicaux et des sciences humaines. De par leur nature de données à caractère personnel relatives à la santé, le traitement de ces informations impose de garantir les droits des participants à la recherche en adoptant des mesures techniques et organisationnelles adéquates et, dans certains cas, en soumettant directement le projet à l'autorisation de la Commission nationale de l'informatique et des libertés. Encadrée par le règlement général sur la protection des données et la loi informatique et libertés, la démarche de mise en conformité dès le montage du projet nécessite la collaboration entre le chercheur et le Délégué à la protection des données compétent, afin d'articuler l'éthique de la protection de la vie privée avec celle relative à la personne humaine, tel qu'il en ressort de la loi Jardé, tout en tenant compte des aspects relatifs au droit d'auteur, à l'open data et aux archives. C'est à travers une approche didactique que ce mémoire a pour but d'accompagner le chercheur dans l'élaboration de son plan de gestion des données, en abordant, pour chaque étape du cycle de vie des données du projet de recherche, les problématiques juridiques et réglementaires, ainsi que les mesures de sécurisation adaptées.

(2019). "Code de la protection des données personnelles (édition 2020)", Paris : Dalloz <https://www.editions-dalloz.fr/code-de-la-protection-des-donnees-personnelles.html>

À jour de son décret d'application du 29 mai 2019

(2019). "Commission nationale de l'informatique et des libertés. Rapport d'activité 2018". Paris CNIL.

[https://www.cnil.fr/sites/default/files/atoms/files/cnil-39e\\_rapport\\_annuel\\_2018.pdf](https://www.cnil.fr/sites/default/files/atoms/files/cnil-39e_rapport_annuel_2018.pdf)

L'entrée en application du RGPD a marqué une prise de conscience inédite des enjeux de protection des données auprès des professionnels et des particuliers. Cela s'est logiquement traduit par une augmentation considérable des plaintes adressées à la CNIL, avec une tendance à la hausse qui s'installe. La CNIL a aussi reçu un afflux de demandes d'information de la part des professionnels souhaitant s'approprier ce nouveau cadre, et ce, tous canaux confondus.

Les autres rapports annuels de la Cnil sont disponibles à cette adresse :

<https://www.cnil.fr/fr/mediatheque/rapports-annuels>

(2019). "Mémento RGPD. Sensibilisation au règlement général sur la protection des données". Paris MSSPS.

<https://solidarites-sante.gouv.fr/systeme-de-sante-et-medico-social/e-sante/sih/article/memento-rgpd>

L'organisation de l'offre de soins doit évoluer en permanence et s'adapter aux besoins du patient et aux nouveaux enjeux de santé publique. Pour renforcer, par exemple l'accès territorial aux soins, il faut pouvoir tirer tout le bénéfice des technologies de l'information et des communications. Le numérique favorise en effet l'échange, le partage, et donc le décloisonnement entre les différents acteurs (offreurs de soins et usagers). Il favorise la prise en charge ambulatoire en contribuant à des organisations plus efficaces et réduit les distances en permettant de développer la télémédecine. Il porte également, du fait de la masse de données automatisées stockées, un immense potentiel d'information. Tous ces nouveaux outils, ces nouvelles organisations, ne peuvent se déployer au quotidien que si les patients et les professionnels de santé leur accordent une pleine confiance. Or, dans un environnement où la numérisation s'accélère et devient omniprésente, de nouveaux risques apparaissent. Il est donc essentiel d'être en situation de pouvoir les comprendre, les évaluer afin de mieux les maîtriser. Les méthodes et outils de la sécurité numérique ne manquent pas, mais ils perdent toute efficacité s'ils ne sont pas soutenus en permanence. Les promouvoir et accompagner leur mise en œuvre relève de la responsabilité des établissements de santé et notamment de celle de leurs managers. Avec l'entrée en application, le 25 mai 2018, du Règlement Général sur la Protection des Données (RGPD) et la promulgation le 20 juin 2018 de la loi relative à la protection des données personnelles, un nouveau cadre juridique est posé. Il vise à rendre au citoyen le contrôle de ses données personnelles et renforce la responsabilité des organismes qui les gèrent en leur demandant d'assurer une protection optimale à chaque instant. Le présent mémento a pour objectif d'éclairer les décideurs sur les conséquences, pour les établissements de santé, de ce nouveau contexte réglementaire et d'en préciser les enjeux. Un mémo-quizz « RGPD » disponible à la fin du document propose des repères afin de faciliter la mise en œuvre de la conformité.

(2019). "Ouvrage collectif SSI Santé". Paris Apssis

<http://www.apssis.com/#/publications-apssis/4395697>

La rédaction d'articles et de retours d'expériences a été proposée à un large panel de professionnels de la SSI Santé. L'objectif était de donner la parole au terrain, par la plume de celles et ceux qui « exercent la SSI », qui mettent en œuvre les principes de la cybersécurité au cœur des organisations, tant sur le volet « technique » que sur le volet « politique ».

L'ouvrage est organisé en 7 thématiques : Normes et référentiels, Gouvernance de la SSI

Santé, Technologies de sécurité, RGPD et cybersécurité, Processus et procédures, Conformité et audits, et Prospective.

Debiès, É. (2019). "Renforcement des droits des individus sur leurs données personnelles : quelles conséquences sur l'utilisation du numéro d'inscription au répertoire national d'identification des personnes physiques (NIR) ?" *Regards* 55(1): 149-155.  
<https://www.cairn.info/revue-regards-2019-1-page-149.htm>

Le Règlement européen Général sur la Protection des Données (RGPD), en plaçant la personne concernée au centre de l'utilisation de ses données, avec des droits renforcés et une place prépondérante conférée à son consentement, est venu percuter l'équilibre juridique propre à l'utilisation du NIR, qui au fil des années a toujours fait l'objet de dispositions spécifiques encadrant ses usages, sans que le législateur et les opérateurs ne se préoccupent particulièrement de la transparence vis-à-vis des personnes concernées. Certes le cadre juridique propre à l'utilisation du NIR était devenu depuis 2004 (Loi n° 2004-801 du 6 août 2004 relative à la protection des données) et au fil des années, d'une lecture et mise en œuvre complexes. Le RGPD permet, par la suppression générale des formalités préalables et la possibilité spécifique offerte aux Etats de prévoir un régime particulier pour les identifiants nationaux, saisie en France pour encadrer les utilisations du NIR, une plus grande lisibilité et simplification des conditions d'usage de ce numéro. Cette transparence nouvelle répond en outre à un autre objectif du RGPD : un meilleur contrôle de ses données par la personne concernée, qui repose avant tout sur l'information relative aux traitements dont ses données sont l'objet. Cette information sur les traitements comportant l'usage du NIR, accessible notamment grâce au décret du 19 avril 2019 laisse percer de nouvelles interrogations, que nous soulèverons dans la première partie. Nous examinerons ensuite la manière dont ce régime juridique spécifique peut s'articuler avec le renforcement des droits des personnes sur leurs données personnelles.

Gadouche, K. (2019). "Le Centre d'accès sécurisé aux données (CASD), un service pour la data science et la recherche scientifique." *Courrier Des Statistiques*(3): 76-92.  
<https://www.insee.fr/fr/information/4254227?sommaire=4254170>

L'accès des chercheurs aux données individuelles collectées par le Service statistique public constitue un enjeu scientifique majeur. Ces informations très détaillées exigent un très haut niveau de sécurité pour éviter toute dissémination préjudiciable au citoyen, ou toute utilisation par un tiers non autorisé. Pour répondre à ce besoin de sécurité, l'Insee a créé en 2010 le CASD, Centre d'accès sécurisé aux données : les équipes du centre ont conçu un équipement sécurisé, permettant un accès à distance, tout en garantissant une authentification forte des utilisateurs et un confinement des fichiers. Le CASD, désormais autonome, s'est développé au fil du temps, élargissant son champ à d'autres producteurs et à d'autres types de données très détaillées et sensibles telles les données de santé ou les données administratives. Ce service propose de nouvelles solutions à la problématique des appariements et de la reproductibilité des études sur données confidentielles. Le CASD est de plus en plus utilisé par la communauté de la recherche en France, et son expérience originale, bien que relativement récente comparé à celle de ses partenaires étrangers, lui permet de s'ouvrir à l'international.

Le Gléau, J. P. (2019). "Le secret statistique", Les Ullis : EDP Sciences

Pour établir leurs statistiques, l'Insee et les autres services statistiques publics récoltent de nombreuses informations auprès de nous. Ces informations sont-elles bien protégées ?

Finalement, qui peut y avoir accès ? Le secret statistique garantit que nos données ne seront pas divulguées à des tiers. Les règles de mise en œuvre du secret statistique se sont adaptées aux mouvements de la société et à l'évolution des techniques. Il est désormais possible de donner accès à des informations statistiques très détaillées pour les chercheurs, sans rompre la confidentialité. Ces techniques, mises au point pour la statistique, sont maintenant utilisées pour mettre à la disposition des chercheurs des données fiscales, sur la santé, etc. tout en respectant la protection des données personnelles. Cet ouvrage est le premier à faire un bilan complet des règles et pratiques qui permettent de garantir le secret statistique. Une première partie définit l'évolution du secret statistique depuis la guerre. Un chapitre pratique détaille les règles à observer dans la diffusion des statistiques pour préserver la confidentialité des données. Un développement particulier est ensuite fait sur la question difficile de la diffusion de données individuelles. Enfin quelques exemples pris dans d'autres pays donnent un éclairage sur les techniques retenues ailleurs.

Mascret, C. (2019). "Quel statut pour les données de santé ? In : Panorama de droit pharmaceutique - 2018." Revue Générale De Droit Médical(6): 137-147.

La révolution des technologies et l'irruption du digital et du big data en santé ont fait prendre conscience de la valeur que possédaient les données de santé tant sur le plan de la recherche, que l'on connaissait jusqu'alors, mais également sur un plan commercial. Alors que l'Europe vient de réviser sa réglementation afin de prendre en considération cette nouvelle donnée, il est intéressant de s'attarder sur le statut que possèdent ces données de santé tant convoitées.

(2018). "Guide pratique sur la protection des données personnelles". Paris CNIL.

<https://www.cnil.fr/sites/default/files/atoms/files/guide-cnom-cnil.pdf>

Le guide pratique, élaboré et rédigé conjointement par le CNOM et la CNIL, accompagne les médecins dans la mise en œuvre des obligations prévues par la nouvelle réglementation sur la protection des données personnelles. Il propose une approche pragmatique et des fiches thématiques, qui peuvent être consultées distinctement.

Dautieu, T. (2018). "La Commission nationale de l'informatique et des libertés, régulateur des données de santé." Les Tribunes de la santé **58**(1): 47-52.

<https://www.cairn.info/revue-les-tribunes-de-la-sante-2018-1-page-47.htm>

La réflexion concernant la conciliation du développement des systèmes informatiques avec le respect des libertés individuelles et publiques trouve son origine au cours des années 70, dans une réaction à un projet gouvernemental visant à des interconnexions de différents fichiers administratifs à partir d'un identifiant unique, le numéro de sécurité sociale. Le comité institué à cet effet préconisa l'adoption d'une loi et l'instauration d'une autorité administrative chargée de veiller à sa correcte application. Ainsi naquirent la loi du 6 janvier 1978 fixant les principes que devront respecter « les traitements automatisés » et la Commission nationale de l'informatique et des libertés (CNIL), bénéficiant d'une « indépendance » grâce à l'apport du Sénat à l'occasion de la discussion parlementaire. Parmi les grands principes qui sont venus enrichir au fil des années la protection des données en droit français, figure l'interdiction de collecter ou de traiter, sauf exceptions, les données dites « sensibles » au nombre desquelles figurent les données de santé. À travers l'exercice des différents pouvoirs qui lui sont attribués par la loi, la CNIL joue donc un rôle important dans la régulation des données de santé au moment même où les défis, tant technologiques que juridiques, sont de plus en plus nombreux concernant le traitement de ces données.

Debiès, É. (2018). "Big data de santé et autodétermination informationnelle : quelle articulation possible pour une innovation protectrice des données personnelles ?" Revue française d'administration publique **167**(3): 565-574.

<https://www.cairn.info/revue-francaise-d-administration-publique-2018-3-page-565.htm>

Dans le domaine de la santé, l'explosion du big data met en lumière les dangers que peut représenter pour l'individu le regroupement des données et traces le concernant, pour constituer un profil toujours plus fin, qui peut aboutir à la stigmatisation et à la privation de droits. Mais quel choix est laissé à l'individu de partager ses données de santé ? Le consentement peut être un leurre, voire se révéler inopérant pour les traitements de données de santé de la sphère publique. Seule l'inclusion de l'individu informé dans le processus d'utilisation de ses données peut réellement donner sens à l'autodétermination informationnelle comme élément de responsabilisation des membres d'une communauté démocratique libre. On peut ainsi imaginer la construction d'une « protection sociale des données ».

Desgens-Pasanau, G. (2018). "La protection des données personnelles. Le RGPD et la nouvelle loi française", Paris : LexisNexis

Lanna, M. (2018). "Données publiques et protection des données personnelles : le cadre européen." Revue française d'administration publique **167**(3): 501-511.

<https://www.cairn.info/revue-francaise-d-administration-publique-2018-3-page-501.htm>

Les mouvements d'ouverture des données publiques et de protection des données personnelles sont aujourd'hui à une période charnière de leurs développements respectifs. Initiés dès la fin des années 1970 au niveau national, ils ont depuis chacun fait l'objet d'influences juridiques communautaires variées. Dès lors, le mouvement de transparence administrative, raison d'être initiale de l'open data, se trouve confronté aujourd'hui au renforcement des règles relatives à la protection des données personnelles. Si les deux domaines n'avaient à l'origine pas vocation à se croiser, les outils numériques aujourd'hui à disposition des administrations ont nécessité la prise en compte de telles problématiques.

Girard, M. et Polton, D. (2018). "La nouvelle réglementation sur l'accès aux données de santé et sa mise en œuvre deux ans après la loi de modernisation de notre système de santé." Journal du Droit de la Santé et de l'Assurance - Maladie (JDSAM) **20**(3): 25-29.

<https://www.cairn.info/revue-journal-du-droit-de-la-sante-et-de-l-assurance-maladie-2018-3-page-25.htm>

Lukacs, B. (2018). "La CNIL et l'analyse des projets de big data en santé". In: [Santé et intelligence artificielle]. Paris : CNRS éditions: 347-352.

Lukacs, B. (2018). "Les données : protection, conditions juridiques d'accès et de traitement". In: [Santé et intelligence artificielle]. Paris : CNRS éditions: 333-346.

Pellegrini, F. (2018). "La portabilité des données et des services." Revue française d'administration publique **167**(3): 513-523.

<https://www.cairn.info/revue-francaise-d-administration-publique-2018-3-page-513.htm>

Résumé La réduction de l'asymétrie entre usagers et responsables de traitements informatisés nécessite la création de droits adaptés. Tel est le cas du droit à la portabilité pour les services numériques. Ce droit est à la fois un droit économique, prolongeant le droit à l'interopérabilité, et un droit des personnes, participant à garantir leur autodétermination informationnelle. Ses deux sources en France sont, d'une part, le règlement général sur la protection des données, adopté au niveau européen, et, d'autre part, la loi Pour une République numérique. La mise en œuvre pratique de ce droit, tant par les acteurs privés que publics, est de nature à réduire les effets de rente et favoriser la création de services nouveaux.

Picart, E. (2018). "Le règlement général européen sur la protection des données personnelles et la relation médecin-patient." Médecine : De La Médecine Factuelle à Nos Pratiques **14**(10): 471-477.

Le médecin est au cœur de la récolte et du traitement des données de santé. De ce fait, le nouveau règlement général sur la protection des données personnelles lui est applicable. D'un côté, le médecin est soumis à des obligations concernant le traitement des données de santé, alors que de l'autre, le patient détient des droits sur celles-ci.

Simon, P., Douville, T. et Gruson, D. (2018). "Big data et santé : du discours aux applications pratiques. Dossier thématique." Journal de Droit de La Santé et de l' Assurance Maladie(20): 7-36.

A l'heure du développement de l'intelligence artificielle et de la prolifération d'outils permettant d'accompagner la relation médicale, ce dossier met en évidence plusieurs traits caractéristiques. Il souligne l'importance de la prolifération des données en santé, qu'il s'agisse de celles transmises aux instances de santé, aux professionnels de santé, aux patients eux-mêmes ou encore aux acteurs du secteur du bien-être. Pour autant, l'ouverture de ces données de santé s'accompagne inévitablement de nombreuses interrogations : la définition tout d'abord de la notion de donnée de santé, la question ensuite de sa finalité qui conditionne le régime juridique applicable. Celui-ci est-il suffisamment protecteur à l'heure où les GAFAM récupèrent les données de chacune et chacun d'entre nous ? La question du consentement de la personne au prélèvement et au traitement de ses données, est au cœur de la réflexion. Le consentement est l'élément pivot, permettant d'assurer la garantie du droit fondamental à la protection de la vie privée. Les enjeux des big data, au travers notamment des développements de l'intelligence artificielle, font l'objet de travaux dans le cadre de la révision des lois de bioéthique.

Rossi, J. et Bigot, J. É. (2018). "Traces numériques et recherche scientifique au prisme du droit des données personnelles." Les Enjeux de l'information et de la communication **19/2**(2): 161-177. <https://www.cairn.info/revue-les-enjeux-de-l-information-et-de-la-communication-2018-2-page-161.htm>

La disponibilité croissante de masses de traces numériques inspire une nouvelle génération de projets de recherche numériquement instrumentés. Or l'exploitation de ces traces, en tant que le droit les qualifie de données à caractère personnel, entraîne l'application d'un corpus de normes juridiques dont le Règlement général de protection des données récemment adopté. Celui-ci prévoit un cadre dérogatoire pour la recherche scientifique, soutenu par une coalition d'acteurs issus de la recherche médicale, qui s'applique également aux sciences humaines et sociales (SHS). Le présent article présente une synthèse de travaux de recherche sur ce cadre légal et la façon dont il vient interroger le rapport aux traces numériques dans la recherche en SHS.

Vuillet-Tavernier, S. (2018). "Protection des données personnelles et recherche : constat et perspectives d'évolution." *Statistique Et Société* 6(1): 31-35.  
<http://statistique-et-societe.fr/article/view/677>

La recherche moderne a souvent besoin d'utiliser des données personnelles dont la confidentialité est protégée par la loi. Cette contradiction peut être surmontée. La Cnil encourage les dispositifs qui permettent aux chercheurs de travailler tout en fournissant les garanties indispensables.

Bertrand, M. (2017). "La protection des données personnelles comme mode de régulation du big data en protection sociale complémentaire." *Revue Française Des Affaires Sociales*(4): 57-78.

Le big data constitue un défi pour le système français de protection sociale. Un changement de paradigme est à l'œuvre avec l'acquisition de données personnelles de plus en plus nombreuses, permettant de réduire l'aléa moral, entre l'assureur et l'assuré. D'une part, la conception du risque s'éloigne progressivement de la notion d'aléa, alors même que le caractère aléatoire est consubstantiel au contrat d'assurance. D'autre part, le modèle traditionnel de protection sociale consistant à atténuer le risque par l'indemnisation du sinistre s'accompagne désormais d'une nouvelle approche préventive. Or, ce changement de paradigme risque d'entraîner une segmentation excessive des risques, susceptible de porter atteinte à la protection des personnes. Si le règlement (UE) no 2016/679 et la loi no 2016-1321 pour une République numérique tendent à redonner aux personnes la maîtrise de l'usage de leurs données personnelles, le cadre normatif en vigueur laisse apparaître une tension entre diffusion et rétention des données. Ce constat impose la recherche d'un nouvel équilibre entre la protection des droits et des libertés des personnes physiques et les enjeux propres au secteur de la protection sociale, à la fois concurrentiel et réglementé.

Bossi Malafosse, J. (2016). "Les nouvelles règles d'accès aux bases médico-administratives : quel effet sur l'ouverture des données ?" *I2D - Information, données & documents* 53(3): 61-64.  
<https://www.cairn.info/revue-i2d-information-donnees-et-documents-2016-3-page-61.htm>

La révision des conditions d'accès aux données médico-administratives prévues par la loi du 26 janvier 2016 intervient dans un contexte marqué par un besoin accru d'accès aux données de santé, motivé en particulier par les nouvelles méthodes de recherche induites par les techniques du data mining.

Padiou, R. (2005). "Grandes bases de données et protection des personnes." *Journal de La Société Française de Statistique* 146(3): 89-93.

La mobilisation de grands fichiers de données administratives pour établir des statistiques, d'une part, et la possibilité d'enchaîner ou apparier de tels fichiers aussi bien que ceux des collectes statistiques pour constituer des panels, d'autre part, présentent un grand intérêt. Mais elles supposent une grande rigueur et l'observation de règles strictes pour protéger la vie privée des personnes concernées. En effet, ces possibilités qui ont connu un fort développement depuis un demi-siècle ont suscité de vives controverses. Il semble que nous soyons aujourd'hui parvenus à un stade de maturité. Les questions techniques, morales et juridiques ont été grandement clarifiées. Tout, certes, n'est pas résolu ; mais nous avons désormais les moyens de traiter ces problèmes de façon satisfaisante. Cet article revient sur la crise de croissance qui a été traversée, puis énonce les enjeux.

Ducimetiere, P., Brocas, A. M., Colladon, et al. (1997). "La confidentialité des données médicales." Echanges Sante - Social(86): 73-75.

## Quelques liens utiles

Site d'Etalab : <https://www.etalab.gouv.fr>

Documentation collaborative du SNDS : <https://documentation-snds.health-data-hub.fr/>

Présentation du SNDS sur le site de l'assurance maladie : <https://assurance-maladie.ameli.fr/etudes-et-donnees/presentation-snds/presentation-systeme-national-donnees-sante-snds>

Site du Health Data Hub : <https://www.health-data-hub.fr/>

Ecosystème du SNDS : <https://ecosysteme-snds.health-data-hub.fr/>

Thématique santé sur le site de la CNIL : <https://www.cnil.fr/fr/sante>