

Intérêt, utilisation et enjeux du SNDS

Vision pratique et juridique

>>> *Le développement de la puissance informatique, la multiplication des données, et les développements des réseaux à la fin des années 1990, ont accru de façon considérable les capacités d'analyse et donc de réutilisation des données pour des finalités autres que celles ayant présidé à leur collecte initiale.*



Ludovic Lamarsalle
Co-fondateur,
Heva



Jeanne Bossi Malafosse
avocate associée Données
personnelles et Sciences du
Vivant, DELSOL Avocats

En 2020, 20 milliards de sites ont été visités par Google chaque jour, 100 millions de Go ont été stockés sur les seuls serveurs de Google, et 6,9 milliards de requêtes ont été effectuées chaque seconde⁽¹⁾. Ainsi, nous apprécions chaque jour un peu plus le degré de réutilisation de nos données, par exemple lors de l'utilisation de gmail ou d'autres outils gratuits, où nous sommes parfois surpris des publicités affichées.

Le domaine de la santé n'échappe pas à ce phénomène, même si la spécificité de la donnée de santé a conduit le législateur à encadrer la réutilisation de ces données de santé. Des millions et des millions de données peuvent être déjà utilisées par les acteurs, mais ces informations sont souvent fragmentées, collectées dans le cadre du suivi médical (dossiers médicaux), d'un essai clinique, pour la constitution d'un registre ou encore pour la prise en charge par l'assurance maladie. Le développement en France de bases de données médico-administratives, d'abord caractérisées par la mise en place dans l'ensemble des établissements de soins du Programme de Médicalisation des Systèmes d'Information (PMSI), a été élargi au fil du temps par l'ajout d'autres bases de données aujourd'hui rassemblées dans le Système National des Données de Santé (SNDS) géré sous la responsabilité du Health Data Hub (HDH).

Le phénomène du Big Data et le recours aux techniques de l'Intelligence Artificielle (IA) conduisent bien sûr aujourd'hui à vouloir exploiter le maximum de données et sur ce point les bases de données du SNDS, d'abord conçues dans un but de suivi de dépenses, constituent une richesse

Les phénomènes du Big Data et de l'Intelligence Artificielle conduisent à exploiter le maximum de données.



importante pour la recherche dans le domaine de la santé.

Le système retenu qui permet, par le recours à un identifiant pseudonymisé construit à partir du numéro d'inscription au Répertoire national d'identification des personnes physiques (NIR ou numéro de sécurité sociale), de suivre les données d'un même patient d'une base à l'autre et donc de pouvoir apparier les données pour obtenir un suivi longitudinal est un atout majeur.

Il convient toutefois pour les acteurs intéressés d'être en mesure d'exploiter ces données sur le plan technique, et de comprendre les conditions d'accès.

Le présent article a pour objet de rappeler ce que contient le SNDS et quelles sont les règles d'accès aux données avant de rappeler les enjeux qui entourent l'exploitation de ces données.

1 Le SNDS, sa création et ses caractéristiques principales

Le système national des données de santé (SNDS) a été créé par l'article L. 1461-1 du code de la santé publique, issu de la loi n°2016-41 du 26 janvier 2016. Il est mis en œuvre sous la responsabilité de l'Etat en concertation avec les organismes responsables des systèmes d'information qui le composent.

Initialement, le SNDS rassemblait 5 bases de données :

- les données de l'assurance maladie, gérées par le système national d'information inter-régimes de l'Assurance Maladie (SNIIRAM) ;
- les données des établissements de santé, publics ou privés, gérées par le programme de Médicalisation des Systèmes d'Information (PMSI) ;
- les causes médicales de décès (CepiDC de l'INSERM) ;
- les données médico-sociales du système d'information des maisons départementales des personnes handicapées (MDPH-CNSA) ;
- un échantillon de données en provenance des organismes complémentaires.

Cette liste des bases de données dites «historiques» composant le SNDS a été élargie par la loi n°2019-774 du 24 juillet 2019 en visant dorénavant également :

- les données produites par les professionnels de santé à l'occasion des activités de soins et de prévention ;
- les données relatives à la perte d'autonomie, lorsque ces données sont appariées avec les bases précédentes ;
- les données à caractère personnel des enquêtes dans le domaine de la santé, lorsque ces données sont appariées avec des bases précédentes ;
- les données recueillies lors des visites médicales et de dépistage obligatoires ;
- les données recueillies par les services de protection maternelle et infantile ;
- les données de santé recueillies lors des visites d'information et de prévention, telles que définies à l'article L. 4624-1 du code du travail.¹

Cette volonté d'élargir les données du SNDS traduit ainsi la volonté des pouvoirs publics d'être en mesure de mettre à disposition des acteurs des données dites de «vie réelle» utiles à la recherche dans un contexte de Big Data, alors que les bases dites «historiques» ne sont disponibles qu'avec un certain décalage dans le temps.

Les données issues de ces nouvelles bases pourront ainsi venir nourrir le Health Data Hub (HDH) sous la forme de «catalogues» de données.

La loi du 24 juillet 2019 a ainsi créé la plateforme des données de santé ou Health Data Hub sous la forme d'un groupement d'intérêt public qui gère la mise à disposition

des données du SNDS. Dès lors aujourd'hui, les termes de plateforme des données de santé ou de HDH ou de SNDS désignent le même système d'information comportant les bases dites historiques et les catalogues de données provenant notamment des nouveaux types de données ajoutés par la loi de 2019.

La publication prochaine attendue en janvier 2021 du décret venant modifier le décret n° 2016-1871 du 26 décembre 2016 relatif au traitement du SNDS actera le remplacement du SNDS par le Health Data Hub et définira de façon plus claire et pérenne le fonctionnement du HDH.

Présentation de quelques bases de données du SNDS

Le SNIIRAM

Le Système National d'Informations Interrégimes de l'Assurance Maladie est une gigantesque base d'informations qui contient les données individuelles pseudonymisées utilisées pour la facturation et le remboursement de la consommation de soins effectués en ville et en établissements de santé privés [DCIR (Données de Consommation Interrégimes)], chaînées à la base de données d'hospitalisation [PMSI (Programme de Médicalisation des Systèmes d'Information)] au moyen d'un identifiant unique anonyme attribué à chaque individu, le numéro de sécurité sociale [NIR (Numéro d'inscription au Répertoire)].

Il contient ainsi les données individualisées et anonymes de l'ensemble des demandes de remboursements des dépenses de santé de plus de 99% des résidents du territoire français, soit près de 65 millions de personnes⁽²⁻⁵⁾.

Le DCIR contient les données individuelles de remboursements des patients de tous les soins effectués en ville et en établissements de santé privés à l'échelle nationale. Ces informations enregistrées sont à la fois administratives et médicales.

Les données administratives sont principalement :

- les données sociodémographiques : date de naissance, sexe, couverture maladie universelle complémentaire (CMU-c), département et commune de résidence, indice de défaveur sociale de la commune de résidence ainsi que date de décès ;
- les informations sur l'organisme d'affiliation du bénéficiaire (Caisse Nationale d'Assurance Maladie (CNAM), Mutualité Sociale Agricole (MSA), Sections Locales Mutualistes (SLM)) ;
- les congés maladie et les pensions d'invalidité ;
- les consultations (nombre de consultations et date de soins) auprès de médecins généralistes ou spécialistes du secteur libéral, ainsi que tous les soins ambulatoires exécutés par un professionnel de santé (infirmiers, kinésithérapeute...) ;

- les actes médicaux (nombre d'actes et date de soins) effectués en mode libéral (utilisation du code de la classification commune des actes médicaux (CCAM)) ;
- la réalisation d'un acte biologique (nombre d'actes et date de soins) hors du secteur hospitalier (utilisation du code nomenclature des actes de biologie médicale (NABM)) ;
- la délivrance de médicaments (nombre de boîtes et date de soins) à partir de la classification Anatomique, Thérapeutique et Chimique (ATC) et du Code Inter-Pharmaceutique (CIP) pour les médicaments délivrés en ville et du code de l'Unité Commune de Dispensation (UCD) pour les médicaments en rétrocession hospitalière ;
- la délivrance des dispositifs médicaux (nombre de dispositifs et date de soins) à partir du code de la Liste des produits et des prestations (LPP) ;
- les remboursements de transports.

Même si la base ne renseigne pas directement sur le motif et le diagnostic de chaque remboursement de soin, les informations médicales du DCIR incluent les pathologies chroniques donnant lieu à une exonération du ticket modérateur pour l'assuré, appartenant à la liste des 30 Affections de Longue Durée (ALD) et les ALD hors liste, dont le motif d'exonération (diagnostic) est codé en utilisant la classification internationale des maladies 10^{ème} édition.

Les données hospitalières du PMSI

Depuis 2006, les données d'hospitalisation du PMSI des champs Médecine-Chirurgie-Obstétrique (MCO), Hospitalisation à Domicile (HAD), Soins de Suite et Réadaptation (SSR) et Psychiatrie (PSY) collectées par l'Agence Technique de l'Information sur l'Hospitalisation (ATIH) sont chaînées et intégrées en routine à celles du SNIIRAM. Depuis 2007, le taux d'appariement entre les données de ville et les données d'hospitalisations sont de plus de 97% pour le régime général. Les PMSI renseignent sur les séjours hospitaliers (établissements privés ou publics) des patients et apportent des informations médicales supplémentaires :

Pour le champ MCO : le motif d'hospitalisation (il est renseigné par le diagnostic principal (DP) ou le diagnostic relié (DR) du séjour, qui sont codés à partir de la CIM-10), les actes médicaux (codés en CCAM) réalisés au cours du séjour, les comorbidités ayant conduit à des soins au cours du séjour (diagnostics associés significatifs, DAS) ou autres DP issus des résumés d'unités médicales (si plus d'une au cours du séjour), les médicaments et dispositifs implantables coûteux et d'autres renseignements médicaux comme par exemple : prélèvements d'organes, transfusions, greffes, séjours en unité de soins intensifs ou en réanimation.

Pour les autres champs du PMSI (HAD, PSY, SSR), les motifs d'hospitalisations peuvent être complétés le cas échéant par des classifications propres.

Les causes médicales de décès

Les données individuelles sur les causes médicales de décès survenant en France (métropole et DOM) sont gérées par le Centre d'épidémiologie sur les causes de décès (CépiDc) de l'INSERM en collaboration avec l'Insee. Les causes médicales de décès sont appariées aux données du SNIIRAM en utilisant une méthode d'appariement indirect. À ce jour, les causes médicales de décès sont appariées aux données du SNIIRAM pour 2013, 2014 et 2015. Pour 2015, 90,1% des décès de CépiDc ont été appariés (535 231 sur un total de 594 192 décès). Les causes de décès (cause initiale et toutes les causes) sont codées conformément à la classification CIM-10.

Les finalités pour lesquelles les données du HDH peuvent être mises à disposition sont énumérées à l'article L.1461-1, III du code de la santé publique et doivent présenter un caractère d'intérêt public, pour contribuer :

- à l'information sur la santé ainsi que sur l'offre de soins, la prise en charge médico-sociale et leur qualité ;
- à la définition, à la mise en œuvre et à l'évaluation des politiques de santé et de protection sociale ;
- à la connaissance des dépenses de santé, des dépenses d'assurance maladie et des dépenses médico-sociales ;
- à l'information des professionnels, des structures et des établissements de santé ou médico-sociaux sur leur activité ;
- à la surveillance, à la veille et à la sécurité sanitaires ;
- à la recherche, aux études, à l'évaluation et à l'innovation dans les domaines de la santé et de la prise en charge médico-sociale.

La notion d'intérêt public est entendue largement puisqu'elle inclut la finalité de « la garantie de normes élevées de qualité et de sécurité des soins de santé et des médicaments ou des dispositifs médicaux » visée à l'article 66 de la loi Informatique et Libertés modifiée.

Certaines garanties sont également posées par le législateur sur l'utilisation des données du HDH :

- aucune décision ne peut être prise à l'encontre d'une personne physique identifiée sur le fondement des données la concernant et figurant dans l'un de ces traitements ;
- le responsable du traitement, ainsi que les personnes le mettant en œuvre ou autorisées à accéder aux données qui en sont issues, sont soumis au secret professionnel ;
- l'accès aux données s'effectue dans des conditions assurant la confidentialité et l'intégrité des données et la traçabilité des accès et des autres traitements, conformément à un référentiel défini par arrêté des ministres chargés de la santé, de la sécurité sociale et du numérique, pris après avis de la Commission nationale de l'informatique et des libertés ;
- les données sont conservées pour une durée maximale de 20 ans, elles sont ensuite archivées pendant 10 ans.

Les données à caractère personnel, présentes dans le système national des données de santé, sont rattachées à chaque personne concernée par un pseudonyme. Elles ne sont donc pas directement nominatives.

2 Les conditions d'accès au SNDS

Une procédure spécifique est prévue pour pouvoir accéder aux données du HDH. Deux types d'accès sont ainsi reconnus :

- un accès permanent au bénéfice de certains services publics ou organismes publics pour l'accomplissement de leurs missions (exemples : DRESS, ARS, HAS, ANSM, INCA, HDH, etc.) ;
- un accès ponctuel, soumis à autorisation de la CNIL, pour les autres organismes (structures privées par exemple) ou les organismes publics hors du cadre fixé, dans un but de recherche répondant à un motif d'intérêt public. Sous réserve de démontrer que l'accès n'a pas pour finalité la promotion de produits ou l'exclusion de garanties des contrats d'assurance.

Une demande d'autorisation doit être déposée à la CNIL conformément aux dispositions de la section 3 du Chapitre III du Titre II de la loi Informatique et Libertés modifiée qui comportent les dispositions relatives aux traitements de données à caractère personnel dans le domaine de la santé.

Dans le cadre de ces demandes d'accès, le HDH est notamment chargé :

- de réunir, organiser et mettre à disposition les données du SNDS et de promouvoir l'innovation dans l'utilisation des données de santé ;
- d'informer les patients, de promouvoir et de faciliter leurs droits, en particulier concernant les droits d'opposition ;
- d'assurer le secrétariat unique ;
- d'assurer le secrétariat du CESREES ;
- de contribuer à l'élaboration, par la CNIL, de référentiels et de méthodologies de référence. Il facilite la mise à disposition de jeux de données de santé présentant un faible risque d'impact sur la vie privée ;
- de procéder, pour le compte d'un tiers et à la demande de ce dernier, à des opérations nécessaires à la réalisation d'un traitement de données issues du SNDS pour lequel ce tiers a obtenu une autorisation ;
- de contribuer à diffuser les normes de standardisation pour l'échange et l'exploitation des données de santé, en tenant compte des standards européens et internationaux ;
- d'accompagner, notamment financièrement, les porteurs de projets sélectionnés dans le cadre d'appels à projets lancés à son initiative et les producteurs de données associés aux projets retenus.

La procédure d'accès est plus stricte pour les organismes du secteur privé puisqu'ils doivent en particulier recourir à un laboratoire de recherche ou à un bureau d'études, publics ou privés, pour réaliser le traitement qui eux-mêmes doivent présenter à la Commission nationale de l'Informatique et des Libertés un engagement de conformité à un référentiel incluant les critères de confidentialité, d'expertise et d'indépendance, arrêté par le ministre

chargé de la santé, pris après avis de la CNIL (arrêté du 17 juillet 2017).

La plateforme a également élaboré des procédures d'accès simplifiées aux données mises à disposition par le SNDS.

L'objectif pour les acteurs est tout à la fois d'accéder aux données de santé du système national des données de santé (SNDS) mis à disposition par le Health Data Hub² (HDH) et d'apparier le cas échéant leurs données à celles du SNDS et de pouvoir pour certains, dans un second temps, constituer un système fils du SNDS afin de pouvoir mettre à disposition eux-mêmes des jeux de données également nourris des bases de données du SNDS.

Il est à noter que toute sortie de données des bases constitutives du SNDS mises à disposition par le HDH s'apparente à la constitution d'un système fils.

Il est à cet égard nécessaire de respecter l'homologation au référentiel SNDS approuvé par arrêté du 22 mars 2017.

L'homologation au regard du référentiel SNDS est une procédure auto-réalisée par le responsable de traitement. Il est à noter que l'arrêté approuvant le référentiel SNDS du 22 mars 2017 sera également modifié au cours de l'année 2021.

Pour rappel, le référentiel rappelle les prérequis nécessaires avant la mise en œuvre du système :

- la réalisation d'une analyse de risques ;
- la réalisation d'une analyse d'impact relative à la protection des données personnelles (AIPD) ;
- la mise en œuvre des mesures de couverture des risques associées ;
- la réalisation d'étapes de recette et de tests pour s'assurer de la bonne couverture des risques ;
- la réalisation d'une homologation de sécurité sur le périmètre considéré ;
- le suivi opérationnel de la sécurité du système d'information.

Enfin, les données non anonymes du SNDS ne peuvent être hébergées que sur des systèmes homologués vis-à-vis du présent référentiel.

NOTES

1. L'article R. 1461-4 du code de la santé publique détaille les catégories de données personnelles réunies au sein du SNDS.
2. Les bases dites « historiques » et la collection de bases composant le « catalogue », conformément aux termes de la Charte Responsables de Données publiée par le HDH.

La prise en compte du contexte de la crise sanitaire liée au Covid 19

Afin de disposer de données sur l'évolution de l'épidémie, le périmètre d'accès aux données de santé a été temporairement élargi. En effet, des fichiers de données ont été spécifiquement créés afin de rassembler les données issues de la crise sanitaire. Ces fichiers nommés Contact COVID et SI-DEP ont respectivement pour but de recenser les informations sur les cas contact et les chaînes de contamination, tandis que le second centralise les résultats des tests au SARS-CoV-2 réalisés par les laboratoires publics ou privés.

L'accès aux données recueillies par le fichier contact COVID a été inclus dans le SNDS par le décret n° 2020-551 du 12 mai 2020, article 3, VI, 3°, tandis que le système SI-DEP n'a pas été intégré dans la base de données. Cependant, il convient de souligner que l'accès à ces données reste temporaire.

En plus des données du fichier contact COVID, le HDH donne également accès temporairement à de nouvelles données de santé (données de pharmacie, aide médicale d'urgence, EHPAD, résultats d'examen biologiques, d'imagerie...)

Afin de veiller au respect de la protection de ces données, la CNIL élabore et rend public son avis trimestriel sur le sujet, qu'elle adresse au Parlement. Elle a rendu son premier avis par une délibération en date du 10 septembre 2020, pour la période de mai à août 2020. Le niveau de protection des données pour le fichier SI-DEP a été jugé comme étant globalement satisfaisant. Toutefois, quelques améliorations peuvent être apportées concernant la gestion des comptes d'administration et la traçabilité des accès. Concernant le fichier Contact COVID, la CNIL a relevé des insuffisances, notamment concernant l'information des personnes ou encore l'exercice des droits des personnes. La deuxième vague de contrôle a débuté fin septembre. Ainsi, le prochain avis trimestriel devrait être publié courant janvier 2021.

3 Les enjeux du Health Data Hub

Ces enjeux sont de natures très diverses et l'objectif n'est pas ici de les énumérer de façon limitative, mais d'attirer l'attention sur quelques sujets qui sont susceptibles de les illustrer.

La question de l'hébergement des bases du HDH

La plateforme a signé avec une filiale irlandaise de la société américaine Microsoft un contrat d'hébergement des données de santé en date du 15 avril 2020. Cette décision a fait l'objet de nombreuses critiques, mais également de recours devant le Conseil d'Etat, notamment concernant le choix de l'hébergeur, Microsoft, qui aurait été effectué sans appel d'offre. A ce jour, deux recours ont déjà été déposés à l'encontre de la plateforme.

Le premier recours a été déposé notamment par l'association le Conseil national du logiciel libre (dite CNLL). Dans son

ordonnance du 19 juin 2020, le Conseil d'Etat a rejeté le recours pour l'essentiel, mais a ordonné à la plateforme des données de santé de fournir à la Commission tous éléments relatifs aux procédés de pseudonymisation utilisés, afin que la CNIL puisse vérifier si ces techniques assurent une protection suffisante des données en question, en empêchant, dans la mesure du possible, toute réidentification des personnes physiques concernées.

Un second recours a été déposé par les mêmes associations, dont le CNLL. Ce recours est intervenu à la suite de l'arrêt Schrems II rendu par la Cour de justice de l'Union européenne (CJUE) le 16 juillet 2020, invalidant le Privacy Shield, la décision d'adéquation de la Commission européenne relatif au niveau de garantie de protection des données aux Etats-Unis. En effet, le HDH est hébergé par la société Microsoft, société de droit américain, hébergeant les données de santé aux Pays-Bas. Dans son ordonnance en date du 13 octobre 2020, le Conseil d'Etat estime que le risque de communication de ces données aux services de renseignement américains ne peut être exclu, même si un arrêté ministériel a été pris le 9 octobre 2020 et interdit, tout transfert de données à caractère personnel dans le cadre de ce contrat. Le Conseil d'Etat juge cependant, que ce risque ne justifie pas pour autant la suspension de la plateforme en urgence. Le Conseil d'Etat décide donc d'imposer de prendre des précautions particulières, sous le contrôle de la CNIL.

Ces mesures particulières ont pour objectif de renforcer certains points du contrat entre le HDH et Microsoft et de rechercher des mesures additionnelles pour mieux protéger les données qu'il héberge. Cependant, ces mesures additionnelles sont temporaires. En effet, le secrétaire d'Etat au numérique, Cédric O, a annoncé qu'une solution pérenne était recherchée pour l'hébergement du HDH sur une plateforme française ou européenne.

L'utilisation du SNIIRAM : un exemple des enjeux d'utilisation de ces données

Cette masse d'informations, avec un suivi inégalé de 2006 à 2019, permet la réalisation d'études épidémiologiques, économiques, de pharmacovigilance et bien d'autres types d'études.

Pour mieux comprendre l'intérêt d'une telle base, une simple recherche sur pubmed avec comme mot clé SNIIRAM, renvoie à 61 études publiées sur les 3 dernières années dans des revues à comité de lecture. Il s'agit principalement d'études épidémiologiques (incidence, prévalence, mortalité), économiques et de morbidité.

Rappelons-nous que les études sur le Mediator[®] (6), sur la Dépakine[®] (7) n'auraient pas pu être menées sans cette base de données. Malgré l'absence des résultats des examens médicaux, les informations médicales reliées ensemble apportent suffisamment de vraisemblance pour pallier

certaines faiblesses de la base. Par exemple, une hypertension pourra être détectée par des traitements médicamenteux délivrés au long cours, comme le stade d'un cancer pourra être apprécié à la fois par les médicaments, mais aussi par la prise en charge préalable proposée au patient (par exemple chirurgie ou radiothérapie).

La possibilité de choisir la date d'inclusion, le suivi rétrospectif et prospectif avec la quasi-intégralité des patients traités en France apporte un regard nouveau avec lequel les cliniciens vont devoir compter dans les prochaines années et qui pourrait les amener à évoluer sur leurs connaissances des données épidémiologiques de beaucoup de pathologies, pour lesquelles parfois peu d'études existent ou pour lesquelles les extrapolations à la population générale qui avaient été faites doivent être confirmées.

Alors bien sûr des précautions d'utilisations doivent être prises dans l'exploitation de ces données, une connaissance fine des types de données est indispensable, par exemple les remontées d'informations au cours du temps ne sont pas identiques au sein des régimes d'affiliation, ainsi les régimes autres que celui de la CNAM ne diffusaient pas l'information au sein du SNIIRAM de manière homogène avant 2014. Seul le régime général (RG) de l'Assurance Maladie le faisait, et rappelons-le, il couvre environ 75% de la population métropolitaine.

L'enjeu n'est pas ici de rappeler les incertitudes qu'engendrent les résultats statistiques extrapolés à partir d'études basées sur quelques dizaines ou centaines de patients, mais tout le monde comprendra l'intérêt de compléter ces études qui, de par leurs données massives, produisent souvent des intervalles de confiance resserrés.

Mais au-delà de ces études classiques, de nouveaux enjeux de connaissance se font jour, la compréhension des lignes de traitement dans la cancérologie, l'association des traitements médicamenteux délivrés, l'impact territorial d'une prise en charge d'une maladie, et surtout la connaissance du risque en lien avec un profil de patient...

Les enjeux à venir

Ces enjeux s'amplifieront sans aucun doute dans les années à venir, en lien avec plusieurs nécessaires évolutions vraisemblables du SNDS.

1. La réforme de l'accès aux données, qui pour accélérer la recherche, devra s'orienter d'un régime d'autorisation a priori vers un système plus souple de contrôle a posteriori. A noter sur ce point que les méthodologies de référence adoptées par la CNIL permettent déjà de simplifier certains accès aux bases de données du HDH.

2. Un accès payant aux données qui permettra aux acteurs institutionnels (CNAM) et Health Data Hub de fournir un service plus rapide qu'aujourd'hui aux acteurs de la recherche.

3. Une automatisation du traitement des données. Il est indispensable d'automatiser la gestion des données afin de pouvoir travailler plus vite ces informations, pour les analyser plus facilement.

4. Le recours quasi systématique aux outils de l'intelligence artificielle, afin de pouvoir déterminer des risques par profil de patient et par maladie.

5. Le développement de nouveaux modèles statistiques.

La réutilisation des données de santé a ouvert une nouvelle ère pour les études de santé publique. Le développement de ces données va encore s'accélérer dans les années à venir, avec pour enjeu de lier toutes ces informations entre elles, et ainsi créer des bases de données de plus en plus fines et médicalisées (les données génomiques par exemple). Le législateur devra donc innover pour à la fois protéger les données de nos concitoyens, mais aussi développer l'utilisation de ces données afin d'affronter les enjeux cruciaux en lien avec l'essor des modèles de machine ou deep learning basés sur l'intelligence artificielle. Ces enjeux seront potentialisés par l'arrivée de nouveaux réseaux comme la 5G, qui va poser la question de l'utilisation de ces modèles en temps réel pour détecter la probabilité d'un risque lors de la prise en charge de la maladie d'un patient.

La confiance restera également la clef de réussite de ce modèle très français de mise à disposition des données qui doit à la fois assurer la sécurité et la confidentialité de nos données individuelles pour permettre leur mise à disposition afin de contribuer à améliorer la santé publique, notre bien commun.

RÉFÉRENCES

1. <https://www.blogdumoderateur.com/chiffres-google/>. Accès 28/12/2020
2. Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, de Roquefeuil L, et al. Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Rev Epidemiol Sante Publique*. oct 2017;65 Suppl 4:S149-67.
3. Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P, et al. The national healthcare system claims databases in France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology. *Pharmacoepidemiol Drug Saf*. août 2017;26(8):954-62.
4. Moulis G, Lapeyre-Mestre M, Palmaro A, Pugnet G, Montastruc J-L, Sailler L. French health insurance databases: What interest for medical research? *Rev Med Interne*. juin 2015;36(6):411-7.
5. Tuppin P, de Roquefeuil L, Weill A, Ricordeau P, Merlière Y. French national health insurance information system and the permanent beneficiaries sample. *Rev Epidemiol Sante Publique*. août 2010;58(4):286-90.
6. https://fr.wikipedia.org/wiki/Affaire_du_Mediator. Accès 28/12/2020
7. Exposition à l'acide valproïque et ses dérivés au cours de la grossesse en France de 2007 à 2014 : une étude observationnelle sur les données du SNIIRAM. <https://solidarites-sante.gouv.fr/ministere/documentation-et-publications-officielles/rapports/sante/article/rapport-d-etude-exposition-a-l-acide-valproique-et-ses-derives-au-cours-de-la>. Accès 28/12/2020