La Gouvernance des Données Médicales : de la Protection Individuelle à une Ressource Collective en Open Data

Cécilia Darnault

LA GOUVERNANCE DES DONNEES MEDICALES : DE LA PROTECTION INDIVIDUELLE A UNE RESSOURCE COLLECTIVE EN *OPEN DATA*.

Darnault Cécilia,
Docteure en droit,
cecilia.darnault@gmail.com

Résumé : La gouvernance des données médicales représente un enjeu majeur de l'évolution des nouvelles technologies dans le domaine médical dans la mesure où elles constituent le matériau d'apprentissage des modèles algorithmiques. En tant qu'éléments personnels sensibles, les données de santé sont gouvernées à l'aune de la protection des droits fondamentaux individuels pour éviter les usages abusifs potentiels. Néanmoins, l'utilité sociale de ces données de santé invite à envisager leur gouvernance à travers une doctrine collectiviste, les considérant en tant que ressources communes en *open data*, plus favorable pour la recherche, les statistiques et l'amélioration de la qualité des soins. Un tel système reposant sur une base de données de santé en *open data*, où les patients deviennent des *donneurs de données* anonymisées repose tout de même sur des moyens qui ne sont pas sans risques envers les droits fondamentaux.

Mots-clefs : Technologies ; Données de santé ; Gouvernance ; Risques juridiques ; droits fondamentaux.

INTRODUCTION

Selon le rapport *Statista Digital Economy Compass*[1], trente-trois zettaoctets de données numériques ont été créées dans le monde en 2019. Un chiffre impressionnant que chacun d'entre nous alimente quotidiennement, et le domaine de la santé n'y fait pas exception. En effet, « le secteur des soins de santé est submergé de données, dont l'étendue et le volume augmentent de façon exponentielle »[2], un flux de données qui nourrit les algorithmes d'intelligence artificielle et permet le développement de nouvelles technologies dans le secteur médical. En oncologie par exemple, créé par une équipe du *Massachussetts Institute of Technology* (MIT, États-Unis), un algorithme est capable d'identifier sur des radiographies de tissu mammaires en apparence sain la zone précise où se développerait un cancer quatre an plus tard. Un résultat qui n'a pu s'obtenir qu'en fournissant les données nécessaires à l'apprentissage de l'algorithme, les chercheurs ayant nourri « un système de deep learning de 72 000 mammographies en les associant aux données cliniques évaluant le risque de cancer du sein (alimentation, génétique, hormones, poids, grossesses, allaitement, ...) de 30 000 patientes »[3]. Des informations personnelles relatives à la santé des patients qui constituent une base de données essentielle au développement, à l'apprentissage et au succès du fonctionnement des algorithmes d'intelligence artificielle en matière de santé. Néanmoins, il advient que l'accessibilité et l'utilisation de ces données fait encore débat, souvent par crainte d'usages abusifs. A ce titre, la gouvernance des données médicales représente un enjeu sociétal majeur, tributaire des choix opérés par les acteurs des politiques publiques, qui va modeler l'évolution de la recherche et des nouvelles technologies médicales. En effet, l'orientation du système de gouvernance des données médicales que les régulateurs juridiques choisissent d'appliquer revêt une importance capitale dans le sens où il impacte directement les possibilités offertes par l'exploitation des données de santé. Constituée d'une grande diversité (les dossiers médicaux et hospitaliers, les demandes de remboursement, les enquêtes, les biobanques, les rapports de laboratoire, les transactions en pharmacie, les recherches et les appareils ou applications de surveillance, les éléments d'imagerie, etc), les données de santé pourraient permettre d'améliorer la qualité des soins pour les patients et d'appuyer directement le développement de

---

[1] T. Gaudiaut, « *La totalité des données créées dans le monde* », Statista Digital Economy Compass 2019 in Statista.com, 24 avril 2019, https://fr.statista.com/infographie/17793/quantite-de-donnees-numeriques-creees-dans-le-monde/ (consulté le 1er avril 2020).

[2] J. Oderkirk, E. Ronchi, « *La gouvernance des données au service de la santé et des soins* », L'observateur OCDE, n°309, T1, 2017.

[3] H. Jalinière, « *Des cancers bientôt révélés par l'imagerie intelligente ?* » in L'intelligence artificielle en 50 questions, Sciences et Avenir, Hors-série, n°199, Octobre-Novembre 2019, p. 26.

la recherche médicale (notamment la découverte de nouveaux traitement, l'amélioration des diagnostics, la progression de la médecine personnalisée, etc). Malgré les possibilités qu'elles offrent, ces données restent encore majoritairement inutilisées en raison d'une gouvernance globale axée sur la protection des droits fondamentaux individuels des patients considérant les données à caractère personnel comme étant propres aux personnes qu'elles concernent. Cependant, le caractère d'intérêt général que revêt le domaine médical et l'utilité sociale de ces informations, invite à interroger ce modèle de gouvernance individualiste et questionner les principes juridiques sollicités pour envisager le modèle différemment.

Cet article propose d'élaborer *une analyse de notre système de gouvernance des données personnelles pour l'explorer sous de nouvelles perspectives*. Non plus à travers un regard tourné vers l'individu porté originellement aux données de santé mais en tant que ressource collective au service de l'intérêt général sanitaire et de l'amélioration des soins, en étudiant le cas du régime juridique français qui entremêle normes européennes et spécificités nationales. L'appréhension des données médicales, considérées comme des informations personnelles particulièrement sensibles, a conduit à l'application de principes juridiques privatistes centrés sur les droits individuels. Une vision qui tend à évoluer au prisme d'une doctrine dissidente, s'appuyant davantage sur des notions publicistes telle que la primauté de l'intérêt général que ces données peuvent représenter et prône une gouvernance juridique plus collective où les données médicales deviennent une ressource commune à disposition de tous. A la croisée des dissidences entre les paradigmes privatiste et public, il est possible de concevoir une gouvernance publique des données médicales en tant que ressource commune tout en garantissant le respect des droits fondamentaux individuels (**I**). La problématique ne s'arrête pas à un simple débat juridique sur la catégorisation des données de santé, en des termes privatiste ou publiciste, ou sur leur régime actuel, mais s'intéresse aux évolutions en cours qui s'intègrent dans un véritable projet d'élaboration d'un système de données de santé en *open data*, tant au niveau national qu'européen. Une base commune de données médicales ouverte à tous qui, malgré un développement national progressif et pédagogique afin de se développer dans les meilleures conditions, soulève tout de même quelques interrogations et exposent les droits fondamentaux des personnes concernées à de nombreux risques non négligeables (**II**).

I. UN MODÈLE HYBRIDE DE GOUVERNANCE JURIDIQUE DES DONNEES MEDICALES

Les intentions règlementaires initiales convergent vers une idéologie privatiste consistant à analyser les données à caractère personnel comme propres aux personnes qu'elles concernent, et donc à gouverner leur traitement par exception dans des conditions spécifiques. Cependant, la perception privée de ces données s'altère lorsque l'on prend en considération des paramètres d'intérêt général que recouvrent certaines de ces informations, notamment les données de santé, qui érigées en ressources communes peuvent constituer des opportunités non négligeables. Une gouvernance collectiviste qui n'est envisageable qu'à travers le respect d'un régime juridique adapté qui permette la nécessaire conciliation, à la fois des droits personnels fondamentaux et des objectifs communs d'intérêt général public en matière de santé.

A. La spécificité du modèle de gouvernance des données médicales

Le modèle de gouvernance juridique des données à caractère personnel est construit sur la base d'une idéologie privatiste considérant les données comme propre à l'individu, en tant qu'éléments composant sa vie privée à protéger d'un traitement extérieur. Un regard qui évolue lorsque l'on conçoit les données de santé, particulièrement sensibles, sous l'angle de l'intérêt général, non plus comme une information individuelle relevant de la vie privée mais en tant qu'ensemble public agrégé qui promet des opportunités pour l'avenir de la médecine.

1) *Les fondements du modèle de gouvernance des données : une logique d'élément privé*

Les données numériques. S'agissant plus précisément des *données à caractère personnel*, elles sont définies légalement comme toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres[4]. « Nouvel or noir de l'internet et nouvelle monnaie du monde digital »[5], elles représentent la genèse de la valorisation économique de l'information à l'ère de la révolution numérique. Les agents économiques ont vite compris l'intérêt de ces données et en ont constitué une matière première exploitable. La

[4] Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, JO du 7 janvier 1978 (article 2) ; Règlement (UE) n°2016/679 du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (règlement général sur la protection des données), JOUE L127 2 du 23/05/2018 (article 4.1).

[5] V. M. Kuneva, Commissaire Européen à la Consommation, Keynote Speech, Roundtable on Online Data Collection, Targeting and Profiling (31 mars 2009), citée dans Personal Data : the emergence of a new asset class, World Economic Forum, jan. 2011, p.5.

collecte et le traitement de cette masse d'informations en accroissement constant sont une source de création de valeur constituant une réelle opportunité économique[6] au cœur d'une *économie de la donnée*[7]. Ainsi, en tant que matière première d'un marché économique en pleine expansion, sous le couvert d'une influence juridique privatiste, la donnée numérique a rapidement été exploitée en tant que bien marchand. Un bien échangeable sur le marché, et cela avant que les individus réalisent la nature, l'existence même des données qu'ils produisent, leur valeur et l'exploitation, qu'elle soit économique, publique ou sociale, qui en résultent. Le fait que le marché se soit ainsi saisi de la question, du traitement et de l'exploitation des données numériques, et qu'elles proviennent des utilisateurs, en tant qu'élément attaché à l'individu qui les génère par nature, les a profondément ancré dans la sphère privée. Une mercantilisation d'un élément personnel qui n'a pas manqué d'engendrer des difficultés. La succession des révélations largement médiatisées relatives à l'exploitation abusive des données personnelles a permis de sensibiliser les individus et suscite un intérêt grandissant au sein du grand public pour le droit des données personnelles. Parmi ces scandales, comment ne pas mentionner l'onde de choc produite par les dénonciations d'Edward Snowden en 2013[8] concernant les programmes de surveillance américains, mais plus récemment et dans ce contexte, impossible de passer à côté de Facebook et son dirigeant, notamment suite aux déboires suscités lors de l'affaire dite *Cambridge Analytica* au cours de laquelle l'entreprise est accusée d'avoir utilisé des données de 30 millions à 70 millions d'utilisateurs de Facebook, recueillies sans leur consentement, ensuite exploitées aux fins de démarchage politique ciblé dans le cadre de la campagne électorale américaine remportée par Donald Trump[9]. Deux affaires particulièrement connues tirées du panier qui ne doivent pas faire oublier les nombreuses autres atteintes de ces dernières années, celles qui sont encore à découvrir, et celles qui viendront. Des nombreux lanceurs d'alerte qui se succèdent aux différentes condamnations qui ont été prononcées, ce sont des faits d'armes qui marquent particulièrement les esprits aujourd'hui mais dont le combat à commencer il y a quelques dizaines d'années maintenant, à l'aune d'un modèle de gouvernance prônant la protection des droits individuels des personnes numérisées.

---

[6] J-M Deltorn, « *La protection des données personnelles face aux algorithmes prédictifs* », RDLF 2017, chron. n°12 (www.revuedlf.com).

[7] Commission Européenne, Vers une économie de la donnée prospère, Bruxelles, 2 juillet 2014, COM(2014) 442 final, http://ec.europa.eu/transparency/regdoc/rep/1/2014/FR/1-2014-442-FR-F1-1.Pdf .

[8] V. not. le documentaire de L. Poitras, Citizenfour, 2015.

[9] A. William, « *Ce qu'il faut savoir sur Cambridge Analytica, la société au cœur du scandale Facebook* », in Le Monde, 22 mars 2018, https://www.lemonde.fr/pixels/article/2018/03/22/ce-qu-il-faut-savoir-sur-cambridge-analytica-la-societe-au-c-ur-du-scandale-facebook_5274804_4408996.html (consulté le 1er avril 2020).

Pour reprendre l'économiste J. Stiglitz, nous sommes une communauté *«* et comme toutes les communautés, il nous faut respecter des règles pour pouvoir vivre ensemble », des règles qui se doivent d'être « justes et équitables, et cela doit se voir clairement (et) accorder toute l'attention nécessaire aux pauvres comme aux puissants, et témoigner d'un sens profond de l'honnêteté et de la justice sociale »[10]. En tant que source de renseignement, que ce soit sur l'identité, le comportement, les habitudes ou les préférences, l'exploitation des données entre immanquablement en conflit avec certains droits des personnes concernées. Pour prévenir les mauvais usages, ou ceux qui pourraient être mal intentionné, des données personnelles dans le cadre du développement des nouvelles technologies, et éviter qu'un algorithme ne se transforme en un outil discriminatoire ou d'atteinte aux droits fondamentaux des personnes, une intervention règlementaire était plus que nécessaire. Laquelle s'est manifestée par l'adoption d'un modèle de gouvernance dont l'objectif était de mettre en œuvre un ensemble de processus, rôles, règles, normes et métriques permettant d'assurer une exploitation efficace et efficiente des informations, tout en définissant les procédures et les responsabilités garantissant le respect des droits fondamentaux des utilisateurs en préservant la sécurité des données collectées par les entreprises ou les organisations[11]. Pour cela, et depuis quarante ans maintenant, un arsenal juridique se déploie tant au niveau national qu'européen. La loi dite *informatique et libertés*[12], modifiée par la loi n° 2004-801 du 6 août 2004 pour transposer les dispositions de la directive 95/46/CE ; modifiée à nouveau par la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles suite à l'adoption du désormais célèbre Règlement (UE) 2016/679 dit Règlement général de protection des données (RGPD)[13], constituent les textes de référence en la matière. Chacune de ces dispositions vient instituer un droit fondamental à la protection des données à caractère personnel au nom de la sécurité des personnes, considérant les données comme propres à la personne, en tant qu'élément individuel privé per se, en vue de garantir un encadrement de leur exploitation qui assure le respect de la vie privée. Dans ce sens, la Convention 108 soulignait déjà en 1981 que « dans certaines conditions, l'exercice d'une complète liberté de traiter les informations risque de nuire à la jouissance d'autres droits fondamentaux ou à d'autres intérêts personnels légitimes (par exemple en matière d'emploi ou de crédit à la consommation). C'est pour maintenir un juste équilibre entre les différents droits

[10] N. Ferry-Maccario, Gestion juridique de l'entreprise, Paris, Pearson Education, 2006.
[11] Terme « Gouvernance », S. Guinchard et T. Debard, *Lexique des termes juridiques*, 25e éd., Dalloz, 2017-2018 ; J. Dionne-Proulx et G. Larochelle, *Éthique et gouvernance d'entreprise*, in Management & Avenir, 2010, Vol. 32, no 2, p. 36 ; S. Pearlman, « *Qu'est-ce que la gouvernance des données et pourquoi en avez-vous besoin ?* », Talend.com, 10 juin 2019, https://fr.talend.com/resources/what-is-data-governance/ (consulté le 2 avril 2020).
[12] Loi n° 78-17 du 6 janvier 1978, *ibid*.
[13] Règlement (UE) 2016/679, *ibid*.

et intérêts des personnes que la Convention impose certaines conditions ou restrictions au traitement d'informations »[14]. Une garantie reprise dès le 1er considérant du Règlement qui dispose ainsi que « la protection des personnes physiques à l'égard du traitement des données à caractère personnel est un droit fondamental. L'article 8, paragraphe 1, de la Charte des droits fondamentaux de l'Union européenne et l'article 16, paragraphe 1, du traité sur le fonctionnement de l'Union européenne disposent que toute personne a droit à la protection des données à caractère personnel la concernant »[15]. Une reconnaissance réitérée par la Cour Européenne des Droits de l'Homme qui considère que par principe, le simple fait de mémoriser des données relatives à la vie privée d'un individu constitue une ingérence au sens de l'article 8, de la Convention européenne des droits de l'homme qui garantit le droit au respect de la vie privée et familiale, du domicile et de la correspondance, peu importe que les informations mémorisées soient ou non utilisées par la suite[16]. La protection des données à caractère personnel est ainsi envisagée, disons même élevée en tant que droit fondamental de l'Homme dont l'abus constitue une violation, une atteinte à la vie privé. De la protection des droits individuels des personnes à leur exploitation par des organismes privés ou publics en tant que matière première, les données numériques à caractère personnel s'inscrivent pleinement dans un modèle de gouvernance juridique à consonance privatiste. Cependant, en raison de leur nature et de leur utilité sociale, certaines données comportent une dimension plus collective.

2) *La singularité des données médicales : une logique de ressource commune*

Parmi les données à caractère personnel, le Règlement distingue entre les données entendues de manière générale et certaines données dites particulièrement *sensibles*. En effet, le traitement des données à caractère personnel qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique sont interdits[17]. A la lecture de la nature même des informations faisant exception, la raison d'une interdiction de traitement par principe est évidente compte tenu de leur impact éventuel sur la protection des

[14] Rapport explicatif à la Convention 108 pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, Strasbourg, 28 janv. 1981 (point 25).
[15] Règlement (UE) 2016/679, *ibid* (considérant 1er).
[16] CEDH 4 décembre 2008, *S. et Marper c. Royaume-Uni*, Requêtes nos 30562/04 et 30566/04.
[17] Règlement (UE) 2016/679, *ibid* (article 9).

droits et libertés des personnes qui subiraient une atteinte[18]. Parmi les exceptions de traitement, il est fait mention des données à caractère personnel concernant la santé. Plus précisément, la catégorie des données de santé recense les informations relatives à la santé physique ou mentale, passée, présente ou future, d'une personne physique (y compris la prestation de services de soins de santé) qui révèlent des informations sur l'état de santé de cette personne[19]. Une définition particulièrement large qui comprend ainsi de nombreux renseignements « relatifs à une personne physique collectés lors de son inscription en vue de bénéficier de services de soins de santé ou lors de la prestation de ces services : un numéro, un symbole ou un élément spécifique attribué à une personne physique pour l'identifier de manière unique à des fins de santé ; les informations obtenues lors du test ou de l'examen d'une partie du corps ou d'une substance corporelle, y compris à partir des données génétiques et d'échantillons biologiques ; les informations concernant une maladie, un handicap, un risque de maladie, les antécédents médicaux, un traitement clinique ou l'état physiologique ou biomédical de la personne concernée (indépendamment de sa source, qu'elle provienne par exemple d'un médecin ou d'un autre professionnel de santé, d'un hôpital, d'un dispositif médical ou d'un test de diagnostic in vitro) »[20]. Des informations singulières qui représentent une ambivalence sui generis qui touche à l'essence même de ce que les textes cherchent à protéger, la vie privée des personnes, devant ainsi bénéficier d'une protection renforcée par l'interdiction de traitement, mais qui, d'un autre côté, pourraient incarner l'avenir de la démocratie sanitaire par le développement de la recherche médicale et des offres de soin.

Il s'avère effectivement que le traitement de la masse des données collectées, « bien que déjà développé dans les sciences, a conduit à des changements radicaux de paradigmes dans le domaine de la santé où il reste émergent », passant d'une recherche d'hypothèse à une recherche déduite de l'exploitation des données. Des traitements par principe interdits, pour lesquels « les différentes législations (nationales et européennes) prévoient cependant un certain nombre d'exemptions à ce principe d'interdiction fondées, soit sur l'intérêt individuel (notamment pour les soins) sous réserve d'obtenir le consentement de la personne dont les informations sont recueillies, soit sur l'intérêt collectif comme celui de la recherche »[21]. Ainsi, au nom de la notion

---

[18] A. Banck, « *RGPD : la protection des données à caractère personnel, 19 fiches pour réussir et maintenir votre conformité* », 3ème éd., Droit en poche, Gualino, Lextenso, 2020, 88p.
[19] Règlement (UE) 2016/679, *ibid* (article 4, Considérant 15).
[20] CNIL, « *Qu'est-ce qu'une donnée de santé ?* », cnil.fr, https://www.cnil.fr/fr/quest-ce-ce-quune-donnee-de-sante (consulté le 3 avril 2020) ; Règlement (UE) 2016/679, *ibid* (article 4, Considérant 35).
[21] E. Rial-Sebbag « *Chapitre 4. La gouvernance des Big data utilisées en santé, un enjeu national et international* », Journal international de bioéthique et d'éthique des sciences, 2017/3 (Vol. 28), p. 39-50.

d'intérêt général, fondamentale pour le droit français mais également reprise en droit européen, autrement dit de ce qui est bien pour le public[22], pour la communauté dans son ensemble, le modèle de gouvernance juridique des données personnelles change de paradigme et prend une toute autre dimension. Véritable enjeu du développement de la recherche médicale, les données ne constituent plus une matière première exploitable au sens économique du terme, mais davantage un matériau, une ressource commune servant de base pour l'enrichissement de la santé tant individuelle que collective[23]. Au sein d'un modèle construit autour de la protection des droits fondamentaux des personnes, où les données sont considérées comme relevant du domaine intime propre à l'individu et de sa vie privée, les données de santé viennent établir une exception qui fait basculer la conception de la donnée d'une matière à caractère personnel à la notion d'une ressource commune pour le bien public. En dépit des risques envers les droits des personnes, l'exploitation des données de santé renfermerait une réserve de trésors pour la sécurité sanitaire dans la mesure où elle « concourt, notamment, à la transparence et l'efficacité du système de santé (;) la collecte et la diffusion de ces informations permettent, en effet, d'alimenter le débat public sur la santé, et, en particulier, de nourrir l'élaboration, la conduite et l'évaluation des politiques publiques de santé (...) d'apporter une aide déterminante à la vigilance pharmaco-épidémiologique, d'améliorer l'efficacité des parcours de soin[16], de favoriser la recherche de longue durée sur les protocoles de soin et de permettre une veille sanitaire renforcée »[24]. Les données personnelles de santé sortent ainsi de l'approche individualiste traditionnelle, par la masse globale d'informations qu'elles constituent, pour être considérées en tant que *bien commun*, qui per se n'appartient à personne et dont l'usage est commun à tous. Des données personnelles communes, un bel oxymore qui ne l'est en réalité que d'apparence, celles-ci n'étant plus entendues de manière individuelle mais sous le prisme d'un réseau qui leur confère une dimension collective. Cela dit, si l'intérêt général préconise de considérer les données de santé comme une ressource publique commune au nom du bien collectif, ces informations particulièrement sensibles doivent être utilisées avec la plus grande prudence au sein d'un cadre juridique qui vient précautionneusement en délimiter les conditions. Sur ce point, le Règlement européen a pris le parti de laisser une importante marge

---

[22] V. Intérêt général *in* G. Cornu (dir.) et Association Henri Capitant, *Vocabulaire juridique*, Paris, Presses universitaires de France, coll. « Quadridge », 2005, 7e éd., 970 p.

[23] F. Lesaulnier, « *Recherche en santé et protection des données personnelles à l'heure du RGPD* », *in* E. Netter, V. Ndior, J-F. Puyraimond, S. Vergnolle, « *Regards sur le nouveau droit des données personnelles* », Centre de droit privé et de sciences criminelles d'Amiens, 2019.

[24] J-M. Sauve, « *Intervention de Jean-Marc Sauvé lors des septièmes entretiens du Conseil d'État en droit social* », Santé et protection des données, Conseil d'État, 1er décembre 2017, https://www.conseil-etat.fr/actualites/discours-et-interventions/sante-et-protection-des-donnees (consulté le 3 avril 2020).

de manœuvre aux États pour déterminer au niveau national le régime juridique applicable aux autorisations dérogatoires concernant les traitements qui portent sur les données de santé et les traitements à des fins de recherche scientifique[25]. Une liberté nationale mise à profit par le législateur français via l'adoption d'un nouveau chapitre IX de la loi n° 2018-493 du 20 juin 2018[26] consacré aux traitements de données de santé applicable aux recherches, études et évaluations déterminant le régime juridique français applicable.

B. Le régime juridique dérogatoire de la gouvernance des données médicales

« L'informatique doit être au service de chaque citoyen. (...) Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques »[27]. Des propos de 1978 qui marquent cependant l'état d'esprit général du Règlement sur la protection des données et des normes nationales en la matière. Si la dérogation d'interdiction de traitement des données sensibles a été accordée aux données de santé en raison de la dimension d'intérêt général qu'elles représentent, cela n'est possible que dans le cadre d'un régime juridique d'une clarté suffisante et qui réponde à un principe de proportionnalité entre l'intérêt du traitement et l'atteinte à la vie privée engendrée. En effet, le Conseil Constitutionnel a rappelé que le respect des droits fondamentaux[28] implique le droit au respect de la vie privée ; ainsi la collecte, l'enregistrement, la conservation, la consultation et la communication de données à caractère personnel doivent être justifiés par un motif d'intérêt général et mis en œuvre de manière adéquate et proportionnée à cet objectif de façon à concilier les différents intérêts en jeu[29]. Une décision qui fait écho à la jurisprudence européenne sur le sujet, consacrant l'importance de la protection des données médicales pour la jouissance par une personne du droit au respect de sa vie privée, et jugeant que le droit applicable doit définir avec une netteté suffisante l'étendue et les modalités d'exercice du pouvoir d'appréciation accordé aux autorités compétentes dans le cadre du traitement des données médicales. La Cour ayant ainsi conclu à la violation de la vie privée pour imprécision des dispositions du droit interne autorisant l'accès d'un organisme public au dossier médical de la requérante[30].

---

[25] Règlement (UE) 2016/679, *ibid* (article 9.4).
[26] Loi n°2018-493 du 20 juin 2018 relative à la protection des données personnelles, JORF n°0141 du 21 juin 2018.
[27] J-M. Sauve, « *Intervention de Jean-Marc Sauvé lors des septièmes entretiens du Conseil d'État en droit social* », Op. cit.
[28] Mentionnant spécifiquement la Liberté proclamée par l'article 2 de la Déclaration des droits de l'homme et du citoyen de 1789.
[29] Cons. Const. n°2019-797 QPC du 26 juillet 2019, *Unicef France et autres*.
[30] CEDH 29 avril 2014, *L.H. c. Lettonie*, Requête 52019/07 : La Cour a constaté que la loi applicable n'avait pas indiqué avec suffisamment de clarté l'étendue du pouvoir discrétionnaire conféré aux autorités compétentes et les modalités de son exercice pour collecter des données médicales.

Conformément au principe de proportionnalité qui gouverne la conciliation du respect de la vie privée et de l'intérêt général sanitaire du traitement des données de santé, celles-ci ne peuvent être exploitées que dans des circonstances déterminées par le Règlement européen comme suit :

- Si la personne concernée a donné son consentement explicite au traitement de ces données à caractère personnel pour une ou plusieurs finalités spécifiques ;
- Si le traitement est nécessaire aux fins de l'exécution des obligations et de l'exercice des droits propres au responsable du traitement ou à la personne concernée en matière de droit du travail, de la sécurité sociale et de la protection sociale ;
- Si le traitement est nécessaire à la sauvegarde des intérêts vitaux de la personne concernée ou d'une autre personne physique, dans le cas où la personne concernée se trouve dans l'incapacité physique ou juridique de donner son consentement ;
- Si le traitement est nécessaire aux fins de la médecine préventive ou de la médecine du travail, de l'appréciation de la capacité de travail du travailleur, de diagnostics médicaux, de la prise en charge sanitaire ou sociale, ou de la gestion des systèmes et des services de soins de santé ou de protection sociale sur la base du droit de l'Union, d'un État membre ou en vertu d'un contrat conclu avec un professionnel de la santé ;
- Si le traitement est nécessaire pour des motifs d'intérêt public dans le domaine de la santé publique, tels que la protection contre les menaces transfrontalières graves pesant sur la santé, ou aux fins de garantir des normes élevées de qualité et de sécurité des soins de santé et des médicaments ou des dispositifs médicaux ;
- Si le traitement est nécessaire à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques de manière proportionné à l'objectif poursuivi[31].

En dehors des circonstances mentionnées par le Règlement qui peuvent justifier le traitement des données médicales, il en ressort des conditions de traitement tant de fond que de procédure, au respect desquelles il faut veiller. Il s'avère effectivement que le traitement des données de santé doit *être nécessaire*, la personne concernée par le traitement devant fournir un *consentement explicite* pour *une ou plusieurs finalités spécifiques*.

1) Un consentement explicite

Les personnes dont les données médicales sont collectées disposent de droits. Ainsi, tout traitement de données de santé doit reposer, pour être licite, sur le fondement du consentement

[31] Règlement (UE) 2016/679, *ibid* (article 9.2).

des personnes dont les données sont collectées et utilisées[32]. Pour être conforme, le Règlement exige que le consentement de la personne concernée soit réalisé à travers *toute manifestation de volonté, libre, spécifique, éclairée et univoque* par laquelle la personne concernée accepte, par une déclaration ou par un acte positif clair, que des données à caractère personnel la concernant fassent l'objet d'un traitement[33]. *Libre*, dans le sens où la personne concernée doit pouvoir être en mesure de refuser ou de retirer son consentement sans subir d'influence, de contrainte ou un préjudice quelconque. Tel serait le cas, par exemple dans une situation où les rapports de force ne seraient pas égaux (notamment dans le cadre des relations de travail via le lien de subordination entre l'employeur et son employé) ; mais encore d'imposer un coût supplémentaire ou des sanctions aux personnes qui souhaiteraient ne pas donner leur consentement ou le retirer[34]. *Spécifique* ensuite, dans la mesure où le consentement est donné par les personnes concernées pour une finalité précise ou plusieurs de manière granulaire, dans l'hypothèse de la nécessité d'une pluralité de traitement pour une même finalité. *Éclairé*, puisque les personnes concernées doivent avoir été informées préalablement au recueil de leur consentement afin de pouvoir le donner, ou le refuser, en toute connaissance de cause. Cela via une information complète[35] communiquée de manière claire, accessible et compréhensible pour une entière intelligibilité. Autrement dit, l'information doit avoir été délivrée de façon concise, transparente et abordable par le grand public, qui plus est de façon adaptée, en fonction de la pathologie de la personne, de son âge, des circonstances du recueil des données (mineurs, majeurs incapables, etc) ; l'objectif étant de permettre à ces personnes de conserver la maîtrise des données utilisées les concernant. Cette obligation peut être remplie dès lors qu'elle est assurée par une mention figurant sur le site internet du responsable du traitement, des organismes d'assurance maladie et sur des supports permettant de la porter à la connaissance des personnes, notamment des affiches dans les locaux ouverts au public ou des documents qui leur sont remis[36], ou encore dans le livret d'accueil remis au patient à l'occasion de son hospitalisation. Et enfin, le consentement doit être *univoque* ; donné par un acte express, positif clair qui se distingue de tous les autres ne laissant place à aucune ambigüité sur la teneur et l'étendue du consentement. Celui-ci peut ainsi se matérialiser par un courrier électronique ou

[32] Règlement (UE) 2016/679, *ibid* (article 6.1).
[33] Règlement (UE) 2016/679, *ibid* (article 4.11).
[34] A. Banck, « *RGPD : la protection des données à caractère personnel, 19 fiches pour réussir et maintenir votre conformité* », Op. cit.
[35] Règlement (UE) 2016/679, *ibid* (article 13.1).
[36] CNIL n°2018-365 du 20 décembre 2018, *Union nationale des organismes d'assurance maladie complémentaire*, Saisine n°918103.

une lettre[37], via les cases à cocher en ligne bien connues ; « par exemple, si un responsable de traitement envoie un courrier électronique à une personne l'informant de son intention de traiter des informations médicales, ce message devra formaliser une demande de consentement et expliquer les caractéristiques et finalités du traitement (...) la personne concernée devra consentir expressément à l'utilisation de ses données, par exemple, en répondant qu'elle consent (par retour d'email) ou en cliquant sur un lien de vérification ou encore en renseignant un code reçu par SMS »[38]. Cependant, il est nécessaire de préciser que le traitement des données personnelles de santé est autorisé, et ce sans consentement préalable de l'utilisateur, s'il poursuit par exemple une finalité de gestion des systèmes et services de santé ou de la protection sociale, de préservation de la santé publique (pour éviter notamment la propagation des maladies), de l'appréciation médicale (soins, diagnostics, médecine préventive) ou encore de préservation des intérêts vitaux d'une personne en incapacité de donner son consentement[39].

2) Un traitement nécessaire pour des finalités spécifiques

Dès lors que les personnes concernées ont été informées et ont exprimé leur consentement, le traitement de données de santé doit être accompagné de modalités de collecte et d'exploitation apportant les garanties appropriées pour garantir la sécurité et la préservation de la vie privée. Pour cela, « cinq principes gouvernent cette protection : 1) les données doivent être collectées de manière loyale et licite 2) en vue d'une finalité déterminée et légitime ; 3) les données collectées doivent être pertinentes et adéquates au regard de cette finalité, 4) elles doivent être complètes et exactes et 5) leur conservation doit être prévue pour une durée définie » [40]. Faisant partie intégrante des principes qui gouvernent de nombreux systèmes juridiques, à l'image de la bonne foi, *la loyauté* « se dit aussi de la bonne qualité des choses, de ce qui a la condition requise par la loi, par l'ordonnance dit le Dictionnaire de Furetière. En ce sens, s'agissant des choses, la notion de loyauté devient très proche de celle de légalité. C'est un point de passage entre la notion morale et la notion juridique »[41]. Aucun étonnement donc qu'elle soit hissée en tant que condition de traitement des données à caractère personnel,

---

[37] Groupe de travail de l'article 29 sur la protection des données (10 avril 2014), Avis 05/2014 sur les techniques d'anonymisation.
[38] C. Galichet, « *Un consentement valide au sens du RGPD* », Village de la Justice, 15 janvier 2019, https://www.village-justice.com/articles/consentement-valide-sens-rgpd,30428.html (consulté le 6 avril 2020).
[39] S. Goldstein, « *Le RGPD et les données de santé* », Legalplace, https://www.legalplace.fr/guides/rgpd-donnees-sante/ (consulté le 6 avril 2020).
[40] J-M. Sauve, « *Intervention de Jean-Marc Sauvé lors des septièmes entretiens du Conseil d'État en droit social* », Op. cit.
[41] W. Cherbonnier, L. Crochet, et Collectif, « *La loyauté : de la règle morale au principe juridique* », Revue juridique de l'Ouest, 2012-3. pp. 327-342.

notamment des données médicales dites sensibles. Associée aux notions de licéité et de transparence, elle est consacrée par le Règlement[42]. Ainsi, pour répondre à l'obligation de traiter les données de manière licite, loyale et transparente au regard de la personne concernée, le traitement doit reposer sur un traitement nécessaire[43] et conforme de l'information reçue aux fins d'éviter les traitements occultes ou cachés[44]. Un traitement nécessaire est également celui qui remplit *une finalité déterminée, explicite et légitime* ; autrement dit qui répond à un objectif précis poursuivi par le responsable du traitement des données[45]. En conséquent, la ou les finalités granulaires doivent être déterminées préalablement au traitement des données, communiquées aux personnes concernées dans le cadre de leur information, et légitimes vis-à-vis de l'activité de l'organisme exerçant le traitement. Encore une fois, et afin de permettre une mise en œuvre efficiente des traitements d'intérêt général, le Règlement et les normes nationales peuvent concéder quelques facilités de gestion ; par exemple, le traitement ultérieur à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques n'est pas considéré comme incompatible avec les finalités initiales[46]. Si le traitement doit avoir une finalité déterminée, il est nécessaire de préciser que *les données collectées doivent être pertinentes et adéquates* au regard de cette finalité. Un principe qui vise à garantir que les données qui seront collectées par l'organisme de traitement sont strictement nécessaires à la finalité, excluant ainsi toutes autres données qui n'auraient pas de lien avec l'objectif poursuivi, sous peine de sanctions[47]. Ces mêmes données pertinentes doivent également être *complètes et exactes*, dans le sens où elles doivent être à jour ; le responsable du traitement doit s'assurer, par la mise en œuvre de mesures appropriées, de traiter une base de données d'actualité et non obsolète. Une mise à jour régulière des données traitées qui emmène au dernier point établissant que *la conservation des données doit être prévue pour une durée définie*. Sauf exception pour motifs d'intérêt public, ces données ne peuvent pas faire l'objet d'un stockage *ad vitam aeternam* mais doivent intervenir en tant que base active lorsqu'elles sont nécessaires pour un projet singulier, dans l'intervalle d'une durée déterminée permettant de répondre à cet objectif[48]. Une durée définie qui n'empêche cependant pas l'exercice de

[42] Règlement (UE) 2016/679, *ibid* (article 5.1).
[43] Règlement (UE) 2016/679, *ibid* (article 6).
[44] A. Banck, « *RGPD : la protection des données à caractère personnel, 19 fiches pour réussir et maintenir votre conformité* », Op. cit.
[45] CNIL n°2018-365 du 20 décembre 2018, *ibid*.
[46] Loi n° 2018-493 du 20 juin 2018, *ibid*. (article 54 I) ; Règlement (UE) 2016/679, *ibid* (article 5.1).
[47] CNIL n° 2011-205 du 6 octobre 2011, *Société X* ; CNIL n° SAN 2019-010 du 21 novembre 2019, *Société Futura Internationale*.
[48] Berlin Commissioner for Data Protection and Freedom of Information v. Deutsche Wohnen SE, October 30th 2019, in "*Berlin Commissioner for Data Protection Imposes Fine on Real Estate Company*", European Data

certains droits des personnes relatifs à la conservation de leurs données. Il s'agit notamment du *droit d'accès, de rectification et d'opposition* permettant aux personnes concernées à la fois d'interroger le responsable du traitement pour savoir si leurs données sont soumises à un traitement et, le cas échéant d'obtenir une copie de ces informations, pour la modification des données qui s'avèreraient incomplètes ou inexactes, mais encore pour s'opposer au traitement des données à l'exception de finalités poursuivant l'exécution d'une mission d'intérêt public ou qui semblent nécessaires aux fins des intérêts légitimes poursuivis par l'organisme[49]. Des facultés qui peuvent s'accompagner d'un droit à l'effacement dit aussi *droit à l'oubli* consistant pour les personnes qui le souhaitent à obtenir l'effacement des informations les concernant dans les conditions prévues par le Règlement[50], notamment lorsque le traitement n'est plus nécessaire, est illicite ou que la personne retire son consentement. Nécessairement, l'exploitation des données médicales implique des *modalités d'exercice de ces droits* qui soient adaptées aux intérêts poursuivis. En matière de recherche scientifique par exemple, définie de manière très large par le Règlement[51], il est concédé une dérogation au principe d'interdiction du traitement des données de santé, une présomption de compatibilité de la finalité de recherche scientifique avec une finalité initiale différente, la possibilité de conservation au-delà de la réalisation du traitement, le recueil d'un consentement pour une ou plusieurs finalités spécifiques, la possibilité de déroger à l'obligation d'information en cas de réutilisation secondaire des données si elle se révèle impossible ou exigerait des efforts disproportionnés, de même que des conditions spécifiques d'application du droit à l'effacement[52].

Tout l'enjeu des conditions juridiques susmentionnées de traitement des données médicales consiste à opérer la délicate, mais pourtant nécessaire, conciliation entre le respect du droit des personnes, qui constitue la priorité du règlement général sur la protection des données (RGPD), et l'intérêt, à la fois personnel en matière d'amélioration des soins, et général aux fins de promotion de la recherche médicale ; cela par la mise en place tantôt de mesures de sécurité, tantôt dérogatoires. Des motifs d'intérêt généraux sanitaires qui poussent à voir les choses en grand, incitant le modèle européen, mais également les systèmes nationaux respectifs, à entrevoir la perspective d'une ouverture généralisée du traitement des données de santé.

Protection Board, 5 novembre 2019, https://edpb.europa.eu/news/national-news/2019/berlin-commissioner-data-protection-imposes-fine-real-estate-company_fr (consulté le 7 avril 2020).

[49] Règlement (UE) 2016/679, *ibid* (articles 15, 16 et 21).

[50] Règlement (UE) 2016/679, *ibid* (article 17).

[51] Règlement (UE) 2016/679, *ibid* (considérant 159).

[52] F. Lesaulnier, « *Recherche en santé et protection des données personnelles à l'heure du RGPD* », Op. cit.

II. LA PERSPECTIVE D'UN MODELE DE GOUVERNANCE DES DONNÉES MÉDICALES EN *OPEN DATA*

Les technologies sont de plus en plus performantes chaque jour, une performance qui demande toujours plus de données nécessaires au développement des systèmes d'apprentissage profond. Ainsi, dans nos sociétés numérisées capitalistes articulées autour d'une production qui se doit d'être toujours plus rapide et le moins coûteuse possible, le modèle de gouvernance des données doit être adapté. Une adaptation qui s'exprime par la volonté de faciliter le traitement des données de santé via la catégorisation en tant que ressources communes dans un mouvement d'ouverture renforcée, dit en *open data*. Cependant, cette libéralisation des données de santé doit être opérée avec la plus grande prudence, de manière pédagogique afin que les utilisateurs puissent s'approprier ces informations conformément au respect des droits des personnes, et de façon progressive pour permettre d'en préciser et d'en ajuster les modalités de mise en œuvre. Un système de gouvernance des données de santé qui va devoir s'adapter aux difficultés de sécurisation et aux risques engendrés envers la protection des droits fondamentaux.

A. L'opportunité d'une gouvernance des données médicales en *open data*

La perspective de sortir du paradigme individualiste traditionnel de gouvernance des données pour les considérer comme une ressource commune ouverte, dont l'usage serait ouvert à tous, s'inscrit dans la volonté d'établir un système de gouvernance global de grande échelle. L'ouverture en *open data* représente des opportunités qui doivent nécessairement être envisagées de manière progressive et pédagogique afin de garantir le respect des droits des personnes, notamment via des procédés de sécurisation permettant de détacher les données médicales de l'identité même des personnes qu'elles concernent.

1) L'ambition européenne d'une base de données de santé anonymisées en open data

Des paradigmes de gouvernance conciliant le traitement des données de santé et le respect des droits fondamentaux des personnes qui ne sont pas incompatibles et certains de leurs concepts respectifs peuvent être réunis par la spécificité des données médicales dans l'hypothèse structurelle d'une sécurisation adaptée. Une conciliation qui n'en est qu'à ses prémices, la volonté étant d'aller bien au-delà du système actuel, promouvant un modèle de gouvernance global, commun à tous les pays membres de l'Union Européenne, qui tend vers l'ouverture généralisée de l'accès aux données de santé. L'idée étant que la gouvernance des données de santé puisse se faire de la manière la plus globale possible, et cela dans le sens d'une

politique commune au niveau européen, mais également en utilisant un spectre de donnée maximal dans le cadre de traitements le plus largement diversifiés. Autrement dit, il s'agit d'intégrer le traitement des données de santé dans le mouvement d'une science ouverte afin de constituer une base, un terreau commun permettant le développement de la recherche médicale et l'amélioration des soins personnalisés en Europe. En effet, « au cœur de ce changement, de nouvelles formes de procédés décisionnels automatisés permettent depuis peu un traitement inédit des données de masse, données brutes, hétérogènes, dynamiques, caractéristiques des *Big Data* »[53] grâce aux systèmes d'intelligence artificielle et aux algorithmes d'apprentissage profond qui sont de plus en plus performants. Un changement qui n'épargne pas le domaine de la santé dans la mesure où, de la même façon que l'Homme acquiert de nouvelles connaissances, la machine offre de nouvelles perspectives[54] en s'aidant des informations mise à sa disposition. Ainsi, plus la machine dispose de données médicales, plus le système est intelligent et s'enrichit lui-même de manière autonome. Tout l'enjeu du développement des nouvelles technologies en matière médicale s'articule autour de l'accès à ses fameuses données de santé ; et l'intention générale et d'en faciliter le traitement via un système collectif ouvert. Si les modèles de réglementation et de recherche actuels reposent sur un accès restreint aux données de santé, y compris les données individuelles des patients, alors le renforcement et l'extension de l'utilisation et de la réutilisation des données de santé sont essentiels pour l'innovation médicale. Ainsi, une libéralisation aiderait les autorités de santé à prendre des décisions quant aux systèmes de santé mais contribuerait également à la compétitivité de l'industrie européenne tout en soutenant de manière significative le travail des organismes de réglementation du système de santé, l'évaluation des produits médicaux, la démonstration de leur innocuité et de leur efficacité, améliorant les soins prodigués aux patients. Pour aller dans ce sens, la Commission Européenne a précisé via la communication du 19 février 2020 relative à la stratégie européenne des données[55], son ambition d'établir un *espace européen commun de données de santé*. Le dessein étant d'élaborer « des mesures législatives ou non législatives sectorielles pour l'espace européen de données sur la santé, complétant le cadre horizontal de l'espace commun de données pour éliminer les obstacles à la fourniture transfrontières de services et de produits de santé numériques » ; mais également de « déployer les infrastructures

---

[53] J-M Deltorn, « *La protection des données personnelles face aux algorithmes prédictifs* », Op. cit.
[54] Collectif (2017), « *Une stratégie pour la France en matière d'intelligence artificielle* », Rapport de synthèse FranceIA, Gouvernement français, présentée le 21 mars 2017 à la Cité des Sciences et de l'Industrie, https://cache.media.enseignementsup-recherche.gouv.fr/file/Actus/85/9/Rapport_synthese_France_IA_738859.pdf .
[55] Commission Européenne, « *Une stratégie européenne pour les données* », COM(2020) 66 final, Bruxelles, 19 février 2020.

de données, les outils et les capacités de calcul pour l'espace européen des données de santé en soutenant le développement des dossiers de santé électroniques nationaux (DSE) » et l'interopérabilité des données médicales. Un espace européen commun des données de santé qui certes, paraît essentiel pour les progrès dans la prévention, la détection et la guérison des maladies, mais ne peut s'opérer dans n'importe quelles conditions. La Commission rappelle que les citoyens européens doivent être rassurés sur la protection de leurs droits fondamentaux dans le cadre de cette mise à disposition commune et ouverte des données de santé qui doit se réaliser d'une manière conforme au RGPD.

Au-delà des garanties offertes par le RGPD, dans l'intention de considérer les données médicales en tant que ressources communes ouvertes, que ce soit au niveau européen ou national, le principal outil de sécurisation des données de santé offert aux personnes concernées réside essentiellement, sur le fondement du droit fondamental à la vie privée, dans la préservation de leur identité. Après tout, lorsque le célèbre Ulysse se fait capturer par le cyclope Polyphème, dans l'illustre ouvrage d'Homère, le roi d'Ithaque prétend s'appeler *Personne* en réponse au géant qui l'oblige à lui indiquer son nom ; une ruse qui lui permet, à lui et à ses compagnons d'infortune de ne pas être pourchassés par les autres cyclopes une fois sortis des griffes de Polyphème. Tout comme l'anonymat à sauver Ulysse et ses compagnons, l'enjeu du traitement ouvert des données de santé réside dans l'identification, ou a contrario dans l'impossibilité d'identifier les personnes concernées par les données médicales diffusées et exploitées. Et cela, que cette identification soit directe ou indirecte, par référence à un identifiant ou tout élément qui lui soit propre et qui, seul ou avec d'autres, permettrait de remonter aux individus. A noter que les données médicales sont des informations particulièrement sensibles au sens du RGPD[56], celui-ci exigeant ainsi que les données soient traitées de façon à garantir une sécurité appropriée à l'aide de mesures techniques ou organisationnelles adaptées, notamment via l'intégrité et la confidentialité. Pour cela, de nombreux outils, dont les techniques et l'efficacité diffèrent, sont à disposition pour apporter des garanties relatives à la sécurité des données et à l'efficience des droits fondamentaux. Parmi ces techniques, il est incontournable de mentionner la pseudonymisation, largement utilisée, et mentionnée dans le Règlement comme un traitement de données à caractère personnel réalisé de telle façon que celles-ci ne puissent plus être attribuées à une personne précise sans avoir

[56] Règlement (UE) 2016/679, *ibid* (article 5.1).

recours à des informations supplémentaires[57]. Une définition qui couvre différentes techniques couramment utilisées en matière de recherche en santé tels que « le recours à une table de correspondance entre le jeu de données pseudonymes (codées) et les données d'identité conservées séparément, classiquement utilisée dans les essais cliniques; les fonctions de hachage utilisées avec un secret qui permettent de chaîner des données relatives à un individu et de suivre son parcours dans le temps sans permettre de l'identifier (Déclaration obligatoire des maladies, PMSI, SNIIRAM...) »[58]. Ces données restent soumises au RGPD dans la mesure où, étant simplement pseudonymisées, elles permettent tout de même la réidentification des personnes, que ce soit via une clef ou un code de cryptage. A l'inverse, l'anonymisation, quant à elle, consiste à rendre totalement impossible, en pratique, toute identification de la personne par quelque moyen que ce soit et ce de manière irréversible ; une irréversibilité qui, si elle est effective, implique que les données anonymisées ne soient plus à caractère personnel, et ne s'inscrivent donc plus dans le champ d'application du RGPD. Cependant, pour qu'une solution d'anonymisation soit construite de manière efficiente, le G29[59] propose trois critères indicatifs qui sont le test de l'individualisation, c'est-à-dire l'impossibilité d'isoler une partie ou la totalité des enregistrements identifiant un individu dans l'ensemble de données ; la corrélation, autrement dit l'incapacité de relier entre eux, au moins deux enregistrements se rapportant à la même personne concernée ou à un groupe de personnes concernées (soit dans la même base de données, soit dans deux bases de données différentes) ; et enfin l'inférence, l'impossibilité de déduire, avec un degré de probabilité élevé, la valeur d'un attribut à partir des valeurs d'un ensemble d'autres attributs[60]. Quoiqu'il advienne, « aucune technique ne permettant d'anonymiser à 100% les données, il est recommandé soit d'informer les personnes de manière complète, soit de recueillir l'avis de la CNIL sur le procédé d'anonymisation envisagé avant d'avancer dans un tel projet », notamment en matière de données sensibles telles que les données de santé[61]. Une technique particulièrement intéressante puisque dans le cadre de l'*open data*, l'anonymisation pourrait représenter, selon la CNIL, une piste envisageable de traitement ouvert des données de santé par la mise en œuvre d'un système de publication en ligne des

---

[57] Règlement (UE) 2016/679, *ibid* (article 4.5).
[58] F. Lesaulnier, « *Recherche en santé et protection des données personnelles à l'heure du RGPD* », Op. cit.
[59] Groupe de travail de l'article 29 sur la protection des données (10 avril 2014), Avis 05/2014 sur les techniques d'anonymisation ; CNIL, « *Le G29 publie un avis sur les techniques d'anonymisation* », cnil.fr, 16 avril 2014, https://www.cnil.fr/fr/le-g29-publie-un-avis-sur-les-techniques-danonymisation (consulté le 9 avril 2020).
[60] C. Galichet, « *Données personnelles : anonymisation ou pseudonymisation ?* », Village de la Justice, 17 janvier 2017, https://www.village-justice.com/articles/donnees-personnelles-anonymisation-pseudonymisation,26194.html (consulté le 9 avril 2020).
[61] C. Galichet, « *L'anonymisation des données personnelles selon le conseil d'État : arrêt JCDecaux du 8 février 2017* », Village de la Justice, 21 mars 2017, https://www.village-justice.com/articles/anonymisation-des-donnees-personnelles-selon-Conseil-Etat-arret-JCDecaux,24541.html (consulté le 9 avril 2020).

informations publiques sans données personnelles[62]. L'idée générale de l'anonymisation s'exprime à la fois par la transformation des données pour qu'elles ne se réfèrent plus à une personne réelle et par la généralisation de façon à ce qu'elles ne soient plus spécifiques à un individu mais communes à un ensemble de personnes[63]. Une technique qui s'accorde relativement bien à l'ambition d'un système européen de données médicales en vue d'un traitement de celles-ci en tant que ressources communes ouvertes à tous. Ainsi, « bien que la contrainte, incontournable, de l'anonymisation limite dans certains cas les usages des données, la mise à disposition du public est complémentaire de la mise à disposition des acteurs plus spécialisés et porte également un potentiel important d'utilité sociale et de valorisation »[64] tout en apportant les garanties nécessaires au respect des droits fondamentaux en ne permettant pas l'identification des personnes. Cependant, le déploiement d'une telle base de données exploitable par tous en *open data* ne se fait pas du jour au lendemain, mais doit nécessairement être mise en œuvre pas à pas pour garantir l'efficience et la sécurité du système envisagé.

2) Une démarche française pédagogique et progressive pour la protection des personnes

Un système de base de données de santé en *open data*, à la disposition de chacun, qui ne laisse pas indifférent, constituant d'innombrables opportunités pour la recherche et le développement médical pour certains, et une source de craintes et d'interrogations pour d'autres. Il est clair que « les évolutions technologiques et sociales, comme la montée en puissance des intérêts généraux ou privés autour de l'accès aux données médicales, créent une situation particulièrement instable et appellent des réglages fins qui relèvent, dans leur principe, du législateur et, pour leur mise en œuvre, de l'autorité de régulation et du juge »[65]. Des réglages effectués de manière pédagogique par le législateur français au fil des réformes qui se sont succédées depuis la première règlementation en matière de régulation des données en 1978[66]. Envisager une base de données de santé globale en *open data* est une chose, mais disposer de ces fameuses données en est une autre ; et sur ce point, la source a été toute trouvée dans la valorisation du système d'assurance maladie compte tenu de la multitude et de l'hétérogénéité

---

[62] CNIL, « *L'anonymisation des données, un traitement clé pour l'open data* », cnil.fr, 17 octobre 2019, https://www.cnil.fr/fr/lanonymisation-des-donnees-un-traitement-cle-pour-lopen-data (consulté le 9 avril 2020).
[63] CNIL, « *Le G29 publie un avis sur les techniques d'anonymisation* », cnil.fr, 16 avril 2014, https://www.cnil.fr/fr/le-g29-publie-un-avis-sur-les-techniques-danonymisation (consulté le 9 avril 2020).
[64] H. Caillol, « *Ouverture des données de santé : l'expérience de l'Assurance maladie* », Informations sociales, vol. 191, no. 5, 2015, pp. 60-67.
[65] J-M. Sauve, « *Intervention de Jean-Marc Sauvé lors des septièmes entretiens du Conseil d'État en droit social* », Op. cit.
[66] Loi n° 78-17 du 6 janvier 1978, *ibid*.

des informations à sa disposition. La démarche française, qu'elle ait été anticipée sur le long terme ou simplement améliorée au fil des textes, s'est construite sur la création de base de données graduellement enrichie par des ouvertures progressivement renforcées. Ainsi, la création d'un Système National Inter-Régimes d'Assurance Maladie (SNIIRAM) a été prévue par la loi de financement de la sécurité sociale pour 1999 et « après plusieurs années de travail technique et l'accord de la CNIL, un entrepôt de données a été constitué à partir de 2003, puis complété et enrichi (notamment via la création du dossier médical partagé et du dossier pharmaceutique) au fil des années, constituant aujourd'hui une source d'information très riche sur la santé de la population et le fonctionnement du système de soins »[67]. Des modalités de mise en œuvre de l'accès, de la collecte et du traitement des données de santé qui ont été précisées par la loi du 26 janvier 2016[68] redéfinissant complètement la politique d'accès aux données de santé au profit d'une ouverture renforcée ; se traduisant par le regroupement de l'ensemble des bases de données de santé en un seul fichier, le système national des données de santé (SNDS), composé entre autres du SNIIRAM, et la création d'un institut national pour traiter les demandes d'autorisation d'accès aux données conformément aux finalités de traitement autorisées. Un mouvement vers l'*open data* poursuivit récemment avec l'adoption de la loi portant l'organisation et la transformation du système de santé[69] qui prévoit à la fois l'élargissement du système national des données de santé (SNDS) et la création du Health Data Hub. Autrement dit, le texte prévoit l'élargissement du SNDS à un grand nombre de sources de données, notamment cliniques et collectées lors des actes pris en charge par l'assurance maladie (résultats d'analyse de biologie, imagerie, compte-rendu médicaux, etc), par la création d'un *Health Data Hub* constituant à la fois une plateforme technologique et l'institut qui l'administre. Il est nécessaire de préciser qu'il ne s'agit pas encore de « mettre en place une base de donnée nationale unique mais de prévoir que pour toutes ces sources de données, les règles d'accès et de traitement sécurisé soient les mêmes pour permettre une meilleure lisibilité pour l'ensemble des parties prenantes »[70].

---

[67] C. Gissot, D. Polton, « *Les bases de données de l'assurance-maladie : un potentiel pour l'amélioration du système de santé et pour la recherche* », Statistique et Société, Vol. 2, n°2, mai 2014, p. 19-24.
[68] Loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé, JORF n°0022 du 27 janvier 2016.
[69] Loi n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé, JORF n°0172 du 26 juillet 2019.
[70] Institut National des Données de Santé (INDS), « *Impact de la loi relative à l'organisation et à la transformation du système de santé sur les données de santé* », Mise en place du Health Data Hub, Plateforme des données de santé, https://www.indsante.fr/fr/impact-de-la-loi-relative-lorganisation-et-la-transformation-du-systeme-de-sante-sur-les-donnees-de (consulté le 9 avril 2020).

Une ouverture renforcée au fil des textes, réalisée progressivement par étape, qui a permis de tirer profit des différentes expériences ultérieures. Il s'avère effectivement des suites de l'expérience produite par le système du Sniiram que l'ouverture des données en elle-même ne suffit pas, mais qu'il est nécessaire de prévoir un programme d'accompagnement des utilisateurs et des difficultés techniques de protection de la confidentialité des données personnelles. Matériellement, « les utilisateurs ont besoin, pour pouvoir utiliser et interroger les bases de données de manière pertinente et efficace, d'une meilleure connaissance des données, lesquelles proviennent principalement des remboursements opérés par l'Assurance maladie et n'ont pas été recueillies initialement dans un objectif d'études » ; ce qui implique de sensibiliser et mettre en œuvre une offre pédagogique complète et spécifique « sur le cadre d'usage des données, de fournir une documentation, un accompagnement et des outils informatiques et de traitement des données adaptés à un usage en dehors de l'institution »[71]. Concernant les difficultés techniques posées par le respect de la confidentialité des données, des mises à l'épreuve sont régulièrement opérées. Par exemple, un *Hackaton Santé* a été organisé le 26 janvier 2015 en collaboration avec la mission Etalab[72], permettant de mettre en lumière les défis à relever concernant les compétences à développer dans l'analyse des risques de réidentification pour l'anonymisation ainsi que dans l'accompagnement des utilisateurs[73]. Une base de données de santé en *open data* qui reposerait donc sur deux piliers fondamentaux qui sont la garantie de la confidentialité des données via des techniques d'anonymisation ne permettant pas la réidentification des personnes, ainsi qu'une politique de responsabilisation des acteurs du système à deux niveaux. Le premier à travers la sensibilisation du rôle des personnes concernées par les données de santé qui sous couvert de l'anonymat, et sur le même schéma que la règlementation relative au don d'organe[74], deviennent des *donneurs de données* de santé pour participer à l'amélioration de leur propre système de soin. Le second à travers la responsabilisation des utilisateurs de données qui favorise un système d'accompagnement puis de coopération, au lieu et place d'un système d'analyse et d'autorisation de traitement, coûteux en termes de moyens institutionnels et de temps, préconisant ainsi davantage la confiance via une technique de formation et de mise en conformité structurelle des organisations quant aux traitement des données médicales. Une orientation du modèle de gouvernance juridique des

[71] H. Caillol, « *Ouverture des données de santé : l'expérience de l'Assurance maladie* », *Op. cit*.
[72] Etalab coordonne au sein du secrétariat général de la modernisation de l'action publique l'action des services de l'État et de ses établissements publics en matière d'ouverture des données.
[73] Voir les détails de l'expérience in H. Caillol, « *Ouverture des données de santé : l'expérience de l'Assurance maladie* », *Op. cit*.
[74] Loi n°76-1181 du 22 décembre 1976 relative aux prélèvements d'organes.

données de santé pour une utilisation en *open data*, initialement gouverné par la protection des droits individuels des personnes concernées, qui tend dorénavant vers une ouverture renforcée et une responsabilisation des utilisateurs ; un choix de gouvernance qui n'est pas sans risques.

B. Les risques d'une gouvernance des données médicales en *open data*

Des données de santé en ressources communes dans un mouvement d'*open data* afin de permettre une effusion scientifique et technologique dans le domaine médical pour l'amélioration des soins des patients, pourquoi pas. Cependant, et malgré les systèmes de sécurité mis en œuvre autour de la protection des données, ce mouvement d'ouverture des données est porteur de risques non négligeables qui doivent être pris en compte.

1) Les risques relatifs à la sécurisation des données en open data

Selon le rapport Villani établit en 2018 « la puissance publique doit [...] amorcer de nouveaux modes de production, de collaboration et de gouvernance sur les données, par la constitution de communs de la donnée »[75]. En dotant la France, et plus globalement l'Union Européenne, d'un système de données de santé commun, « il s'agit ni plus ni moins que de contrer le déploiement intensif de géants états-uniens dans le champ de l'intelligence artificielle et de rouvrir un espace compétitif pour des acteurs de petite ou moyenne taille qui pourront avoir accès à ces données »[76] en évitant ainsi la perte d'autonomie des systèmes nationaux publics face à la montée en puissance continuelle de ces nouveaux acteurs mondiaux du numérique. Des perspectives de système en *open data* qui repose sur un enjeu véritablement central de sécurisation des données dans le cadre de leur traitement, dans la mesure où elle constitue à la fois un outil de conformité pour la protection des données personnelles pseudonymisées, et une obligation légale à la publication de bases de données en *open data* de manière à rendre impossible l'identification directe ou indirecte des personnes concernées[77]. L'anonymisation représente ainsi autant une condition de ploiement des bases de données en *open data* qu'un outil essentiel de confidentialité des données de santé censé garantir la protection de la vie privée des personnes concernées. Cependant, l'anonymisation ne fait pas l'unanimité tant au niveau des techniques utilisées que d'un point de vue de la sécurité des données vis-à-vis des garanties juridiques à apporter. En effet, malgré les apparences et les

---

[75] Voir « *Donner un sens à l'intelligence artificielle, pour une stratégie française et européenne* », par Cédric Villani, sur https://www.aiforhumanity.fr.
[76] V. Peugeot, « *Données de santé : contours d'une controverse* », L'Économie politique, 2018/4 (N° 80), p.30-41.
[77] V. article L. 1461-2 du code de la santé publique.

vertus qui lui sont attribuées, l'anonymisation étant particulièrement complexe et rarement « efficace à 100% » [78], elle véhicule de nombreux risques.

Des critiques d'abord, dans le sens où l'anonymisation des données, qu'elles soient médicales ou non, repose sur des techniques qui consistent à apporter des modifications aux bases de données. Des techniques diverses, tantôt perturbatrices par l'altération de la véracité des données dans le but de limiter le lien entre les données et un individu, celles-ci deviennent suffisamment incertaines pour ne plus être reliées à un individu en particulier, en modifiant directement les données (par exemple, les données ne seront plus au centimètre ou au kilogramme mais à la dizaine près) ; tantôt non perturbatrices, notamment via la technique de k-anonymisation visant à diluer les attributs des personnes concernées en modifiant leur échelle ou leur ordre de grandeur (spécifiées mensuellement plutôt que de manière hebdomadaire, à l'échelle d'une région plutôt que d'une ville, etc). Des techniques nécessaires à la confidentialité des données mais qui, en revanche posent des difficultés scientifiques d'exploitation. Il s'avère effectivement que les données de santé ne sont pas anodines, et des données biaisées ne peuvent amener qu'à des résultats dont le manque de fiabilité peut largement affecter l'utilité médicale attendue des résultats. Ainsi, le dessein étant de permettre le développement de la recherche médicale et l'amélioration de l'offre de soins aux patients, lesdites techniques d'anonymisation appliquées à ce contexte particulier peuvent avoir pour finalités, bien que ce ne soit pas l'objectif de départ, de rendre les données de santé inutilisables, voir aboutir à des résultats potentiellement dangereux pour les patients. Outre les risques pour la sécurité sanitaire liés à l'exploitation de résultats issus de données médicales biaisées, la validité de ces techniques est également critiquée dans un environnement d'insécurité numérique et de massification des données dit de *big data*.

Le premier risque d'un tel système est celui lié à la sécurisation extérieure de la base de donnée. En effet, l'accessibilité à tous dans le cadre de l'*open data* et la dispersion des bases de données, cela « entre les mains de nombreuses entreprises plus ou moins habituées à manipuler des données sensibles, induisent mécaniquement une augmentation des risques de fuite de données, par inadvertance – un sous-traitant technique peu exigeant – ou à la suite de manœuvres crapuleuses »[79]. En matière de cybercriminalité, de nombreuses failles de sécurité

[78] C. Galichet, « *Données personnelles : anonymisation ou pseudonymisation ?* », *Op. cit*.
[79] V. Peugeot, « *Données de santé : contours d'une controverse* », *Op. cit*.

et fuites via des vols de données par exemple sont révélées régulièrement[80], plus ou moins graves, bien que toutes inacceptables, notamment au regard de la sensibilité particulière des données médicales. Ensuite, il est incontestable que « les données de santé sont générées de plus en plus massivement, collectées et stockées par un nombre croissant d'acteurs, utilisées à des fins qui ne cessent de se diversifier (; une) triple envolée qui n'est pas sans soulever de nombreux problèmes » notamment vis-à-vis de la sécurisation des informations personnelles[81]. Une situation que n'arrangera pas l'intention de mise à disposition des données en *open data* combinée à la massification générale des informations numériques. En effet, « les données de santé peuvent provenir de différentes sources, sont de différente nature et sont conservées sur différents supports qui disposent le plus souvent de règles de recueil et d'accès qui leurs sont propres (;) les données médicales, provenant de patients sont conservées dans les dossiers médicaux qui comportent des données biologiques, génétiques, d'imagerie, etc (;) les données relatives au style de vie sont le plus souvent recueillies par voie de questionnaires diffusés au sein du système de santé ou pour les besoins de la recherche »[82], mais encore les données dites de « bien-être » issues des multiples applications ou objets connectés (montres, pèse-personnes et applications de suivi de poids, du sommeil ou d'activité physique, etc) qui fournissent une importante quantité d'informations sur nos habitudes et notre état de santé général. Ainsi « plus le nombre de base de données est important, plus les croisements entre bases sont possibles (voir inévitables), plus les risques de réidentification de données pourtant anonymisées sont grands »[83] ; il s'agit d'un risque d'interopérabilité entre les différentes sources de bases de données à disposition des organismes, produisant ainsi une « cartographie »[84] retraçant toutes les traces de l'individu numérisé. Un croisement d'information permettant ainsi de remonter jusqu'aux personnes concernées[85] et rendant inefficace la technique d'anonymisation censée garantir la confidentialité des données. Éviter la réidentification des individus étant un enjeu central de la mise en *open data* des données de santé, le modèle est interrogé à la fois techniquement et juridiquement aux fins d'apporter des solutions. Parmi les pistes envisagées, il est question d'évaluer l'anonymisation, non plus sur la base d'une distinction stricte entre

---

[80] Voir « Ransomware attack on Hancock Health drives providers to pen and paper », Healthcareitnews, 15 janvier 2018 ; Voir le site https://www.cyberveille-sante.gouv.fr .
[81] V. Peugeot, « *Données de santé : contours d'une controverse* », *Op. cit*.
[82] E. Rial-Sebbag, « *Chapitre 4. La gouvernance des Big data utilisées en santé, un enjeu national et international* », *Op. cit*.
[83] V. Peugeot, « *Données de santé : contours d'une controverse* », *Op. cit*.
[84] S. Paricard, « *Le corps numérique* », Les affres de la qualification juridique, LGDJ, actes de colloque, 2015.
[85] Voir l'expérience vécue par le gouvernement australien en 2016 concernant la réidentification des données de remboursements de dépenses médicales anonymisées de 2,9 millions d'Australiens, couvrant une période allant de 1984 à 2014 in « Research reveals de-identified patient data can be re-identified », phys.org, 18 décembre 2017.

données anonymes et données personnelles, fondée sur l'impossibilité pure et simple de réidentification des personnes concernées ; ni en des termes purement techniques reposant sur les critères d'individualisation, de corrélation ou d'inférence qui, en l'état du contexte actuel de *big data*, compromettraient de manière trop importante l'accès aux ressources et l'utilité sociale sanitaire qui en est attendue. Il s'agirait donc de ne plus se référer à une condition d'impossibilité stricte de réidentification, mais plutôt d'adopter une approche fondée sur le risque de réidentification. Une vision qui défend l'idée « d'une révision des mécanismes de protection des données qui irait au-delà de la distinction entre données personnelles et données anonymes, pour se diriger vers une évaluation plus quantitative de la probabilité de réidentification »[86] dont l'objectif serait d'opérer un calcul de proportionnalité entre l'utilité du traitement envisagé et le respect de la vie privée des personnes concernées. Une piste intéressante dans la mesure où la juxtaposition du *big data* et de l'*open data* met en péril la garantie de sécurité des données de santé. Une approche par le risque permet effectivement d'adopter une démarche par la limitation des atteintes graves pour la vie privée des personnes. Cela dit, cette approche par le risque porte bien son nom, et comporte des risques, qu'ils soient mineurs ou majeurs pour la sécurité des données. De plus, l'exigence de consentement répondant déjà à de nombreuses dérogations en la matière, notamment vis-à-vis des services publics du système de santé qui fournissent majoritairement les bases de données mises en *open data*, l'individu n'est à nouveau pas consulté sur le calcul du risque. Il s'agit donc de déterminer avec soin les critères d'évaluation déterminant à quel moment l'atteinte à la vie privée est acceptable vis-à-vis de l'utilité du traitement, mais également d'opérer des contrôles de cette décision subjective, voir arbitraire pour apporter des garanties suffisantes. Une problématique d'autant plus intéressante confrontée aux autres risques inhérents à l'*open data* tel qu'établit.

2) Les risques relatifs au choix d'un modèle de gouvernance en open data

Au regard des difficultés vues précédemment à assurer la sécurité des données de santé dans un modèle en *open data* par des moyens techniques, le système de gouvernance doit nécessairement mettre en œuvre des moyens juridiques forts. Ainsi, si la sécurisation des données mérite encore de s'améliorer, « sans ouvrir le débat sur la propriété des données, qu'il n'est pas utile de chercher à résoudre pour se positionner par rapport à *l'open data*, il est fondamental de comprendre que la reconnaissance du droit et de l'intérêt à accéder à ces

[86] H. Tanghe, P-O Gibert, « *L'enjeu de l'anonymisation à l'heure du big data* », Revue française des affaires sociales, p. 79-93.

données est davantage une question de responsabilité que de sécurité »[87]. Et dans ce domaine, la tendance depuis quelques années n'est plus tellement celle de la responsabilité juridique traditionnelle, qui n'intervient que dans un second temps, mais davantage celle de la *responsabilisation* sociale ; un mouvement auquel le modèle de gouvernance des données, qu'elles soient de santé ou non, n'échappe pas. Si le concept connaît un essor particulier en matière environnementale, notamment via le développement de la responsabilité sociale des entreprises, en réalité aucun domaine juridique n'est exempté, pas même les matières pénales. Ainsi, la gouvernance par la responsabilisation doit s'entendre dans un champ d'application conjuguant différentes composantes tenant respectivement à la responsabilisation en tant qu'intégration de *mœurs* éthiques, via une gouvernance coopérative prenant en considération les différentes parties prenantes ainsi que, le cas échéant, dans sa définition judiciaire ayant vocation à engager la responsabilité des acteurs concernés[88]. Dans ce sens, « le principe de la responsabilité qui consiste à répondre de ses actes devant l'autre connait des adaptations très diverses, tantôt comme responsabilité juridique exerçant une contrainte normative coercitive, tantôt comme mécanisme économique, tantôt comme impératif moral, tantôt comme mécanisme de gouvernabilité »[89]. Un parti pris intéressant qui s'articule autour d'une gouvernance par le risque et l'autonomie des acteurs auquel ni le RGPD, ni les loi nationales en matière de traitement des données ne font exceptions, mais qui est vecteur de risques. Il s'avère que « traditionnellement, le mécanisme le plus protecteur repose sur la technique de l'autorisation préalable consistant à ne permettre l'usage d'une liberté qu'après avoir obtenu l'aval de l'administration qui dispose d'un véritable pouvoir de censure (alors qu'à) l'opposé, le législateur peut faire le choix de ne permettre à l'administration qu'une intervention a posteriori consistant à sanctionner un détournement de l'usage légal des données médicales (, mais encore faire le choix d''un) régime intermédiaire qui consiste, lui, à instaurer un système déclaratoire obligeant l'administré à s'identifier afin de faciliter un éventuel contrôle ultérieur »[90]. Ainsi parmi les options possibles, le RGPD[91] et les dispositions françaises modifient l'approche de la protection des données personnelles, en passant d'une méthode

---

[87] M. Léo, « Patient connecté et données de santé : les vrais risques », I2D – Information, données & documents, 2016/3 (Volume 53), p. 65-66.

[88] C. Darnault, « *Les pme face au contentieux économique - Essai de guide* pratique », Sciences de l'Homme et Société, Aix-Marseille Université (AMU), 2018. Français.

[89] O. Costa, N. Jakbo, C. Lequesne, P. Magnette, « *La diffusion des mécanismes de contrôle dans l'Union Européenne : vers une nouvelle forme de démocratie ?* », Revue française de science politique, 2001, Vol. 51, n° 6, p.859.

[90] E. Pechillon, « *L'accès ouvert aux données de santé : la loi peut-elle garantir tous les risques de dérives dans l'utilisation de l'information ?* », L'information psychiatrique, 2015/8, Vol. 91, p. 645-649.

[91] Règlement (UE) 2016/679, *ibid* (article 24 et 74).

reposant en grande partie sur l'existence de formalités préalables, pour une logique d'*accountability*, fondée sur les concepts de conformité et de responsabilité. Ce qui signifie concrètement qu'il ne s'agit plus « seulement pour les acteurs d'effectuer des demandes d'autorisation auprès de la CNIL, ils doivent aussi s'assurer, au moment du montage d'un projet qui implique un traitement de données personnelles, puis tout au long de la vie du projet, en pleine responsabilité, du respect des principes de protection des données et, surtout, ils doivent à tout moment être en mesure de le démontrer en cas de contrôle de la CNIL »[92].

Une politique de responsabilisation qui comporte certes des avantages en termes de facilité d'accès aux données et d'accroissement des traitements potentiels au nom de la recherche médicale. Néanmoins, miser sur la bonne foi, la conformité et la responsabilisation des acteurs n'est pas sans risque dans la mesure où cette politique de gouvernance, bien qu'elle ait permis des avancées innovantes dans d'autres domaines, a tout aussi largement démontré ses limites. Une telle politique de responsabilisation par la conformité dans un contexte d'*open data*, si l'objectif reste toujours la protection des droits des personnes, nécessite d'employer des moyens colossaux en terme de contrôle. En effet, envisager l'*open data* via un régime de conformité peut donner des vertiges dans un environnement où la lisibilité et la compréhension des normes ne sont pas toujours des plus aisées pour tout un chacun, en particulier pour les petites structures, ou lorsque les règlementations concernant la protection des données sont appliquées de manière purement mécanique dans le cadre de contrats de masse dans un marché du numérique dominé par des mégastructures internationales. Celles-ci même qui, par certains aspects, outrepassent les souverainetés nationales respectives qui voient leurs différents modèles de gouvernance des données s'affronter au lieu de s'allier. En effet, si via des demandes d'autorisation préalable, l'autorité de régulation conserve un certain contrôle de la légalité des traitements, dans le cas d'un accès libre aux données pour tous, la problématique est bien différente. S'en remettre ainsi à la conformité doit s'accompagner d'un système de contrôle considérable, disposant de moyens à la hauteur de l'opération, afin de dissuader tout manque de conformité ou d'abus. Sur ce point, il semblerait que les régulateurs s'appuient sur un régime de sanction qui permette de dissuader pour mieux prévenir les risques de violations à la règlementation. Ainsi, « l'augmentation spectaculaire des seuils d'amendes pouvant être imposés par la CNIL pour violation de la protection accordée aux données personnelles, couplée au renforcement des contrôles » laisse supposer que « le respect des dispositions

[92] F. Lesaulnier, « *Recherche en santé et protection des données personnelles à l'heure du RGPD* », Op. cit.

légales devient un enjeu économique majeur pour les responsables du traitement ». En effet, « bien que la CNIL prenne en compte la taille de la structure, la gravité de la violation, et son caractère répété ou non, pour déterminer le montant des sanctions, sa récente amende de 50 millions d'euros contre Google montre qu'elle entend faire pleinement usage des latitudes qui lui sont octroyées et qui devrait servir d'exemple pour intimer aux géants du net une mise en conformité réelle et sérieuse »[93]. Cependant, bien qu'une célèbre maxime rappelle que l'on peut obtenir beaucoup plus avec un mot gentil et un revolver, qu'avec un mot gentil tout seul[94], le risque pesant sur les contrevenants suffira-t-il pour garantir les droits des personnes ? L'autorité de régulation disposera-t-elle de suffisamment de revolver pour les pointer de manière efficiente sur les organismes traitant les données de santé ? Avec des données accessibles à tous, la CNIL pourra-t-elle garantir l'effectivité des droits à tous les niveaux de circulation organisationnelles ou géographiques des données médicales ? Tant d'interrogations qui restent en suspens dans un modèle de gouvernance pour des données de santé en *open data* qui évolue progressivement, tant au niveau technique que juridique, pour concilier au mieux l'utilité sociale du traitement des données médicales et le respect de la vie privée des personnes concernées. Cela pour définir au mieux les modalités d'un système de gouvernance d'accès responsable des données de santé, à dissocier d'une captation sans conditions, déjà soumis à des difficultés techniques propres aux technologies d'intelligence artificielle en elles-mêmes. En effet, bien que ces problématiques ne soient pas propres aux traitements des données de santé mais communes à l'ensemble des technologies issues des systèmes statistiques et algorithmiques, elles ne doivent pas être éludées pour autant. En dehors des failles techniques et juridiques de l'ouverture des données de santé, se pose également la question des biais intrinsèques de leurs traitements. En sus d'une politique de risque et de conformité au régime de protection des données, la communauté scientifique met en lumière les biais sociaux que comportent les données elles-mêmes, en responsabilisant également les acteurs du traitement des données au respect d'un principe de précaution en les incitant à la vigilance et à une implication grandissante dans la transparence des systèmes algorithmiques. Il s'avère que « les algorithmes s'immiscent de plus en plus dans notre quotidien à l'image des algorithmes d'aide à la décision (algorithme de recommandation ou de scoring), ou bien des algorithmes autonomes embarqués dans des machines intelligentes (véhicules autonomes) » ; ainsi, s'ils sont largement « déployés dans de nombreux secteurs et industries pour leur efficacité, leurs

[93] I. Fréret Iris, « *Une responsabilité accrue pour les acteurs du RGPD* », I2D – Information, données & documents, p. 21-24.
[94] E. Netter, « *Regards sur le nouveau droit des données personnelles* », Op. cit.

résultats sont de plus en plus discutés et contestés (dans la mesure où) ils sont accusés d'être des boites noires et de conduire à des pratiques discriminatoires liées au genre ou à l'origine ethnique »[95]. Matériellement, les systèmes apprenant sur la base des données fournies, elles-mêmes issues de nos sociétés avec tous ce qu'elles peuvent comporter d'inégalitaire et de discriminatoire, cumulées aux choix opérés par leur utilisateurs, proposent des résultats constituant le reflet des échantillons sur lesquels ils sont construits, autrement dit posant des difficultés au regard des droits fondamentaux. Lorsqu'on voit que certains algorithmes confondent encore des individus afro-américains avec des gorilles[96], que le taux de succès des systèmes de reconnaissance de la parole fonctionne de manière discriminatoire en fonction des langages ou des tonalités sous-représentées[97], que des systèmes de justice prédictive condamnent davantage certaines catégories de population[98], il est nécessaire d'appréhender le traitement des données de santé, mais également les résultats issus des algorithmes qui en apprennent, avec une attention particulière. Si les règlementations en matière de protection des données visent principalement la protection de la vie privée pour des raisons évidentes, elles ne s'arrêtent pas là. En effet, il y a plusieurs décennies déjà, la Convention 108 mettait en garde sur les difficultés soulevées par l'exercice d'une complète liberté de traiter les informations qui, dans certaines conditions, risque de nuire à la jouissance d'autres droits fondamentaux que celui de la vie privée tels que les droits à la non-discrimination et à un procès équitable[99]. Un droit fondamental à la protection des données qui ne doit donc pas s'entendre simplement de la sécurisation des données mais doit être envisagé par le système de gouvernance des données, qu'elles soient de santé ou autres, de manière global en tant que droit fondamental qui ne doit pas porter atteinte à la vie privée des individus, entraîner ou reproduire des discriminations, ou engendrer des limites à la liberté d'expression.

---

[95] P. Bertail, D. Bounie, S. Clémençon, P. Waelbroeck, « *Algorithmes: Biais, Discrimination et Équité* », 2019.
[96] O. Robillard, « *L'intelligence artificielle distingue mal les femmes noires* », L'ADN Innovation, 12 février 2018, https://www.ladn.eu/tech-a-suivre/ia-machine-learning-iot/sexisme-racisme-quand-lintelligence-artificielle-se-trompe/ (consulté le 15 avril 2020).
[97] Y. Demeure, « La reconnaissance vocale automatisée coupable de discrimination raciale ? », SciencePost.fr, 7 avril 2020, https://sciencepost.fr/la-reconnaissance-vocale-automatisee-coupable-de-discrimination-raciale/ (consulté le 15 avril 2020).
[98] J.M. SAUVE, Colloque organisé à l'occasion du bicentenaire de l'Ordre des avocats au Conseil d'État et à la Cour de cassation, Cour de Cassation, 2018 ; The United States Solicitor General also filed a Brief to defend COMPAS. See Brief for the United States as Amicus Curiae, Loomis v. Wisconsin, 137 S. Ct. 2290 (2017) (No. 16-6387), 2017 WL 2333897.
[99] J-M Deltorn, « *La protection des données personnelles face aux algorithmes prédictifs* », Op. cit.

En conclusion, si l'on analyse le système de gouvernance juridique des données à caractère personnel de ses débuts, il y a maintenant déjà plus de quarante ans, les desseins gouvernementaux ont évolué, prenant également en considération l'explosion du développement des nouvelles technologies et des systèmes d'intelligence artificielle. Alors que la loi *Informatique et Liberté* de 1978 établissait que l'informatique dans son ensemble, et le traitement des données à caractère personnel ne devait porter atteinte ni à l'identité humaine, ni aux droits de l'homme, des principes fondamentaux réitérés au sein du RGPD ; l'ambition d'un système de données anonymisées en *open data* et les difficultés liées au développement des technologies (*big data*, identité numérique, etc) viennent mettre à mal ces fondations juridiques centrées sur la protection individuelle. Dans ce contexte, et notamment dans le cas particulier du traitement des données de santé en raison de l'intérêt général que véhicule leur exploitation, ces informations ne peuvent plus simplement être considérées comme privées, propres aux personnes qu'elles concernent, mais doivent, au nom du bien commun, s'inscrire dans une logique publique collective, où elles sont détachées des individus qu'elles identifient, en tant que ressources communes anonymes, n'appartenant plus aux personnes de manière individuelle, mais étant à disposition de tous. L'interdiction de principe du traitement des données de santé, considérées comme particulièrement sensibles, n'est au final qu'une protection en demi-teinte. Cela dans la mesure où l'interdiction qui semble être une protection stricte et radicale des individus est en réalité une exception à toutes les exemptions de traitement prévues par le règlement lui-même d'une part, et par la marge de manœuvre laissées aux États membres quant à la règlementation du traitement des données dans le domaine spécifique des informations médicales. Un point qui est en réalité laissé à l'entière discrétion des systèmes nationaux plus qu'une simple marge de manœuvre pour des spécificités nationales. D'autant plus que le parti pris de l'*open data* est l'exploitation de données anonymisées qui, contrairement à d'autres techniques telle que la pseudonymisation, n'intègrent pas le champ d'application du RGPD dès lors qu'elles ne permettent plus théoriquement la réidentification des individus. Ainsi dans les faits, le changement s'opère d'un système de gouvernance construit autour de la protection des individus et du respect de leur droit fondamental à la vie privée, à une gouvernance par le risque où l'objectif est la possibilité de traitement des données de santé, au nom de l'utilité sociale et de la démocratie sanitaire, qui doit être réalisée, dans la mesure du possible, via une politique du moindre risque d'atteinte aux droits des personnes. Un système de gouvernance juridique qui n'est plus défini par la protection de données à caractère personnel dans le cadre de traitements autorisés ponctuellement dans des conditions juridiques bien spécifiques ; mais davantage d'après un modèle axé autour d'une mise à disposition

totalement libéralisée en *open data* de données anonymisées qui doivent être exploitées de manière proportionnée et responsable pour limiter les risques de réidentification et d'atteinte aux droits des personnes concernées. Il s'agit là d'une prise de position bien différente. Et bien que les nombreux arguments mis en avant soient au nom de l'intérêt général, les données sont considérées comme un enjeu de premier plan pour la recherche dans le domaine de la santé, constituant le matériau de base pour le développement de la recherche médicale, représentant de nombreuses opportunités individuelles pour l'amélioration des soins des patients et collectives pour la démocratie sanitaire, le parti pris bien qu'intéressant et porteur d'espoir, n'en reste pas moins discutable au regard des risques qu'il engendre. Il reste à compter sur une application consciencieuse du principe de proportionnalité qui soit véritablement responsable, et cela dans tous les sens du terme ; tant d'un point de vue d'une évaluation des risques qui pèsent sur les droits des personnes au regard de l'apport du traitement vis-à-vis de l'utilité sociale, que dans la définition judiciaire du terme via des mesures de contrôle et des sanctions attingentes aux potentiels traitements qui s'avèreraient manifestement disproportionnés. Le véritable enjeu de ce modèle de gouvernance du traitement des données de santé anonymisées en *open data* réside ainsi dans l'amélioration des outils techniques et des principes juridiques définissant concrètement l'anonymisation. Mais pas seulement, il est également nécessaire de prévoir la mise en œuvre de politiques publiques permettant la lisibilité des règlementations, un accompagnement pédagogique des organisations et de leur responsable de traitement, devenu l'acteur clef du traitement des données, aux fins qu'ils puissent appréhender au mieux la conformité de leurs systèmes. Il s'agira également de prévoir les moyens nécessaires, que ce soit au niveau financier qu'en termes de ressources humaines, pour les institutions de régulation en la matière afin de mettre en place un contrôle véritablement efficient qui, dès lors que les données sont mises à disposition en *open data*, constitue le dernier garde-fou aux violations potentielles des droits des personnes. Cependant, et pour terminer sur une note davantage optimiste, le développement des moyens technologiques et l'amélioration des algorithmes d'apprentissage profond est tel que nous ne sommes pas en mesure d'imaginer quels seront les outils à notre disposition demain. La communauté scientifique étant déjà à l'œuvre pour travailler sur les difficultés techniques liées au traitement des données, notamment d'anonymisation, et aux biais potentiels issus des systèmes algorithmiques, pour le développement de technologies médicales qui soient au service de chaque citoyen, au service du bien commun dans le respect des droits fondamentaux.