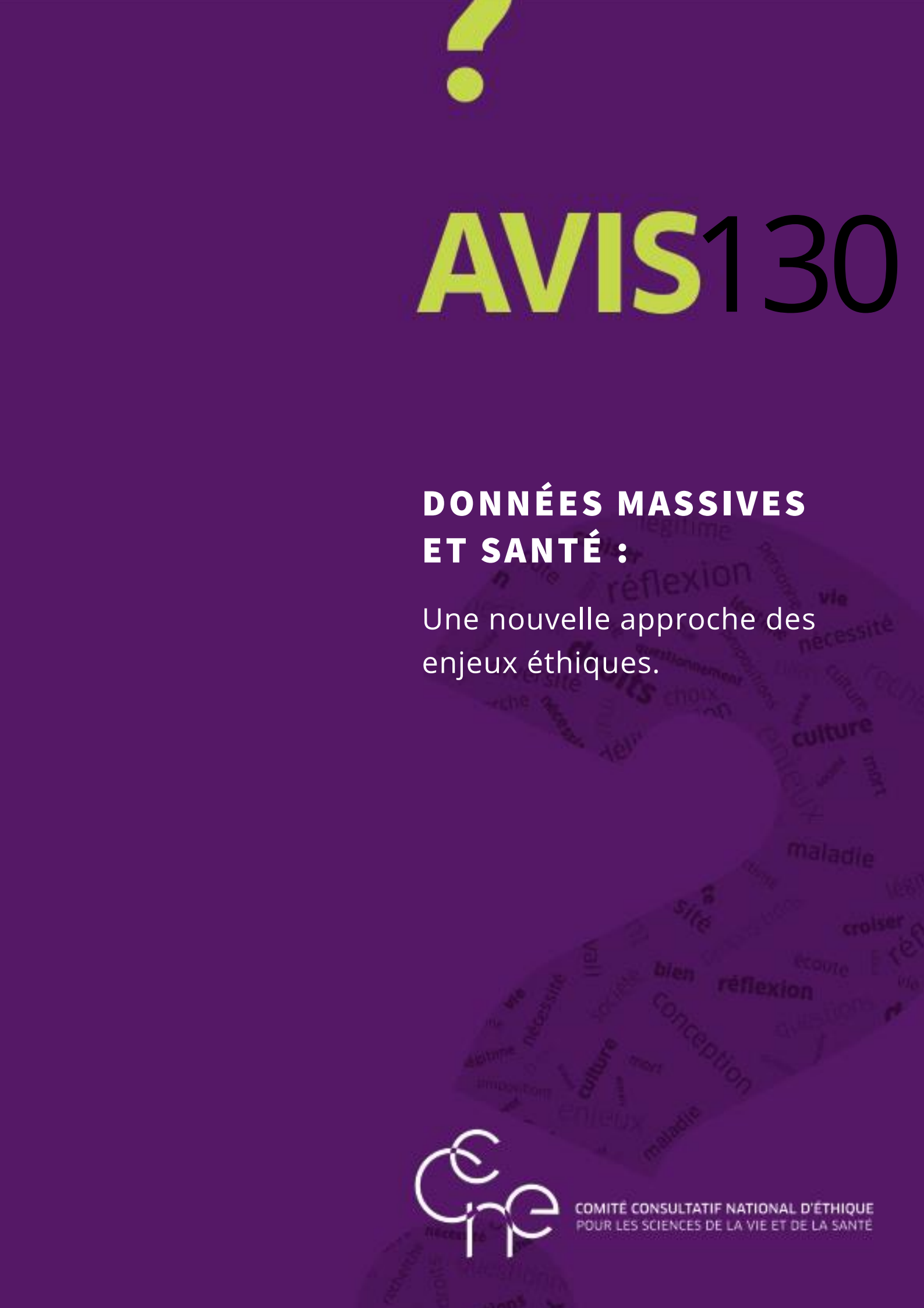
AVIS 130

DONNÉES MASSIVES ET SANTÉ :

Une nouvelle approche des enjeux éthiques.



COMITÉ CONSULTATIF NATIONAL D'ÉTHIQUE
POUR LES SCIENCES DE LA VIE ET DE LA SANTÉ

AVIS 130

AVIS 130

DONNÉES MASSIVES ET SANTÉ :

UNE NOUVELLE APPROCHE DES ENJEUX ÉTHIQUES

Avis rendu public le 29 mai 2019

AVIS 130

TABLE DES MATIERES

AVIS 130

RÉSUMÉ

Poursuivant une réflexion antérieure sur les questions que posent le recueil de données personnelles et les nouvelles possibilités de leur utilisation dans le domaine des sciences de la vie et de la santé, et à la suite des États généraux de la bioéthique[•], le CCNE consacre son Avis 130 aux enjeux éthiques que soulève l'utilisation des « données massives »[••] (*Big Data*). Dans le contexte de mutations technologiques et culturelles accélérées liées au traitement de ces « données massives », le CCNE souligne combien l'accumulation massive de données issues de personnes, comme la capacité accrue qu'a le traitement de ces données de produire de la valeur, nécessitent débat et réflexions éthiques.

Cet avis :
- énonce des principes éthiques *communs* à tous les contextes d'utilisation des données massives dans le champ de la santé (*chapitre 2*) ;
- identifie des enjeux éthiques *spécifiques* aux situations de soin, de recherche, de gestion des soins, ou de vie personnelle qui utilisent les données massives (*chapitre 3*) ;
- propose 12 recommandations indispensables au respect des principes éthiques fondamentaux permettant, sans les freiner, le développement des technologies fondées sur les données massives.

Les mutations induites par les données massives

Le CCNE estime qu'on ne saurait adopter une position hostile à ces nouvelles technologies numériques à raison des risques dont elles sont porteuses, car il serait contraire à l'éthique de ne pas favoriser leur développement si elles peuvent bénéficier à la santé de tous et aider à la rationalisation des coûts. Mais le CCNE rappelle que les grands principes qui fondent l'éthique médicale – respect de la personne (incluant le respect

[•] Voir l'Avis 129 « *contribution à la révision de la loi de bioéthique* », ainsi que le rapport « *numérique et santé. Quels enjeux éthiques pour quelles régulations ?* » commandé par le CCNE et consacré à l'intelligence artificielle.

[••] Par « données massives », on désigne la disponibilité, soit d'un nombre important de données, soit de données de taille importante, que seuls les outils du numérique alliant l'algorithmique à la puissance de calcul des ordinateurs permettent de traiter efficacement. Outre *le changement d'échelle*, qui fait que seule la machine - et non plus l'humain - est capable d'assurer la collecte, la conservation et l'analyse des données, celles-ci se caractérisent principalement par trois propriétés : leur *pérennité* (elles peuvent être copiées et réutilisées indéfiniment) ; leur *diffusion* dans le temps et l'espace, qui permet leur partage rapide et sans distinction de frontières ; la *génération de données secondaires, nouvelles informations* obtenues par le traitement et le croisement des données initiales avec d'autres sources, qui fait de ces données un matériel exploitable bien au-delà des finalités du recueil initial.

de son autonomie), justice, pertinence et bienfaisance (incluant ici l'obligation de non-nuisance) – ne doivent pas être affaiblis par le développement des technologies numériques.

Les mutations induites par les données massives modifient notre façon de formuler des hypothèses de recherche, d'acquérir des savoirs et de les utiliser pour une prise de décision intéressant l'individu et la collectivité. Elles induisent des changements dans nos représentations et bouleversent notre façon de considérer la mesure de nos paramètres de santé et de bien-être. Ces données personnelles tendent à devenir un outil relationnel dynamique et un instrument d'autonomie visant à un meilleur contrôle personnel de notre état de santé.

Le CCNE appelle à la vigilance pour que les personnes qui n'ont pas aisément accès aux technologies du numérique bénéficient aussi des avancées dans le domaine de la santé et ne subissent ni stigmatisation ni discrimination dans leur accès aux soins. (*Recommandation 9*).

De ces mutations, le CCNE tire deux constats :

- Le premier est que toute donnée primaire issue d'une activité humaine – même sans lien apparent avec la santé – peut contribuer – par son croisement avec d'autres données qui ne lui sont pas liées – à la création d'une information nouvelle relative à la santé d'une personne. Une donnée de santé ne peut plus se limiter aux seules données personnelles recueillies dans le cadre d'une prise en charge médicale (mesures biologiques, caractéristiques génomiques, données cliniques, etc.).

- Le second est que les questionnements éthiques se développent dans un contexte instable en raison d'innovations technologiques extrêmement rapides et rendu complexe par la diversité des situations dans lesquelles les données relatives à la santé sont recueillies et traitées, ce qui nécessite une vigilance continue et une évaluation périodique de la mise en œuvre effective des dispositifs de protection (*recommandation 2*) ; la réflexion éthique doit prendre en compte le fait que certaines innovations peuvent ne pas avoir pour finalité le soin, mais l'exploitation d'un marché se présentant comme relevant du bien-être.

Une perception nouvelle des enjeux éthiques et des propositions de réponses adaptées

Le CCNE montre qu'une des caractéristiques des données massives relatives à la santé est d'effacer les distinctions sur lesquelles repose la mise en œuvre des principes éthiques qui fondent la protection des droits individuels dans le champ de la santé. Ainsi, la séparation s'estompe entre *vie privée et vie publique* par la possibilité de croiser des données sans lien les unes avec les autres, mais aussi parce que notre

représentation de l'intime change. Le *rapport entre l'individuel et le collectif* évolue : l'autonomie de chacun s'accroit, mais la connaissance très précise des personnes et de leur état de santé induit le risque d'un profilage ; celui-ci met en cause la protection de la vie privée et peut aboutir à la stigmatisation de personnes ou de groupes. Celle-ci menace la vie privée, mais aussi les principes de solidarité et d'équité qui fondent notre système de santé ; *soin et commerce* deviennent plus difficiles à distinguer, conséquence de la transformation du soin et du marché de la santé.

Le CCNE remarque, en outre, que la notion même de consentement libre et éclairé, exigé *a priori* comme assurant la protection d'une personne face à une décision la concernant, est remise en cause par les conditions mêmes d'exploitation des données massives (finalités incertaines et incompréhension, voire inaccessibilité, du processus d'analyse). La nécessité d'une protection de la personne doit être réaffirmée et ses modalités doivent être redéfinies, afin d'éloigner la menace d'une société de surveillance et de contrôle par de multiples opérateurs agissant à des fins diverses. Même si le consentement reste l'un des fondements majeurs de licéité du traitement des données personnelles, le RGPD[¶], qui se veut ferme sur ses principes mais pragmatique, prend acte de ce que cette exigence n'est pas réalisable dans tous les cas, notamment dans l'hypothèse d'une réutilisation des données ; il reconnaît comme licites d'autres modalités de protection lorsque sont poursuivies des finalités d'intérêt général. Le CCNE prend acte de ce passage progressif d'une volonté de contrôle exhaustif *a priori* par l'individu à une logique d'intervention et de contrôle *a posteriori* fondée sur une recherche d'intelligibilité et de responsabilisation. Cette logique exige une loyauté de comportement des responsables du traitement, une transparence de leurs processus, et la possibilité de contrôler leurs possibilités d'accès aux données et leur démarche déontologique. Le CCNE réaffirme l'importance d'une gouvernance identifiée et d'engagements qui doivent pouvoir être vérifiés. L'information précise et loyale, adaptée aux différents contextes d'utilisation, devient un critère éthique majeur. ***(Recommandations 1 à 3).***

Le CCNE estime qu'une *garantie humaine* des différentes étapes du processus de l'analyse des données est fondamentale. Elles sont, en effet, la « matière première » nécessaire à la conception d'algorithmes d'aide à la décision, qui prennent une part toujours plus importante dans l'exercice médical et la définition des politiques de santé. Une « garantie humaine » est donc essentielle pour répondre de la rigueur méthodologique des différentes étapes du processus de recueil et de traitement des données que sont : (i) la qualité et l'adéquation des données sélectionnées pour entrainer les algorithmes ; (ii) l'adéquation du choix des traitements algorithmiques à la question posée ; (iii) la vérification sur un jeu de données indépendantes de la robustesse et de l'exactitude du résultat donné par l'algorithme. Sa mise en œuvre exige des ac-

[¶] Règlement général sur la protection des données, règlement européen entré en vigueur dans les États de l'Union européenne le 25 mai 2018.

tions fortes dans trois domaines : la formation, une évaluation qualitative des sites, applications et objets connectés, une recherche scientifique de haut niveau. ***(Recommandations 4, 5, 6, 8).***

Une spécificité des questionnements éthiques selon les contextes

À la diversité des sources de données primaires correspond une diversité tout aussi importante des champs de leur utilisation dans le domaine de la santé (soin, conception de nouveaux médicaments ou dispositifs médicaux, amélioration des connaissances, gestion des essais cliniques, meilleure performance économique, amélioration de la santé publique, objectif commercial). Quelques enjeux éthiques spécifiques à trois de ces situations sont explicités :

- Le CCNE rappelle que la relation de soin se fonde sur une relation humaine directe, basée sur la confiance et un ensemble de décisions véritablement partagées entre le médecin et le patient, même si l'informatisation des systèmes de soin est maintenant généralisée. Trois principes éthiques peuvent être fragilisés par l'utilisation des données massives : *le secret médical*, par la multiplication des informations partagées et échangées entre divers acteurs, dont certains ne relèvent pas du milieu médical ; *la responsabilité de la décision médicale,* par le risque d'automatisation que crée la multiplication des logiciels algorithmiques ; *la relation personnelle* entre le médecin et son patient, qui est menacée d'appauvrissement avec les innovations attendues du traitement des données massives, le patient risquant d'être réduit à un ensemble de données à interpréter, semblant rendre inutile son écoute. Le CCNE rappelle que les technologies du numérique doivent rester une aide à la décision et estime que le temps ainsi gagné devrait être mis à profit pour libérer du temps d'écoute, d'échange, de prise en compte personnelle du patient. (***Recommandation 7***).

- Dans le cadre des protocoles de recherche**,** le CCNE rappelle que l'enjeu éthique principal est de trouver le bon équilibre entre le risque d'une sous-exploitation des données limitant des recherches menées dans l'intérêt général et celui d'un partage des données trop large et insuffisamment contrôlé mettant en cause les droits fondamentaux de la personne. Le CCNE souligne que chacun des acteurs intervenant dans ce champ appréhende différemment les questionnements éthiques auxquels il est confronté. Le titulaire des données, par exemple, ne disposera que de peu d'informations précises sur ces recherches au moment du recueil et ne tirera pas nécessairement de bénéfice pour lui-même des résultats ; la confiance accordée repose, dans ce contexte, avant tout sur le processus de gouvernance et la manière dont l'accès aux données est contrôlé. La qualité d'information délivrée au titulaire des données est un critère majeur de cette qualité de la gouvernance. ***(Recommandations 11, 12).***

Le CCNE analyse plus spécifiquement le cas des données génomiques – qui se distinguent entre autres par les caractères identifiant et, pour partie, prédictif, de la séquence génomique pour l'individu et ses apparentés – ainsi que par les questionnements liés à la constitution de grandes banques de données génomiques et à leur partage dans le cadre de la recherche. Si leur exploitation a permis de spectaculaires avancées des connaissances et une notable amélioration de la prise en charge des patients, elles illustrent également les risques induits par l'exploitation des données massives : diffusion incontrôlée, possibilité d'identification, de perte de confidentialité, et donc de sécurité. Le CCNE recommande le maintien de la réglementation spécifique et, en particulier, la nécessité d'un consentement exprès et la prise en compte de la parentèle ; il relève le risque accru de découvertes incidentes, qui pose un problème éthique spécifique, et incite à la vigilance quant aux biais possibles de sélection des populations étudiées (***recommandation 4***).

- Les ***réseaux sociaux, les applications et objets connectés, les plateformes internet*** de partage d'informations de santé destinées aux patients, sont devenus une source très importante de données, précieuse notamment pour le suivi médical, la pharmacovigilance, mais aussi pour la recherche ou les politiques de prévention ou de veille sanitaire. Toutefois, quand elles sont recueillies hors parcours de soin, la diffusion et l'utilisation de ces données portant sur des événements de leur vie réelle fragilisent la protection des patients, notamment leur droit au respect, tant d'une information loyale que des limites de leur consentement à la diffusion, à l'hébergement et à la réutilisation de ces données potentiellement relatives à la santé.

Le CCNE note enfin que les technologies reposant sur la numérisation mettent en évidence la nécessité – pour qu'un pays garde la maîtrise de sa politique de santé et de sa capacité à l'innovation scientifique et médicale – d'affronter les défis technologiques du stockage et de la sécurité, ainsi que d'assurer un haut niveau scientifique et technologique pour l'exploitation des données. ***(Recommandation 10).***

INTRODUCTION

Le 25 janvier 2017, la ministre des Affaires sociales et de la Santé a saisi le CCNE d'une demande d'avis induite par le développement de la médecine de précision et l'utilisation de plus en plus fréquente et habituelle des *Big Data*[1]. Relevant qu'ils offrent des opportunités majeures d'amélioration de la qualité et de la sécurité des soins, elle soulignait qu'ils posent néanmoins des questions éthiques, qui portent en particulier sur les modalités de l'information et du consentement des personnes concernées, ou de l'expression de leur droit d'opposition, en fonction notamment des finalités de l'utilisation des données générées. Était également évoquée la nécessaire protection de ces données. La ministre s'interrogeait sur la recherche d'un juste équilibre entre, d'une part, les opportunités de développement, au profit de la collectivité, des patients et de leur parentèle, et, d'autre part, la nécessité de protéger la vie privée des citoyens. Cette saisine s'intègre donc dans le mouvement général d'interrogation sur les enjeux éthiques majeurs que posent les activités de recueil et de traitement des données massives dans le domaine de la santé.

Le CCNE s'était déjà auparavant intéressé aux questions posées par le recueil et les nouvelles possibilités d'utilisation de données personnelles dans le domaine des sciences de la vie et de la santé[2]. Notamment, l'avis n° 124 du 21 janvier 2016, qui porte spécifiquement sur la génétique, s'est intéressé aux problèmes posés par l'information contenue dans la séquence du génome, ainsi qu'aux procédures de consentement, notions qui sont au cœur de notre étude[3].

---

[1] Ce terme de *Big Data* sera remplacé dans la suite de l'avis par celui de données massives qui correspond à la traduction française la plus usuelle.

[2] Avis n° 46 du 30 octobre 1995 sur génétique et médecine : de la prédiction à la prévention.

Avis n° 77 du 20 mars 2003 sur les problèmes éthiques posés par les collections de matériel biologique et les données d'information associées.

Avis n° 91 du 16 février 2006 sur les problèmes éthiques posés par l'informatisation de la prescription hospitalière et du dossier du patient.

Avis n° 98 du 26 avril 2007 sur biométrie, données identifiantes et droits de l'homme.

Avis n° 104 du 29 mai 2008 sur le « dossier médical personnel » et l'informatisation des données de santé.

Avis n° 116 du 23 février 2012 sur les enjeux éthiques de la neuro-imagerie fonctionnelle.

Avis n° 124 du 21 janvier 2016 : Réflexion éthique sur l'évolution des tests génétiques liés au séquençage de l'ADN humain à très haut débit.

[3] S'interrogeant sur le défi de la gestion des données (pages 19 et 20), cet avis 124 annonçait le présent avis dans les termes suivants : « *La réflexion sur la gestion des masses importantes de données (« big data ») est de plus en plus présente au niveau international. Le Conseil de l'Europe, par exemple, réfléchit aux questions éthiques posées par une médecine « big data » connectée, dont le sujet déborde très largement des seules données génétiques issues du séquençage d'ADN humain à très haut débit. Une des caractéristiques de ce mouvement est que les opérateurs principaux en sont de grands groupes (Google, Amazon, Facebook et Apple, par exemple) qui n'ont pas de tradition de travail avec les médecins et les biologistes. Une autre spécificité est que la puissance requise pour l'analyse et le stockage de ces données sélectionne un petit nombre d'entreprises qui, seules, sont en capacité de les assurer, créant à la fois une concentration de pouvoir qui peut aller jusqu'à apparaître hégémonique, et une forme d'appropriation de ces données qui est, de fait, contradictoire avec le fondement et la justification de ces analyses de masses importantes de données de santé, à savoir le libre partage des informations et leur ouverture en accès libre. Cette question est au cœur des préoccupations du CCNE, et sera reprise dans une étude indépen-*

La révolution numérique s'étend à tous les domaines et elle partage, avec l'économique et l'environnement, la caractéristique d'être un champ transversal qui innerve tout le champ de la santé. Le CCNE s'est engagé dans une réflexion sur les enjeux éthiques de cette révolution numérique, dont témoigne le rapport numérique et santé missionné par le président du CCNE, paru en novembre 2018[4]. Le présent avis poursuit ce travail en l'approfondissant sur la question spécifique de l'utilisation des *données massives* en matière de santé. Ces deux documents s'intègrent dans une réflexion plus vaste que mène le CCNE sur les questions bioéthiques que posent les transformations du système de santé, caractérisé par « *une tension entre une grande technicité et des enjeux fondamentaux qui touchent chaque être humain dans la représentation qu'il a de lui-même et de son espèce* » (rapport de synthèse sur les États généraux de la bioéthique, et avis 129[5], p. 40).

Mais pourquoi s'intéresser particulièrement aux données massives ? Elles sont certes une composante essentielle des sciences et technologies du numérique et en particulier de l'apprentissage machine, de la robotisation et des nouveaux moyens de communication. Elles font également partie de la révolution numérique qui bouleverse en profondeur notre société. Cet intérêt se justifie par la spécificité des données massives qui mérite assurément une étude particulière.

Ce terme est passé dans le langage courant, bien qu'il n'en existe aucune définition universellement reconnue. Il a de nombreuses acceptions, qui dépendent du domaine dans lequel on l'utilise. D'une manière générale, il sert à désigner la disponibilité soit d'un nombre important de données, soit de données de taille importante, que seuls les outils du numérique allant de l'algorithmique à la puissance de calcul des ordinateurs permettent de traiter efficacement. L'information que contiennent initialement les données, quelle que soit leur origine (scientifique, administrative, personnelle ou autre) est considérablement enrichie par les rapprochements qui peuvent être opérés entre elles. Très diverses, ces données peuvent avoir un lien avec la santé ou le bien-être. Elles peuvent être issues de mesures biologiques, notamment celles issues des champs d'étude tels que la génomique, la protéomique[6] ou la métabolomique[7], de données cliniques, environnementales et comportementales groupées dans de grandes cohortes et collectées ponctuellement ou de façon répétée à partir de dossiers médicaux, mais aussi des individus eux-mêmes, *via* les réseaux sociaux et d'autres moyens de communication. L'originalité de l'analyse des données massives

---

dante du comité sur les questionnements éthiques liés à leur utilisation, au-delà même des questions de génétique qui font l'objet du présent avis. »

[4] « Numérique et santé. Quels enjeux éthiques pour quelles régulations ? ». Rapport du groupe de travail commandé par le CCNE avec le concours de la commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique d'Allistene (CERNA).

[5] Les États généraux ont eu lieu au premier semestre de 2018 et l'ensemble des travaux, arguments, opinions, ont été restitués dans un rapport de synthèse publié en juin 2018. Par ailleurs, le CCNE a publié dans son avis n° 129 une « Contribution du Comité consultatif national d'éthique à la révision de la loi de bioéthique 2018-2019 ».

[6] Étude de l'ensemble des protéines d'un organisme, d'une cellule.

[7] Analyse des métabolites présents dans l'organisme et résultant des processus biologiques survenant au cours du métabolisme.

vient de ce qu'elle ne s'appuie pas obligatoirement sur des structurations préexistantes, et permet de découvrir au sein des données collectées des corrélations, voire des causalités, à l'aide d'une algorithmique spécifique.

Ces technologies numériques, fondées sur des innovations scientifiques dont le développement est très rapide, gouvernent déjà de multiples aspects de notre quotidien (information et documentation, localisation, communication, transactions commerciales et financières, gestion des processus industriels, prédiction et aide à la décision). Elles induisent des modifications très profondes qui portent essentiellement sur deux points :

- la gestion et la communication des données : une fois collectées, elles peuvent être reproduites à l'infini sans perte de qualité ; stockées dans des « gisements[8] » ou des plateformes, elles peuvent être cédées et utilisées très largement, y compris pour des usages différents de ceux pour lesquels elles avaient été initialement fournies ou captées ;

- la prise de décision : le traitement algorithmique extrait de la masse de données dont on dispose des informations dont la précision peut apparaitre suffisante pour guider le raisonnement humain, si ce n'est se substituer à lui.

La très rapide évolution des sciences et technologies du numérique a déjà permis d'importantes innovations pour la prise en charge des patients et l'organisation du système de santé. Cela sera encore plus vrai demain car cette évolution est un fait majeur, irréversible.

Mais ces nouvelles technologies, par les ruptures qu'elles provoquent, induisent des questionnements éthiques spécifiques :

A/ L'information délivrée à la personne et son consentement au recueil et à l'utilisation des données personnelles sont rendus plus complexes dès lors que les données peuvent être aisément dupliquées et réutilisées à des fins non initialement définies et que leur traitement fera apparaître de nouvelles données, dites secondaires, souvent plus sensibles que les données initiales.

B/ L'analyse porte sur un nombre de données très important et qui n'a généralement pas été fixé à l'avance ; leur croisement permet souvent une identification très précise des personnes, et les efforts d'anonymisation des données initiales peuvent ne plus

[8] Une riche terminologie désigne les dispositifs de stockage de données : entrepôts, gisements (par référence aux matières premières auxquelles les données sont comparées), lacs, bases, banques.

être à cet égard une garantie suffisante pour la protection des droits de la personne[9] ; apparaissent dès lors sous un jour nouveau les enjeux éthiques de cette protection, mais aussi l'enjeu de la solidarité. En effet, une connaissance très précise des prédispositions de chacun pourrait induire une gestion individualisée de ces risques et menacer le principe de mutualisation sur lequel repose notre système de protection sociale.

C/ La relation de soin et la prise en charge de la personne dans le cadre de la médecine reposent classiquement sur un rapport humain fondé sur la confiance, l'écoute, l'observation, l'utilisation d'indicateurs physiologiques et l'expérience ; ce rapport ne peut qu'être profondément modifié par la place que prendra dans cette démarche une aide à la décision fondée sur l'exploitation algorithmique des données.

D/ Au-delà de la relation de soin, les paramètres qui influencent la santé et le bien-être deviennent mesurables à l'aide de dispositifs et d'applications connectés. Les individus sont ainsi de plus en plus incités à contrôler eux-mêmes leur état physique et à se regrouper pour comprendre les résultats de la recherche médicale en lien avec la maladie qui les concerne[10]. En conséquence, de nouveaux acteurs interviennent, dont certains peuvent exploiter à des fins commerciales le marché de la santé et du bien-être.

Ces questions ont été abordées lors des États généraux de la bioéthique que le CCNE a organisés en 2018 en préalable à la révision des lois de bioéthique. Elles sont traitées aux pages 73 à 84 de son rapport de synthèse[11]. Trois préoccupations ont été exprimées de manière récurrente :

- la nécessité d'explications et d'informations sur le cheminement des données numériques et l'exploitation des données collectées ;
- la crainte que le développement des outils numériques n'induise une perte de la relation humaine singulière entre le patient et le médecin, avec, à terme, le risque que la décision médicale soit imposée de manière impersonnelle par l'outil numérique ;
- une méfiance partagée sur le devenir des données et le risque de leur exploitation malveillante, en particulier vis-à-vis de personnes vulnérables.

La réflexion du CCNE porte sur l'analyse des enjeux éthiques nouveaux qui accompagnent cette mutation technologique rapide de l'exploitation de données massives, tant

[9] ARTICLE 29. *Data protection working party on the protection of individuals with regard to the processing of personal data. Opinion 05/2014 on anonymisation techniques* (0829/14/EN WP216). Avril 2014. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

[10] Voir notamment les sites *PatientsLikeMe* ou *Carenity*.

[11] Rapport de synthèse des États généraux. Juin 2018. https://www.ccne-ethique.fr/fr/publications/rapport-des-etats-generaux-de-la-bioethique-2018

au regard de la personne que des valeurs sur lesquelles repose notre système de santé. Pour autant, les réponses proposées pour préserver le respect de ces valeurs ne doivent pas nous priver des progrès qui émanent de l'exploitation des données massives dans le domaine de la santé et de la recherche. Comme le souligne l'avis 129 du CCNE « *l'insuffisance du recours au numérique dans la prise en charge des patients, pour la recherche ou pour soutenir le développement du pilotage par les données induit, sur une large échelle, des situations non éthiques au sein de notre système de santé* ».

1. LES DONNÉES DE SANTÉ À L'ÈRE DES DONNÉES MASSIVES

1.1 Une rupture

Depuis l'avènement de la médecine, clinique et biologique, à la fin du moyen âge, les soignants ont collecté des données concernant chaque personne malade, sans consentement explicite des patients. Le document de collection était une « observation » fondant le dossier médical. Cette observation, propriété de l'institution (et non de la personne), était l'outil de base de l'élaboration du progrès dans l'identification des situations morbides par rapprochement des observations de mêmes contenus.

Depuis la seconde moitié du XIX^e^ siècle, les praticiens ont cherché à donner aux signes cliniques et biologiques observés une valeur prédictive d'un diagnostic, voire du succès d'une prescription thérapeutique.

Au cours du temps, ces valeurs séméiologiques médicales ont bénéficié des avancées mathématiques de Bayes (bien longtemps méconnues puisqu'elles datent du milieu du XVII^e^ siècle) sur les cumuls de probabilité prédictive. Les approches bayésiennes en médecine clinique sont récentes[12]. Elles ont considérablement contribué à améliorer la précision diagnostique voire thérapeutique. Elles ont de longtemps précédé les collections massives de données que l'intelligence humaine ne peut plus traiter directement.

[12] La théorie Bayésienne fournit un modèle mathématique de la manière optimale de mener un raisonnement plausible en présence d'incertitudes. La règle de Bayes indique comment combiner, de façon optimale, les [informations] *a priori* issus de notre évolution ou de notre mémoire avec les données reçues du monde extérieur (source : Stanislas Dehaene, cours au Collège de France). L'inférence bayésienne est une démarche inductive qui consiste à calculer la probabilité d'une hypothèse à partir de connaissances a priori données sous la forme de mesures de probabilité. Dans le raisonnement bayésien, au contraire de la statistique classique, les paramètres sont considérés comme des variables aléatoires auxquelles on affecte une densité de probabilité. La probabilité est une mesure du degré de croyance (ou de confiance) dans l'occurrence d'un évènement ou dans la véracité d'une proposition. C'est la traduction numérique d'une connaissance ; elle mesure un degré de certitude dans la vérité d'une hypothèse.

Une innovation majeure vient de ce qu'en France, depuis une quinzaine d'années[13], le dossier médical est devenu la propriété du patient. Son usage à des fins collectives (de recherche médicale ou d'aide à la décision médicale) est donc devenu dépendant du consentement explicite et éclairé du titulaire du dossier, ce qui introduit un changement de logique important dans l'utilisation des données recueillies dans le cadre du soin.

Par ailleurs, la numérisation des informations cliniques et paracliniques collectées dans les institutions de soins (données des observations et des dossiers médicaux comportant les résultats des examens cliniques, biologiques, d'imagerie statique et dynamique, ainsi que les données médico-administratives) a permis la constitution de très grandes collections de données. Elles sont apparues au sein des institutions publiques (grands hôpitaux universitaires voire privés dans le monde anglo-saxon, centres hospitaliers universitaires, Assurance maladie et système national d'information inter régimes de l'Assurance maladie ou SNIIRAM, en France). Elles sont apparues aussi au sein des institutions privées participant au service public de délivrance des soins (centres anticancéreux en particulier).

Les données massives introduisent une rupture par leurs quatre caractéristiques majeures :

- un changement d'échelle, tenant à l'augmentation considérable du nombre des données disponibles[14] et de notre capacité à les analyser ;
- leur pérennité : utiliser les données ne les détruit pas, elles sont donc réutilisables ;
- leur diffusion rapide, qui permet leur partage, et peut s'opérer au-delà de l'équipe médicale et des frontières nationales ;
- leur capacité à générer de nouvelles informations (données secondaires) et de nouvelles hypothèses, par l'effet de leur traitement.

Il s'agit d'une rupture de nature technologique qui modifie profondément la manière d'aborder les questions relatives à la santé et qui oblige à réexaminer la manière dont sont considérées les préoccupations éthiques soulevées depuis trente ans par l'introduction, puis l'évolution, des techniques numériques dans le domaine de la santé, pour les motifs suivants :

[13] Cette innovation date de la loi n°2002-303 du 4 mars 2002 relative aux droits des malades et à la qualité du système de santé, dite aussi loi Kouchner.

[14] Ces données représentent aujourd'hui des terabytes (10^{12}) accumulés chaque jour, et donc des petabytes (10^{15}) si l'on considère l'ensemble des données. La quantité totale de données collectées dans le monde double tous les 2 ou 3 ans.

- Toute information primaire prise isolément, apparemment anodine, doit être aujourd'hui considérée comme pouvant contribuer à une information sur la santé. Existe en effet la possibilité, fondamentalement nouvelle, de la réutiliser et de la croiser, dans un contexte différent, avec d'autres données qui ne lui sont pas liées. Ce traitement fait apparaître de nouvelles informations (données secondaires ou déduites), qui, elles, peuvent être des données personnelles sensibles et identifiantes. Celles-ci sont sauvegardées et peuvent être utilisées à l'insu des titulaires des données initiales, alors même qu'elles peuvent se prêter à une exploitation pouvant s'avérer pour eux bienfaisante mais aussi éventuellement nocive. Cette impossibilité de définir *a priori* une donnée de santé est aggravée par le caractère massif et non sélectionné de la collecte, qui aboutit au stockage de vastes ensembles donnant lieu à une analyse dynamique constamment mise à jour et actualisée, au sein de laquelle le chercheur, le médecin ou même d'autres intervenants peuvent puiser.

- La technologie numérique abolit le particularisme des données collectées et permet de les connecter, quels que soient leur support (écrit, son, image), leur domaine (santé ou autre) et leur origine (réseaux sociaux, applications mobiles, écrits informatiques). Il devient dès lors plus complexe de caractériser une donnée comme sensible et donc de lui apporter une protection spécifique.

- La notion même de santé répond aujourd'hui à une définition élargie pour s'étendre au bien-être, comme le préconise l'OMS (Organisation mondiale de la santé), mais aussi pour prendre en compte des facteurs environnementaux et sociaux, ainsi que des données sur le mode de vie qui peuvent être recueillies en temps réel grâce aux objets connectés.

- Les données primaires sont recueillies dans de multiples situations : elles sont fournies par les individus eux-mêmes (réseaux sociaux, applications mobiles, géolocalisations, objets connectés, paiements numériques), qu'ils en soient conscients ou non, elles sont collectées dans le cadre du soin (entrepôts de données hospitalières, DMP [dossier médical partagé], le futur espace numérique de santé[15]), ou par les chercheurs pour alimenter des bases de données, registres et cohortes ; à quoi il faut ajouter les données collectées par les structures médico-administratives de santé qui constituent le SNDS[16].

[15] Voir le rapport : Stratégie de transformation du système de santé. « Accélérer le virage numérique » (rapport final octobre 2018). « *Il s'agit de créer dès la naissance pour chaque usager un espace numérique de santé sécurisé et personnalisé lui permettant d'avoir accès à l'ensemble de ses données et services de santé tout au long de sa vie* ». https://solidarites-sante.gouv.fr/IMG/pdf/masante2022_rapport_virage_numerique-2.pdf

[16] Le système national des données de santé regroupe : le système national d'information inter régimes de l'Assurance maladie (SNIIRAM) ; les données des hôpitaux et autres établissements de santé (programme de médicalisation des systèmes d'information - PMSI) ; les données statistiques relatives aux causes de décès (BCMD). Les données relatives au handicap devraient le compléter.

Tout ce qui précède montre que les données primaires forment une « matière première » que cherchent à utiliser des opérateurs dont l'origine, la formation et les motivations sont très variées. Ils peuvent poursuivre un objectif de soin (professionnels de santé), de conception de nouveaux médicaments ou dispositifs (firmes pharmaceutiques, *start-up*), d'amélioration des connaissances (chercheurs), commercial (*start-up* et multinationales de l'internet), de meilleure performance économique ou de santé publique (institutions publiques). Il s'agit d'un ensemble hétérogène d'acteurs ne partageant pas une même culture professionnelle, ni les mêmes valeurs, notamment la protection de la vie privée, ce qui ajoute à la complexité du problème.

Citons encore, au nombre des ruptures qu'induit la généralisation du recueil, de la conservation et du traitement des données massives, un défi environnemental qui résulte de ces opérations et concerne plusieurs aspects :

- la consommation d'énergie très élevée des ordinateurs, *data centers* et réseaux, qui représente près de 10 % de la consommation mondiale d'électricité, chiffre qui ne cesse d'augmenter ;
- l'empreinte carbone élevée due au fonctionnement des infrastructures ;
- une consommation des métaux rares nécessaires à la fabrication des ordinateurs et *smartphones* ; leur extraction emprunte des techniques destructives et utilise des produits nocifs pour l'environnement et ils produisent, après utilisation, des déchets dont une partie importante est retrouvée dans des décharges sauvages en Asie ou en Afrique.

Il s'agit d'un sujet très important, qui ne sera pas développé plus avant dans le présent avis car il n'est pas spécifiquement lié à la santé, même si la part de cette consommation liée à la santé est certainement importante. Mais il est certain que, dans ce domaine où la disponibilité et la continuité des soins sont d'une importance primordiale, nous ne pourrons pas nous fier durablement à ces technologies nouvelles si nous n'avons pas l'assurance de maîtriser les conditions de leur fonctionnement. Cette maîtrise devra porter sur une ressource énergétique suffisante (qu'elle passe par une forte réduction de la consommation ou par le développement de sources d'énergie fiables et respectueuses de l'environnement) et sur la disponibilité des composants nécessaires à la fabrication des appareils numériques, qui devront pouvoir être extraits, exploités et retraités de manière non nocive[17]. Cela ne fait que souligner l'importance d'une recherche fondamentale sur les technologies, cherchant à en améliorer le rapport coût/efficacité.

[17] Le lecteur pourra consulter à ce propos le site ecoinfo.cnrs.fr, ainsi que l'article "Impacts environnementaux du numérique, de quoi parle-t-on ?" par Françoise Berthoud, sur le Blog Binaire (l'informatique : science et technique au cœur du numérique) /Le monde, 29 janvier 2019 (http://binaire.blog.lemonde.fr/2019/01/29/impacts-environnementaux-du-numerique-de-quoi-parle-t-on/), ainsi que le rapport du comité éthique de l'Unesco.

1.2 Une mutation technologique qui induit des changements de comportement

Les données massives, par leurs caractéristiques mêmes, remettent en cause des distinctions sur lesquelles nous sommes habitués à raisonner et qui servent classiquement de support à la réflexion éthique : vie publique et vie privée, relation de soin et marché économique de la santé et du bien-être, individualisme et solidarité.

- *Vie publique et vie privée*

La perspective du traitement croisé de données d'origine et de nature différentes, recueillies aussi bien par des opérateurs publics de santé que par des opérateurs privés dans le cadre des nouveaux modes de vie et de relations sociales, efface la distinction traditionnelle entre vie publique et vie privée, puisqu'une donnée en apparence anodine peut, une fois corrélée avec d'autres, donner une information sur la santé. Cette remise en cause est accentuée par l'impossibilité de garantir sur le long terme l'efficacité d'une anonymisation des données sensibles. La perte de confidentialité de la vie privée – résultant parfois d'une divulgation par les personnes elles-mêmes – est une caractéristique majeure de notre époque. Elle donne à l'exploitation des données massives un champ possible d'investigation particulièrement vaste. La notion de données de santé ne peut plus se limiter aux seules données personnelles recueillies dans le cadre du soin.

- *Transformation du soin et du marché de la santé par l'intervention de nouveaux acteurs*

S'ajoutant aux traditionnels acteurs publics et privés de la santé, de nouveaux acteurs interviennent dans la relation de soin, ainsi que sur le marché de la santé et du bien-être.
Il s'agit d'abord des « *data scientists*[18] », qui interviennent quel que soit le domaine d'application. Ils occupent une place centrale puisqu'ils sont responsables de la gestion des données et de leur exploitation en vue de produire de nouvelles informations.

Beaucoup d'initiatives viennent des patients eux-mêmes, qui recherchent et partagent une information médicale, à l'image de la plate-forme *PatientsLikeMe*[19] ; mais ce sont principalement des entreprises privées qui sont le moteur de l'innovation en matière de technologies numériques y compris dans le domaine de la santé. Certaines s'adressent aux personnes et aux patients comme à des consommateurs. Profitant de

[18] Ce terme désigne les experts de la gestion et de l'analyse des données massives. Ils conçoivent les modèles et algorithmes pour collecter, traiter et restituer les données. Il faut y ajouter les curateurs chargés de nettoyer les données.

[19] https://patientslikeme.com

la multiplication des objets connectés[20] et de l'usage généralisé des réseaux sociaux, elles cherchent à attirer leur clientèle en incitant les individus à être acteurs de leur propre santé. Elles le font par une approche individualisée recourant à la notion de performance.

Au-delà des risques de divulgation difficilement contrôlable de données personnelles sensibles, cette évolution marque l'importance croissante, dans le domaine de la santé, d'acteurs qui n'ont pas pour finalité le soin mais l'exploitation d'un marché. Ils ne se sentent pas nécessairement tenus par les obligations déontologiques ni par les règles du secret auxquelles sont spécifiquement soumis les acteurs professionnels de santé, même s'ils veillent souvent à s'assurer leur concours.

Pour autant, il ne s'agit pas de condamner ces nouvelles pratiques. Elles peuvent avoir pour réelle vertu d'inciter nos contemporains à se préoccuper davantage de leur santé ; elles permettent d'en mesurer les paramètres et de prendre en compte, par une observation en temps réel, les facteurs relevant de l'environnement et du mode de vie, données déterminantes pour la prise en charge du patient.

- *Médecine de précision et solidarité*

Le traitement de multiples données intéressant la santé, relatives au soin proprement dit, mais aussi à la génétique, aux habitudes et au mode de vie, ainsi qu'à l'environnement d'une personne, permet une approche très individualisée de la prédiction d'une maladie, de sa prévention, ou, lorsqu'elle survient, de la thérapeutique à appliquer. Avec le concept dit d'une « médecine de précision », chaque patient devient en quelque sorte un cas unique. L'ensemble des informations permet d'atteindre une précision diagnostique et thérapeutique accrue et de meilleure qualité, améliorant le pronostic. Une telle précision favorise la qualité des soins, mais elle n'est pas sans conséquence. Elle permet notamment l'identification d'éventuels facteurs de risque, partagés par ceux qui développeront ultérieurement la maladie. La médecine de précision pourrait ainsi rendre possible un profilage qui pourrait dispenser de recourir à la notion d'aléa pour mutualiser les risques au sein d'une collectivité. Cette mutualisation constitue pourtant un principe sur lequel repose la solidarité nationale, valeur essentielle de notre politique de santé.[21]

[20] A l'ère numérique, chacun de nous est invité à être l'acteur de sa propre santé. De multiples applications proposent de mesurer divers paramètres, tels que poids, rythme cardiaque, tension artérielle, taux de glycémie et de cholestérol, et d'en tirer des enseignements sur l'hygiène de vie et la recherche du bien-être (notion de « *quantified self* »). La diffusion rapide de ces pratiques induit des comportements permettant à chacun de suivre les conseils d'un « coach » et de mesurer ses progrès, en se comparant éventuellement à d'autres.

[21] Cette question sera développée à la rubrique 2.2- Justice : la solidarité au défi de l'individualisation

Les limites devenues imprécises entre vie publique et vie privée, entre commerce et soin, ainsi qu'entre individualisme et solidarité, montrent que nous sommes confrontés à une situation nouvelle marquée par de fortes incertitudes et des risques de comportements non éthiques. Ce constat nous incite à faire preuve de vigilance, particulièrement dans la manière dont nous aborderons en matière de santé les notions essentielles d'information, de consentement et de contrôle.

1.3 La notion de « donnée relative à la santé »

La définition d'une donnée de santé est complexe et mouvante. Elle s'est élargie au fil du temps. Ce qui a été dit à la rubrique 1.1 sur l'extension du champ de la santé et sur la possibilité d'obtenir secondairement des données de santé sensibles à partir de données primaires *a priori* sans rapport direct avec la santé, montre que cette notion ne se limite pas aux données de santé « par nature » (celles recueillies dans le cadre d'une prise en charge médicale). Elle inclut nécessairement aussi celles qui – sans être en elles-mêmes qualifiées de données de santé – le deviennent, soit par leur croisement avec d'autres données qui permet de tirer une conclusion sur l'état de santé ou le risque pour la santé d'une personne, soit « par destination » (parce qu'elles sont utilisées dans un parcours de soin)[22]. Dans ces deux cas, c'est la finalité du traitement qui qualifie de données de santé des données qui ne le sont pas *a priori*. Pour évoquer l'ensemble, c'est l'appellation de « données relatives à la santé » que nous proposons d'employer dans cet avis.

Il n'est dès lors pas étonnant que le règlement général 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 (désigné ci-après le RGPD ou le règlement européen) –, qui sera examiné à la rubrique suivante, en donne en son article 4 cette définition extensive des « données concernant la santé » : « *Aux fins du présent règlement, on entend par « données concernant la santé », les données à caractère personnel relatives à la santé physique ou mentale 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* ».[23]

[22] Voir le site de la Commission nationale de l'informatique et des libertés (CNIL) rubrique « Qu'est-ce qu'une donnée de santé ? ». (cnil.fr).

[23] Le considérant N° 35 du RGPD apporte les précisions suivantes : « *Les données à caractère personnel concernant la santé devraient comprendre l'ensemble des données se rapportant à l'état de santé d'une personne concernée qui révèlent des informations sur l'état de santé physique ou mentale passé, présent ou futur de la personne concernée. Cela comprend des informations sur la personne physique collectées lors de l'inscription de cette personne physique en vue de bénéficier de services de soins de santé ou lors de la prestation de ces services au sens de la directive 2011/24/UE du Parlement européen et du Conseil au bénéfice de cette personne physique; un numéro, un symbole ou un élément spécifique attribué à une personne physique pour l'identifier de manière unique à des fins de santé; des informations obtenues lors du test ou de l'examen d'une partie du corps ou d'une substance corporelle, y compris à partir de données génétiques et d'échantillons biologiques; et toute information concernant, par exemple, une maladie, un handicap, un risque de maladie, les antécédents médicaux, un*

Cette définition faisant des données concernant la santé un sous-ensemble des données à caractère personnel, il convient de se référer à la définition de celles-ci donnée par le même article 4 : « *Aux fins du présent règlement, on entend par « données à caractère personnel », toute information se rapportant à une personne physique identifiée ou identifiable (ci-après dénommée «personne concernée») ; est réputée être une «personne physique identifiable» une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale* ».
La loi informatique et liberté actualisée[24] prévoit une section dédiée aux traitements de données à caractère personnel dans le domaine de la santé (chapitre II, section 3, art. 64-73).

La variété des données relatives à la santé – dans la définition que nous en donnons – ne permet pas une énumération exhaustive. On peut citer, comme catégories essentielles, les données historiques (relatives aux antécédents médicaux familiaux et personnels), les données cliniques, biologiques, celles relatives aux résultats d'analyses et d'investigations (y compris les données d'imagerie), les données relatives aux traitements, aux prescriptions et aux consommations de médicaments, ainsi que les données environnementales, socio-économiques et démographiques, les données comportementales (qui informent sur la qualité de vie et les habitudes).

Dans la suite de cet avis, nous qualifierons donnée de santé toute information relative à la santé et au bien-être physique ou mental, et donnée personnelle de santé toute donnée de santé se rapportant à une personne physique identifiée ou identifiable.

1.4 La protection des données de santé

Dans le cadre de la prise en charge médicale, les données de santé sont soumises à plusieurs législations protectrices, essentiellement le RGPD (art. 9.1), la loi informatique et liberté (LIL) et le code de la santé publique (CSP). La LIL modifiée (art. 6.I), comme le RGPD (art. 9.1), prévoit un principe d'interdiction de traitement des données de santé, qui est toutefois tempéré par une série d'exceptions (voir chapitre 2.1.3).

---

traitement clinique ou l'état physiologique ou biomédical de la personne concernée, indépendamment de sa source, qu'elle provienne par exemple d'un médecin ou d'un autre professionnel de la santé, d'un hôpital, d'un dispositif médical ou d'un test de diagnostic in vitro. »

[24] Cette modification de la loi N° 78-17 du 6 janvier 1978, dite informatique et libertés, résulte de la loi n° 2018-493 du 20 juin 2018, qui transpose en droit français le RGPD. Une ordonnance de décembre 2018 améliore la lisibilité du cadre juridique national.

Le CSP réglemente le secret médical (art. L. 1110-4), l'hébergement des données de santé[25] (art. L. 1111-8 et R. 1111-8-8 et s.), leur mise à disposition (art. L. 1460-1 et s.), la conformité des systèmes d'information (art. L1110-4-1), le partage des données (loi de modernisation de notre système de santé, art. R1110-1 et 1110-3), l'interdiction de procéder à une cession ou à une exploitation commerciale des données de santé (art. L. 1111-8, art. L 4113-7 du CSP).

L'adoption de mesures particulières de sécurité a pour avantage évident de mieux assurer la confidentialité. Celles-ci peuvent avoir l'inconvénient – en particulier pour l'hébergement – de compartimenter les bases de données (cliniques, médico administratives, génomiques) ce qui peut avoir pour effet de limiter les possibilités de communication entre elles. Or, le croisement de ces bases est d'un intérêt majeur pour la recherche médicale et l'efficacité de celle-ci s'en trouve affectée, puisqu'elle ne peut tirer tout le bénéfice de la particulière richesse des données collectées dans notre pays[26]. Pour permettre le croisement des données, les projets européens (dont ceux financés dans le cadre d'Horizon 2020) incitent les gouvernances des entrepôts de données à suivre les principes et critères du FAIR[27] (*findable, accessible, interoperable, re-usable*, ou, en Français, « trouvable, accessible, interopérable et réutilisable »). Les principes du FAIR permettent la construction, le stockage, la présentation et la publication des données afin de disposer, entre autres, d'un partage facilité et encadré. C'est le sens de la politique française de science ouverte énoncée par la ministre de l'enseignement supérieur et de la recherche le 4 juillet 2018[28].

Cette protection spécifique ne joue que pour les données *« recueillies à l'occasion d'activités de prévention, de diagnostic, de soins ou de suivi social et médico-social »* et elle ne recouvre donc pas la notion très large de données personnelles de santé que nous avons retenue à la rubrique précédente.

---

[25] Art. 1111-8 du CSP : *« Toute personne qui héberge des données de santé à caractère personnel (…) réalise cet hébergement dans les conditions prévues au présent article.*

L'hébergement, quel qu'en soit le support, papier ou électronique, est réalisé après que la personne prise en charge en a été dûment informée et sauf opposition pour un motif légitime.

(…) L'hébergeur de données mentionnées au premier alinéa (…) est titulaire d'un certificat de conformité. »

Art. L 1110-4-1 du CSP : *« Afin de garantir la qualité et la confidentialité des données de santé à caractère personnel et leur protection, les professionnels de santé, les établissements et services de santé et tout autre organisme participant à la prévention, aux soins ou au suivi médico-social et social dont les conditions d'exercice ou les activités sont régies par le présent code, utilisent, pour leur traitement, leur conservation sur support informatique et leur transmission par voie électronique, des systèmes d'information conformes aux référentiels d'interopérabilité et de sécurité élaborés par le groupement d'intérêt public mentionné à l'article L. 1111-24. Ces référentiels sont approuvés par arrêté du ministre chargé de la santé, pris après avis de la Commission nationale de l'informatique et des libertés ».*

[26] Cette question sera développée au chapitre 3.2.1.

[27] Voir le site https://www.force11.org/fairprinciples ; et Wilkinson M, *et al*. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3, *Article number* : 160018 (2016).

[28] Le Plan national pour la science ouverte annoncé par Frédérique Vidal, le 4 juillet 2018. http://www.enseignementsup-recherche.gouv.fr/cid132529/le-plan-national-pour-la-science-ouverte-les-resultats-de-la-recherche-scientifique-ouverts-a-tous-sans-entrave-sans-delai-sans-paiement.html.

D'une manière plus générale, notre pays se distingue par l'ancienneté et la fermeté de sa législation protectrice. La loi informatique et liberté a imposé dès 1978, dès lors que l'on était en présence de données personnelles, une obligation de déclaration préalable, voire parfois la nécessité d'un avis ou d'une autorisation de la CNIL (commission nationale de l'informatique et des libertés) pour leur traitement. De façon générale, cette loi, comme aujourd'hui le RGPD, prévoit que les traitements de données personnelles doivent respecter certains principes clés : une finalité déterminée, explicite et légitime, la minimisation de la collecte des données, une durée de conservation des données limitée, une obligation de sécurité, le respect des droits des personnes. La CNIL a été chargée de faire respecter ces principes, qui ont, pour l'essentiel, été depuis lors réaffirmés et complétés. Quelques aménagements ont été apportés pour tenir compte des avancées technologiques issues de la généralisation de la numérisation et de l'exploitation des données massives, notamment avec la loi n° 94-548 du 1er juillet 1994 relative au traitement de données nominatives ayant pour finalité la recherche dans le domaine de la santé, ainsi que la loi n° 2016 du 7 octobre 2016 pour une République numérique. Sont érigés en principes la neutralité de l'internet, le droit à la portabilité et à la récupération des données, la loyauté des plateformes et l'information loyale, claire et transparente des personnes s'agissant de l'utilisation de leurs données. Le droit à la protection des données à caractère personnel est notamment assuré par la reconnaissance du droit de décider et de contrôler les usages qui en sont faits et d'un droit à l'effacement de ces données. Des dispositions spécifiques sont prévues lorsque leur titulaire était mineur au moment de la collecte.

Mais ce qui a été dit plus haut des problématiques nouvelles posées par l'exploitation des données massives démontre que devenait irréaliste l'exigence d'une obligation généralisée d'autorisation ou de déclaration préalable. Cette exigence a été abandonnée par le règlement européen du 27 avril 2016, entré de plein droit en application dans les États-membres à compter du 25 mai 2018.

Ce règlement a adopté une ambitieuse législation qui réaffirme les principes protecteurs et qui prévoit plusieurs dispositifs pour assurer leur mise en œuvre[29]. Ce texte ne remet en cause ni la spécialité du consentement, ni la finalité du traitement des données personnelles, ni la minimisation du nombre des données collectées à ce qui est réellement utile. Les sanctions qu'il prévoit en cas de manquement sont dissuasives.

Pour éviter le contournement qui pourrait résulter du transfert de données personnelles dans un pays ne relevant pas de l'Union européenne et ne disposant pas d'une loi de protection des données personnelles reconnue d'un niveau équivalent au RGPD,

[29] Voir le texte du RGPD en français sur le site de la CNIL : https://www.cnil.fr/fr/reglement-europeen-protection-donnees.

où les opérateurs pourraient contester l'application du règlement européen, l'article 44 édicte qu'un tel transfert ne peut avoir lieu que si le responsable du traitement et ses sous-traitants, y compris en cas de transferts ultérieurs, respectent les dispositions protectrices du RGPD. Les entreprises qui ont leur siège en dehors de l'Europe mais qui offrent des biens ou des services en Europe doivent aussi respecter le RGPD (article 3). La responsabilité du responsable du traitement et de ses sous-traitants est minutieusement réglementée et sanctionnable. Chaque État-membre doit en outre nommer une autorité de contrôle indépendante (articles 50 à 59).

Le RGPD met l'accent sur l'autorégulation et sur une responsabilisation accrue des acteurs.

Son apport essentiel réside en définitive dans la reconnaissance du principe « *d'accountability* », c'est-à-dire l'obligation pour le responsable du traitement de prendre des mesures permettant de garantir, et de prouver, que la protection des données a bien été une préoccupation constante et que les règles pour la garantir ont été respectées.

Le RGPD comporte des dispositions particulières applicables aux traitements de données personnelles visant à faciliter la recherche, entendue dans une conception large incluant le développement et la démonstration de technologies, la recherche fondamentale et appliquée, ainsi que la recherche financée par le secteur privé (voir notamment l'article 89). La loi n° 2018-493 du 20 juin 2018, prise pour l'adaptation à notre pays de ce règlement, institue la CNIL comme autorité de contrôle nationale. Le rôle de la CNIL est celui d'un véritable régulateur. Dans le domaine de la recherche en santé, elle a adopté cinq nouvelles méthodologies de référence adaptées au cadre juridique en matière de données de santé, et un référentiel pour accéder à certaines données du SNIIRAM[30]. Elles permettent d'alléger les formalités liées au traitement de ces données pour les besoins de la recherche en santé puisqu'une demande d'autorisation n'est pas nécessaire en cas de déclaration de conformité[31]. Est créé un comité d'audit qui fait réaliser, par des prestataires sélectionnés, des audits sur l'ensemble des systèmes réunissant, organisant ou mettant à disposition tout ou partie des données du système national des données de santé.

Les comités de protection des personnes (CPP) participent également à la protection des données[32] ; ils sont chargés d'émettre un avis préalable sur les conditions de vali-

[30] https://www.cnil.fr/fr/quelles-formalites-pour-les-traitements-de-donnees-de-sante-caractere-personnel. SNIIRAM : Système national d'information inter-régimes de l'Assurance maladie.

[31] Ces méthodologies de référence concernent les recherches impliquant la personne humaine, les études ou évaluations n'impliquant pas la personne humaine, l'accès aux données du PMSI par les établissements de santé, les fédérations et les industriels du secteur de la santé aux fins de réaliser des études dans des conditions strictes de confidentialité et de sécurité.

[32] LIL, ordonnance décembre 2018, art.76.

dité de toute recherche impliquant la personne humaine, au regard des critères définis par l'article L 1123-7 du CSP. Leur rôle est précisé par les articles L 1121-1 à L 1126-11 du même code. Ils sont sollicités en particulier pour tout projet impliquant la réutilisation de données de santé, et sur la nécessité éventuelle de recontacter le titulaire des données. Le CEREES[33] intervient pour les demandes d'autorisation pour les études, évaluations ou recherches n'impliquant pas la personne humaine.

2. L'ÉTHIQUE À L'ÉPREUVE DES DONNÉES MASSIVES DANS LE DOMAINE DE LA SANTÉ

Les données massives fragilisent les repères qui servent classiquement de support à la réflexion éthique (distinction entre vie publique et vie privée, entre individualisme et solidarité, entre préoccupations économiques et finalité du soin[34]). Outre l'incertitude qui en résulte, cela accentue des tensions entre individu et collectif, entre logique économique et valeurs éthiques qui commandent la protection de la vie privée et l'autonomie des personnes.

Mais il serait caricatural d'opposer individualisme et libéralisme économique d'une part, principes éthiques d'autre part.

Les valeurs éthiques, telles qu'autonomie, respect de la vie privée et de la dignité, égalité, solidarité/fraternité/justice, ne peuvent pas en effet être prises isolément, comme si chacune d'elles constituait un absolu. Sans que l'on puisse établir entre elles une hiérarchie, elles sont elles-mêmes traversées par les tensions que nous venons d'évoquer. C'est ainsi que les principes d'autonomie et de justice peuvent inciter à privilégier la capacité d'action individuelle du sujet, et entrer en conflit avec les valeurs d'égalité, de solidarité/fraternité. C'est dans une perspective non pas statique mais dynamique qu'il faut examiner les enjeux éthiques issus de l'interaction entre ces valeurs. Ces enjeux évoluent tant en fonction des mutations constantes et rapides de notre société que des avancées scientifiques et technologiques. Un point d'équilibre est donc toujours à rechercher et il faut l'adapter aux spécificités de chaque contexte. Cette dernière dimension, plus spécifique et concrète, sera l'objet du chapitre 3.

Mais il convient au préalable de s'interroger sur la manière dont les données massives obligent à renouveler la réflexion éthique dans le domaine de la santé.

Pour conduire cette réflexion, il sera recouru aux grands principes éthiques qui fondent la déontologie médicale : respect de la personne (incluant le respect de son

[33] Comité d'expertise pour les recherches, les études et les évaluations dans le domaine de la santé (CEREES).
[34] Voir §1.2.

autonomie), justice, pertinence et bienfaisance (incluant ici l'obligation de non-nuisance).

2.1 Respect de la personne

2.1.1 Une valeur confrontée à une situation nouvelle

Le CCNE soulignait dans sa contribution à la révision de la loi de bioéthique 2018-2019 (avis 129) les bouleversements récents dans la représentation de l'humain ; il mentionnait ainsi le génome et les données parmi les « *sauts qualitatifs en cours dans la représentation du corps et de son rapport à la personne, qu'il s'agisse de progrès ou de fragilité* ». Il poursuivait en mentionnant que tout ceci « *affecte l'image même de l'espèce et de l'homme autant que la place de l'individu, du patient et du citoyen* ». Ce questionnement a été repris lors de la consultation des États généraux de la bioéthique avec l'interrogation sur une « *nouvelle objectivation du corps humain où le génome et les données de santé s'ajoutent aux caractéristiques corporelles traditionnelles.* » [35]

Cela pose la question du lien entre la donnée et la personne. Faut-il considérer que la donnée est un « objet détachable de la personne », ce qui conduit à ignorer celle-ci et à limiter la protection de la personne due à l'usage qui peut être fait de la donnée ? Doit-on au contraire estimer que celle-ci donne une représentation, une image, de la personne ? On parle alors d'identité ou de profil numérique, ce qu'Antoinette Rouvroy[36] résume ainsi : « *les sujets [qui] n'existent que de manière infra-individuelle (fragmentés dans diverses banques de données) ou supra-individuelle (les « profils » ne s'adressant jamais qu'à des ensembles d'individus, ou, plus exactement, à des ensembles de comportements)* » [37].

Dès lors que le numérique permet une reproduction à l'infini des données sans dépossession de leur titulaire, cela suscite une interrogation sur la pertinence de l'application de la notion de propriété. La question de la patrimonialité des données personnelles est en effet au cœur de débats juridiques entre conception personnaliste et conception patrimoniale. Les données personnelles sont à la fois des prolongements de la personnalité des individus (révélant l'identité, la vie privée, des choix personnels, etc.) et des informations qui peuvent circuler, être cédées, reproduites. « *Sur le plan juridique, [les données] se situent à la frontière entre la personne et la chose,*

[35] Voir sur le site du CCNE le rapport des États généraux de la bioéthique 2018, rendu public le 2 juillet 2018.

[36] Chercheuse qualifiée du FNRS au centre de recherche en information, droit et société (CRIDS), Université de Namur, Belgique.

[37] Antoinette Rouvroy. Pour une défense de l'éprouvante inopérationnalité du droit face à l'opérationnalité sans épreuve du comportementalisme numérique. *Dissensus*. Revue de philosophie politique de l'université de Liège, Avril 2011, pp. 127-149.

et peuvent relever aussi bien de l'une que de l'autre »[38]. Le droit français a plutôt opté pour une qualification extrapatrimoniale des données (notamment la loi du 7 octobre 2016 pour une République numérique qui retient un « droit à l'autodétermination informationnelle »). Même si des voix s'élèvent pour la reconnaissance d'un droit de propriété, retenant l'argument de la valeur économique des données[39], la qualification extrapatrimoniale semble devoir être privilégiée. Une auteure avisée s'insurge contre l'hypothèse de la qualification des données personnelles comme des biens et estime même que « les textes ne vont pas assez loin dans l'adoption de l'approche personnaliste »[40]. C'est la position également retenue par le Conseil d'État qui considère que « *s'il convient de renforcer la dimension de l'individu acteur dans le droit à la protection des données, c'est en envisageant celui-ci comme un droit à l'autodétermination plutôt que comme un droit de propriété* »[41].

Mais puisque la cession des données numériques peut s'opérer sans inconvénient apparent pour le titulaire, on peut se demander pourquoi chercher à le protéger si l'utilisation par des tiers n'entrave en rien sa propre capacité d'agir.

Il faut insister fortement sur la nécessité de cette protection. Nous avons vu en effet que la logique du traitement numérique repose sur les corrélations établies entre une multitude de données qui, prises isolément, peuvent être peu signifiantes, mais qui, par recoupement, donnent des indications très précises sur les droits individuels les plus sensibles (opinions philosophiques et politiques, croyances religieuses, état de santé, orientations sexuelles, mode de vie, personnes et lieux fréquentés). Laisser s'opérer une appropriation sans limite des données personnelles nous éloignerait d'une société démocratique. Cela reviendrait en effet à admettre la menace d'une société de surveillance et de contrôle des individus, par de multiples opérateurs publics ou privés, agissant aux fins les plus diverses, qu'elles soient commerciales, politiques ou sécuritaires.

Mais si un consentement est cohérent avec la vision d'un « droit à l'autodétermination informationnelle », il n'atteint son plein effet que s'il est libre et éclairé, c'est-à-dire si le titulaire des données a connaissance de ceux (acteurs de santé) qui pourront utiliser les données, de l'usage qu'ils entendent en faire dans l'immédiat, et des usages

[38] Philippe Mouron. Pour ou contre la patrimonialité des données personnelles. *La revue européenne des médias et du numérique* 2018, n°46-47, pp. 91.

[39] Le Monde en date du 12 janvier 2019. « Inventer un droit patrimonial sur les données de santé », p.7. Rapport Mes data sont à moi - Pour une patrimonialité des données personnelles. *Génération libre*, janvier 2018 (https://www.generationlibre.eu/wp-content/uploads/2018/01/2018-01-generationlibre-patrimonialite-des-donnees.pdf).

[40] Judith Rochfeld. Contre l'hypothèse de la qualification des données personnelles comme des biens, in : Les biens numériques, éd CEPRISCA, 2015, pp.221-236.

[41] Étude 2014 du Conseil d'État, Le numérique et les droits fondamentaux, pp.264. http://www.conseil-etat.fr/Decisions-Avis-Publications/Etudes-Publications/Rapports-Etudes/Etude-annuelle-2014-Le-numerique-et-les-droits-fondamentaux.

potentiels qu'ils pourraient en faire dans l'avenir. Ces données massives pourraient prendre une dimension « supraliminaire » pour reprendre un terme proposé par Gunther Anders[42], révélant une dimension trop grande pour être appréhendée par l'individu, et questionnant la notion même de consentement.

Quelques spécificités des données massives font en effet que cet objectif d'un consentement individuel libre et éclairé est très difficile, voire impossible, à atteindre dans les situations d'exploitation algorithmique de données massives :

- La technologie numérique permet la réutilisation des données, à des fins qui peuvent différer sensiblement des raisons pour lesquelles elles avaient été collectées initialement. Il en résulte une difficulté de prédéfinir la ou les finalités de l'exploitation des données massives, ce qui explique la notion de « finalités non incompatibles » développée par le RGPD (art. 5.1b et art.6.4).

- Des corrélations signifiantes, ne correspondant pas à une hypothèse prédéfinie, peuvent être mises en évidence par une réutilisation des données qui ne suppose pas nécessairement une interrogation préalable et précise du titulaire pour un consentement à cette nouvelle utilisation. L'absence de prédétermination porte tant sur les données utilisées (il est difficile de savoir à l'avance lesquelles seront concernées par les rapprochements opérés par le traitement algorithmique) que sur les résultats de leur exploitation.

- Le titulaire des données auxquelles un opérateur souhaite avoir accès se trouve dans une situation d'infériorité. Celle-ci résulte tant de l'asymétrie des savoirs sur les technologies utilisées ou l'intérêt des données que de la pression qui résulte – en particulier dans le cas de l'accès aux sites des opérateurs privés d'internet qui offrent une prestation de service - de la subordination de celle-ci à la fourniture des données.

- La complexité des processus mis en œuvre dans l'exploitation des données fait que, même si l'on offrait de communiquer aux personnes intéressées une information loyale sur les algorithmes ou les programmes qui en sont issus, elles seraient difficilement en mesure d'en retirer des informations leur permettant de faire un choix concernant le devenir et l'utilisation des données. Cette difficulté est encore accrue en cas de recours aux techniques d'apprentissage profond (*deep learning*). Les résultats obtenus *via* l'utilisation de ces algorithmes restent en effet aujourd'hui difficilement explicables pour le programmeur lui-même, ce qui en fait une question importante de recherche.

[42] Günther Anders. *Et si je suis désespéré, que voulez-vous que j'y fasse ?* (1977). Allia, collection, Petite collection.

Nous évoquerons ci-dessous (§ 3.3) les réponses possibles à cette question du consentement et quelles pourraient être des alternatives possibles au consentement individuel tel qu'on le définit classiquement.

2.1.2 Une situation nouvelle qui oblige à une autre perception des enjeux

Le règlement européen réaffirme – tout en les aménageant – les principes de la spécialité du consentement[43], de finalités déterminées, explicites et légitimes pour le traitement des données, de la minimisation de la collecte et de la limitation du traitement, ce qui s'imposait d'un point de vue éthique. Dans son rapport d'étude de 2014 sur le numérique et les droits fondamentaux, le Conseil d'État a justifié l'exigence de la finalité du traitement dans ces termes : « *Le principe de finalités déterminées est au cœur de la confiance que les personnes peuvent avoir dans les services de la société numérique. Lorsqu'elles recourent à de tels services et que les données les concernant sont collectées dans ce cadre, elles doivent avoir l'assurance que ces données ne seront pas utilisées pour d'autres finalités que celles du service, sauf à ce qu'elles en aient été informées ou que la loi le prévoie. Le principe de finalités déterminées est ce qui fait que les données personnelles ne sont pas des marchandises ou, du moins, qu'elles ne sont pas des marchandises comme les autres.* »

Un droit d'accès et de rectification et un droit à l'effacement sont reconnus à la personne concernée (articles 16 et 17 du RGPD).

L'autorisation préalable de la CNIL n'étant pas exigée en cas d'utilisation des données conforme à une méthodologie de référence44, l'enjeu est celui du contrôle, qui doit être effectif pour conserver la confiance des usagers, mais qui ne doit pas être paralysant. Il faut en effet insister sur le fait qu'il serait contraire à l'éthique de faire obstacle aux avancées majeures qui sont attendues de l'exploitation des données massives dans les domaines de la recherche médicale et du soin.

2.1.3 Comment assurer maîtrise et contrôle ?

Face à l'impossibilité de tout contrôler sont recherchées des alternatives. C'est ainsi que sont encouragés des codes de bonne conduite pouvant être soumis à l'autorité de contrôle, ainsi que des mécanismes de certifications et de labels, permettant

[43] Lorsque le consentement constitue le fondement de la licéité, ce qui n'est pas toujours le cas : il est en effet non exigé en cas de nécessité pour l'exécution du contrat, d'intérêt légitime, de mission d'intérêt public.

[44] « *La création et la mise à jour de méthodologies de référence ont été rendues nécessaires par l'évolution des textes législatifs nationaux, par l'entrée en vigueur du Règlement général sur la protection des données (RGPD) ainsi que par les retours d'expérience formulés par les acteurs de terrain sur les cadres existants. L'adoption par la CNIL de ces méthodologies vise à créer un cadre protecteur des personnes concernées favorable à la recherche, à l'innovation et la compétitivité.*»(https://www.cnil.fr/fr/recherches-dans-le-domaine-de-la-sante-la-cnil-adopte-de-nouvelles-mesures-de-simplification).

d'apporter une preuve de conformité aux exigences légales. Pour chaque traitement, un responsable doit être désigné. Il en détermine les finalités et les moyens. Il doit prendre des mesures appropriées pour fournir toutes informations utiles à la personne concernée (articles 12 à 15 du RGPD). Il est assisté, le cas échéant, par un délégué à la protection des données (obligatoire pour le secteur public [articles 37 à 39]), auquel est reconnue une indépendance fonctionnelle (article 38).

Est institué un Comité européen de la protection des données[45], qui est chargé d'assurer la cohérence de l'application du RGPD.

Le règlement européen se veut ferme sur ses principes, mais réaliste. Même si le consentement reste l'un des fondements majeurs de licéité du traitement de données personnelles, le RGPD, prenant acte de ce que cette exigence n'est pas réalisable dans tous les cas, notamment dans l'hypothèse de réutilisation de données, prévoit que le traitement de données personnelles est licite pour certaines finalités (voir article 6.1 du RGPD). Dans le cas de catégories particulières de données à caractère personnel, dont les données de santé, (article 9 du RGPD, article 8 de la loi informatique et libertés dans sa version issue de la loi du 20 juin 2018), l'interdiction de traitement est la règle, mais cette interdiction peut être levée dans la mesure où la finalité du traitement l'exige, par exemple pour « *les traitements pour lesquels la personne concernée a donné son consentement exprès* », « *les traitements comportant des données concernant la santé justifiés par l'intérêt public* », « *les traitements portant sur des données à caractère personnel rendues publiques par la personne concernée* » ou encore « *les traitements nécessaires à la recherche publique*[46] ». En tout état de cause, si le consentement n'est pas requis, l'information doit être délivrée « *d'une façon concise, transparente, compréhensible et aisément accessible, en des termes clairs et simples* » et le responsable du traitement doit prendre des mesures appropriées pour y parvenir. Ces informations sont fournies par écrit ou par d'autres moyens y compris, lorsque c'est approprié, par voie électronique (RGPD art. 12).

Tout État membre peut en outre introduire des conditions supplémentaires pour tenir compte de ses spécificités, à condition toutefois de ne pas porter atteinte à l'efficacité

[45] Ce comité « *a vocation à prendre la suite du groupe de l'article 29 (le G29), qui était l'enceinte informelle d'échanges et d'élaboration de doctrine instituée par la directive 95/46. Outre les avis formels qu'il sera appelé à rendre sur le fondement de l'article 64, et les décisions contraignantes qu'il rendra sur le fondement de l'article 65 en cas de litiges entre autorités, l'EDPB poursuivra le travail d'élaboration de la doctrine commune des autorités de protection des données de l'Union européenne au travers de lignes directrices, avis, etc.* » (voir site de la CNIL).

[46] Voir article 8 modifié par la loi n°2018-493 du 20 juin 2018. Dans la mesure où la finalité du traitement l'exige pour certaines catégories de données, ne sont pas soumis à l'interdiction prévue au I : suivent 11 catégories de traitement dont :
« *Les traitements nécessaires à la recherche publique au sens de l'article L. 112-1 du code de la recherche, mis en œuvre dans les conditions prévues au 2 de l'article 9 du règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 précité, après avis motivé et publié de la Commission nationale de l'informatique et des libertés rendu selon les modalités prévues à l'article 28 de la présente loi.* »

du règlement européen et de ne pas en altérer l'esprit. C'est le sens même de la modification, en date du 21 juin 2018, de la loi 78-17 du 6 janvier 1978 ; l'ordonnance n° 2018-1125 du 12 décembre 2018 achève, au niveau législatif, la mise en conformité du droit national avec le règlement général sur la protection des données (RGPD).

On assiste ainsi au passage progressif d'une volonté de contrôle exhaustif *a priori* à une logique d'intervention et de contrôle *a posteriori* fondée sur une recherche d'intelligibilité et de responsabilisation, qui exige des responsables du traitement une loyauté de comportement vérifiable.

Le CCNE ne peut qu'approuver, au plan éthique, cette démarche qui doit se fonder sur une relation de confiance entre le titulaire des données et ceux qui les recueillent, ceux qui y ont accès et ceux qui les traitent. La transparence reste une valeur fondamentale, érigée en principe par l'article 5 du règlement européen, mais peut paraitre parfois peu aisément conciliable avec le principe de fonctionnement sur lequel repose l'exploitation des données massives. Il faut en revanche rechercher une intelligibilité suffisante du processus d'utilisation des données mis en œuvre, ainsi que de ses conséquences possibles. On peut parfois douter de la mise en œuvre d'une telle démarche dans ce qu'elle a d'idéal au regard de la réalité des politiques de certains opérateurs. De plus, en cas de recours aux techniques de programmation fondées sur l'apprentissage profond, il devient impossible de suivre pas à pas le cheminement qu'emprunte la machine pour répondre à la question qui lui a été posée. On n'en est pas moins en droit d'attendre du responsable du traitement qu'il maîtrise la logique suivie et les paramètres pris en compte[47]. Renoncer à cette exigence reviendrait à renoncer à tout contrôle possible sur les biais dont peut être affecté le traitement algorithmique et, partant, à toute responsabilité possible des acteurs. Si l'on comprend que la complexité du processus ne permet pas, sauf exception, à un particulier de se livrer lui-même à ce type de contrôle, il devrait n'accorder sa confiance et ne consentir à l'utilisation de ses données personnelles que si celle-ci se fait dans le cadre d'une gouvernance identifiée par un responsable désigné et ayant pris des engagements clairs. Ces engagements doivent bien entendu pouvoir être vérifiés par une autorité de contrôle disposant du concours d'experts[48].

Il est essentiel de se doter de ces moyens de contrôle et d'assurer une large information de nos concitoyens pour qu'ils sachent :

[47] Conseil constitutionnel, décision N° 2018-765 DC du 12 juin 2018 : « *En dernier lieu, le responsable du traitement doit s'assurer de la maîtrise du traitement algorithmique et de ses évolutions afin de pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard. Il en résulte que ne peuvent être utilisés, comme fondement exclusif d'une décision administrative individuelle, des algorithmes susceptibles de réviser eux-mêmes les règles qu'ils appliquent, sans le contrôle et la validation du responsable du traitement.* »

[48] Voir le rapport de la CNIL : « Comment permettre à l'homme de garder la main ? Rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle », 15 décembre 2017, https://www.cnil.fr/fr

- les précisions qui doivent leur être fournies lorsqu'on les informe de l'utilisation ou de la réutilisation de leurs données personnelles,
- les engagements pris par celui qui fait la demande d'utilisation,
- les contrôles pouvant être exercés pour vérifier le sérieux de ces engagements et leur suivi sur le long terme.

Ce processus peut être de nature à provoquer une surcharge de travail lourde à supporter pour le système de santé, difficulté qui devra être anticipée, prise en compte et surmontée.

L'information doit être adaptée aux différents contextes d'utilisation (soin, recherche ponctuelle, participation à une base de données internationale, etc.) afin que le titulaire des données ait une compréhension suffisante des technologies utilisées et des termes employés alors qu'il peut se trouver dans une situation de vulnérabilité. Lorsque son consentement est requis, cette information doit lui offrir la possibilité de choisir librement. La complexité des usages et des contextes oblige à engager une réflexion sur la notion de consentement et sur de nouvelles modalités du recueil de ce consentement qui garantissent le respect des principes éthiques et les droits des personnes (voir chapitre 3.4.1) (RECOMMANDATION N° 1).

Ce sont là les conditions d'une confiance qui est nécessaire pour répondre à l'exigence éthique du respect de la dignité du titulaire des données personnelles. Elles font écho aux impératifs de loyauté et de vigilance mis en avant par la CNIL dans son rapport précité, déposé en décembre 2017 au terme du débat public qu'il lui avait été demandé d'animer sur les enjeux éthiques des algorithmes et de l'intelligence artificielle.

Même si cette démarche n'est pas spécifique à la santé, elle trouve tout son intérêt dans ce domaine, réserve faite des particularités que nous examinerons dans la troisième partie, notamment en matière de recherche et de soin.

Encore faut-il veiller à ce que cette confiance soit réciproque. Cette condition ne serait pas remplie si le titulaire des données, après avoir consenti à leur utilisation dans des domaines intéressant la collectivité, notamment la recherche médicale, pouvait faire un usage abusif de son droit d'effacement, au risque de fragiliser, voire de remettre en cause, un projet de recherche. La contrepartie de l'obligation de poursuivre dans le temps l'information sur les résultats du traitement devrait être la subordination du droit d'effacement à un motif légitime, tel qu'une mauvaise utilisation avérée.[49]

[49] C'est pour répondre à cette préoccupation que le RGPD exclut le droit à l'effacement lorsque 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 (article 17-3-d). Dans cette hypothèse, doivent être prises des garanties appropriées pour les droits

Si le règlement européen et la législation dérivée visent ainsi à établir un système ambitieux de protection, qui tend à devenir une référence au-delà même de l'Union européenne, ils ne sont pas encore pleinement mis en œuvre et ils n'atteindront leur but que si les principes protecteurs qu'ils prévoient sont effectifs. C'est là un enjeu majeur d'un point de vue éthique.

Compte tenu du rythme particulièrement important des innovations scientifiques et technologiques et des évolutions qu'elles déterminent dans le recueil et l'exploitation des données relatives à la santé, le CCNE estime qu'il est nécessaire d'évaluer périodiquement la mise en œuvre effective des dispositifs juridiques, afin de vérifier le maintien dans le temps de l'efficacité du système de protection des données personnelles qu'ils instaurent.
(RECOMMANDATION N° 2).

Deux points justifient sur ce point une particulière vigilance :

- Le règlement européen ne s'applique qu'aux données dites personnelles et ne concerne pas les informations anonymes, entendues comme celles qui ne concernent pas une personne physique identifiée ou identifiable. Pour déterminer si une personne physique est identifiable (voir note 9), il convient de prendre en considération l'ensemble des moyens raisonnablement susceptibles d'être utilisés par le responsable du traitement ou par toute autre personne pour identifier la personne physique directement ou indirectement, tels que le ciblage (considérant 26). Cette notion de « *moyens raisonnablement susceptibles d'être utilisés* » pourrait constituer un critère peu stable dans un monde marqué par un puissant mouvement continu d'innovations technologiques[50].

- Le règlement ne s'applique pas aux traitements de données à caractère personnel effectués par une personne physique au cours d'activités strictement personnelles ou domestiques et donc sans lien avec une activité professionnelle ou commerciale[51]. En revanche, le règlement s'applique aux responsables du traitement ou aux sous-traitants qui fournissent les moyens de traiter des données à caractère personnel pour de telles activités personnelles ou domestiques (considérant 18).

---

et libertés de la personne concernée, telles que la minimisation des données et leur pseudonymisation (article 89-1).

[50] Il est sans doute concevable que des données un temps anonymes redeviennent à caractère personnel du fait de l'évolution des techniques d'identification, ce qui les ferait alors retomber dans le champ d'application du RGPD.

[51] Ces activités pourraient inclure l'échange de correspondance et la tenue d'un carnet d'adresses, ou l'utilisation de réseaux sociaux et les activités en ligne qui ont lieu dans le cadre de ces activités.

2.2 Justice : la solidarité au défi de l'individualisation

En France, le financement de la santé est assuré en grande partie par la sécurité sociale et particulièrement les caisses d'assurance maladie, et repose sur le mécanisme de la mutualisation des risques sanitaires. Celle-ci est l'une des manifestations fondamentales de la solidarité, valeur éthique essentielle de notre système de santé puisqu'elle contribue à la mise en œuvre de nos principes fondateurs d'égalité et de fraternité.

L'organisation de la prise en charge des soins est très encadrée et repose sur deux types de couvertures : l'assurance maladie obligatoire et l'assurance maladie complémentaire.

La première reste indépendante des revenus et de l'âge, et considère le seul coût des traitements, qui ne doit pas être un obstacle au soin ; elle est complétée par des assurances complémentaires (un peu moins de 14 % de la prise en charge) qui sont segmentées par catégories (revenus, âge) et qui introduisent des sous-groupes. Cependant la prise en compte de ces coûts cache certaines disparités entre types de soins. Pour certains d'entre eux (optique, prothèse dentaire), la prise en charge par les organismes complémentaires devient majoritaire, et peut laisser un « reste à charge » important.

Pour les acteurs qui doivent supporter le coût toujours plus lourd de la prise en charge des dépenses de santé, la tentation est forte de procéder à une personnalisation du risque afin de parvenir à une meilleure maîtrise économique. Également à la base du contrat d'assurance (les assureurs interviennent de manière importante en matière de santé), la mutualisation des risques repose sur l'aléa. Chacun cotise une somme relativement peu élevée et le grand nombre de cotisants fait que les produits ainsi obtenus servent à faire face aux frais, parfois très importants, qui résultent, pour un bien plus petit nombre, de la réalisation du risque assuré. « *Même si l'attachement français aux principes de non-discrimination et de mutualisation du risque est fort, l'émergence de la médecine prédictive pourrait le fragiliser. La connaissance fine des risques individuels, si elle se révélait possible, pourrait ainsi entraîner un délitement ou du moins un affaiblissement de la solidarité et de la couverture mutualisée du risque santé* » (avis 124 du CCNE).

La question que pose en matière de santé la difficile conciliation entre autonomie et liberté individuelle d'une part, égalité d'accès aux des soins et solidarité nationale d'autre part, n'est pas nouvelle[52]. Mais les données massives et leur exploitation algo-

[52] Paul Ricoeur l'a exprimé en ces termes : « *En dernière analyse ce conflit sur le front de la santé publique n'a rien d'étonnant. On pourrait récrire le contrat médical dans les termes d'une série de paradoxes. Premier paradoxe : la*

rithmique, en particulier par les techniques dites d'intelligence artificielle, éclairent cette question d'un jour nouveau.

2.2.1 Une valeur confrontée à une situation nouvelle : la solidarité nationale face à l'individualisation du risque médical

De même qu'une ambivalence intrinsèque du numérique fait qu'une intervention trop rigoureuse destinée à en prévenir les aspects négatifs risque d'en entraver le potentiel positif, de même l'exploitation des données massives dans le domaine de la santé révèle un conflit de valeurs et d'intérêt entre les bénéfices pouvant être espérés par l'individu et ceux qui pourraient résulter, pour la collectivité, d'une approche intrusive visant à favoriser une meilleure rationalité économique.

La disponibilité et le croisement de bases de données – cliniques, biologiques, génomiques, imagerie, environnementales – permettent d'analyser l'état de santé des personnes avec une précision allant jusqu'à l'individualisation : chaque individu est un patient unique du fait des particularités de sa maladie ou de son risque. Cela non seulement en cas de maladie déclarée dont on peut affiner le diagnostic et donc adapter très précisément la thérapeutique, facteur d'amélioration de la prise en charge (médecine personnalisée), mais aussi – hors situation de maladie – pour l'analyse du risque. La voie est ainsi ouverte à une démarche de prédiction et de prévention, visant à prévenir l'émergence de la maladie.

Comment éviter que cette connaissance plus précise du risque et cette personnalisation thérapeutique n'induisent – selon une logique économique – des discriminations et une remise en cause de la solidarité nationale, qui représente l'exigence du vivre ensemble et repose sur l'aléa ? Celui-ci peut être fragilisé dès lors qu'une action de prévention du risque est connue et peut être imposée, ou que la prédiction d'un pronostic défavorable incite à minimiser la prise en charge pour des raisons économiques.

Comme nous le détaillerons plus loin au chapitre 3, l'évolution des tests génétiques est emblématique de cette menace, en ce qu'ils sont des marqueurs prédictifs puis-

---

personne humaine n'est pas une chose, et pourtant son corps est une partie de la nature physique observable. Deuxième paradoxe : la personne n'est pas une marchandise, ni la médecine un commerce, mais la médecine a un prix et coûte à la société. Dernier paradoxe qui recouvre les deux précédents : la souffrance est privée, mais la santé est publique. Il ne faut donc pas s'étonner si ce conflit sur le front de la santé publique ne cesse de s'aggraver, vu le coût de plus en plus élevé de la recherche en biologie médicale, vu celui d'explorations du corps humain et d'interventions chirurgicales hautement sophistiquées, le tout aggravé par l'allongement de la vie humaine, pour ne rien dire des attentes déraisonnables d'une opinion publique qui demande trop à un corps médical dont elle redoute par ailleurs les abus de pouvoir. En bref, le fossé ne peut que se creuser entre la revendication d'une liberté individuelle illimitée et la préservation de l'égalité dans la distribution publique des soins sous le signe de la règle de la solidarité. » (Les trois niveaux du jugement médical – le contrat médical *Esprit* No. 227 - Décembre 1996, pp. 21-33).

sants. C'est ainsi que, dès le 30 octobre 1995, dans son avis n° 46 « Génétique et médecine, de la prédiction à la prévention », le CCNE notait : « *Les tests génétiques apportent des informations sur l'identité des personnes et soulignent leur diversité qui contribue à la richesse de l'humanité. L'utilisation de ces informations à des fins de sélection ou de discrimination dans la vie sociale et économique, que ce soit dans le domaine des politiques de santé, de l'emploi ou des systèmes d'assurance, conduirait à franchir une étape d'une extrême gravité vers la mise en cause des principes d'égalité en droits et en dignité, et de solidarité entre tous les êtres humains, sur lesquels repose notre société. Le CCNE insiste sur la nécessité de respecter ces droits fondamentaux, quelle que soit la finalité de l'utilisation des tests génétiques. Il y va des droits de l'homme* ».

2.2.2 Une situation nouvelle qui oblige à une autre perception des enjeux

Le profilage, en ce qu'il donne une meilleure connaissance de chacun, offre aussi d'utiles perspectives pour aider l'État à améliorer sa gestion de l'économie de la santé.

Mais, par nature même, l'individualisation du risque peut porter atteinte à la mutualisation. Elle fournit une justification à des avantages tarifaires consentis à ceux qui ne présentent pas un risque élevé, au détriment des autres. Elle pousse en outre à agir sur les individus pour les inciter à des comportements susceptibles de prévenir ou de limiter l'importance des maladies qui les menacent. Or, c'est une chose que de contribuer à l'éducation générale de la population par des campagnes incitant à une meilleure hygiène de vie, c'en est une autre que de cibler des personnes ou des groupes déterminés à raison des risques qu'ils présentent. La prévention changerait alors de nature. Elle perdrait son caractère éducatif pour devenir intrusive si elle s'accompagnait d'une surveillance délibérée, par exemple à l'aide d'objets connectés. Elle perdrait son caractère bienveillant si elle instaurait un régime de pénalités ou de récompenses venant sanctionner les résultats obtenus[53]. Une telle dérive menacerait non seulement le principe de solidarité, mais aussi la protection et le soin qu'une politique de santé doit assurer aux personnes les plus vulnérables. Elle porterait une grave atteinte aux libertés individuelles en ne s'adressant pas au libre arbitre des individus mais en exerçant sur eux une contrainte inappropriée. La différenciation entre personnes à « bons risques » et à « mauvais risques » ne pourrait chercher à se justifier qu'en mettant en avant les mérites de ceux qui consentent des efforts pour avoir une hygiène de vie satisfaisante et en stigmatisant ceux ayant une conduite jugée irresponsable. Au-delà même du caractère en soi inacceptable d'une telle classification, la contrainte qu'elle instaure serait injuste car on ne saurait oublier que l'importance des

[53] C'est déjà le cas pour certaines assurances complémentaires (voir note 56).

risques ne tient pas seulement aux comportements individuels, mais aussi à des facteurs que nous ne maîtrisons pas et qui tiennent à l'hérédité et aux accidents de la vie. On ne peut pas davantage faire abstraction de l'incidence du milieu social, dont toutes les études montrent qu'il est déterminant pour l'état de santé et l'espérance de vie des personnes[54].

Dès lors que la logique de l'individualisation des risques de santé tend à établir des corrélations entre les personnes en fonction de critères de comportement, ainsi que de choix et d'hygiène de vie, la communication mondialisée permise par internet et les réseaux sociaux favorise le développement d'affinités et d'un sens de la solidarité entre ceux qui se reconnaissent comme semblables, quelle que soit leur localisation géographique, plutôt qu'entre des concitoyens ; cela peut rendre plus difficile l'acceptation de l'espace politique national comme niveau de référence de la solidarité et des politiques de santé publique. C'est pourtant dans ce cadre national que sont financés les systèmes de santé et que s'exerce principalement la solidarité face à la maladie et aux aléas de la vie.

2.2.3 Comment conserver la maitrise par la solidarité nationale ?

Aujourd'hui en France, l'impératif éthique d'égalité d'accès aux soins interdit aux caisses d'assurance maladie, instituées par le principe de l'assurance maladie publique obligatoire, d'opérer une sélection des risques en matière de santé. Mais, si le coût des maladies aiguës graves est entièrement assuré, des pratiques discriminantes pourraient être recherchées pour les maladies mineures, ou pour celles qui deviennent chroniques et qui concentrent une part importante des dépenses de santé[55]. « *Pourrait ainsi s'estomper la distinction classique entre le risque subi né d'une hérédité non encore contrôlable et le risque choisi relevant du mode de vie.* » (CCNE avis 124).

Le risque d'établissement d'un profil individuel de santé tiendrait moins à une possible inégalité d'accès qu'à une obligation de soin au nom d'une plus grande efficience du système de santé. Le bénéfice attendu par la collectivité serait de réduire les dépenses occasionnées par un diagnostic et un traitement rendus plus tardifs.

Une tentation de discrimination tarifaire en matière d'assurance est un problème réel car elle repose sur une logique économique puissante. L'interdiction de sélection peut

[54] Notamment : INSEE première, N° 1372, 5 octobre 2011 : « l'espérance de vie s'accroît, les inégalités sociales face à la mort demeurent »

[55] L'idée que le comportement rétroagit sur la prise en charge sanitaire n'est pas exclue : par exemple, la prise en charge de la prestation de santé à domicile dans l'apnée du sommeil est soumise à l'observance au traitement de la personne atteinte, évaluée par télésuivi.

être contournée par les assurances privées, notamment parce que les données massives démultiplient les possibilités de profilage, même sans avoir recours aux questionnaires médicaux ou aux données de santé. La disponibilité d'informations sur le comportement quotidien des assurés incite déjà les assureurs privés à établir des partenariats avec des entreprises leur permettant de moduler leurs primes en accordant des « récompenses » à ceux dont le comportement est jugé responsable[56].

Le conflit entre logique économique et intérêt individuel d'une part, exigences du vivre ensemble et de la solidarité d'autre part, ne peut être arbitré que par l'action politique. Seule la loi peut fixer des limites à l'individualisation des risques et édicter les règles nécessaires à la préservation de la solidarité nationale.

Il serait particulièrement utile qu'une même démarche soit suivie à un échelon supranational et notamment dans le cadre de l'UE.

Le CCNE estime que doivent être instaurées les conditions d'une vigilance exercée de manière collégiale par tous les acteurs de santé, afin de s'assurer que les logiques de personnalisation, qui peuvent être bénéfiques, ne transgressent pas les valeurs d'équité et de solidarité en évoluant vers un profilage de nature discriminatoire, notamment pour des raisons économiques. (Voir RECOMMANDATION N° 8).

2.3 Non-nuisance et bienfaisance : les données massives, un facteur d'innovation en santé mais un risque de nuire si la qualité des données n'est pas assurée

2.3.1 La qualité des soins et l'accès à l'innovation confrontés à une situation nouvelle

Ne pas tirer parti de la collecte et de l'analyse des données des patients (ou des individus sains participant à une recherche) avec les technologies informatiques dont on dispose au bénéfice de la santé du patient concerné comme de la collectivité ne serait pas éthique. Ignorer les risques de nuire qui peuvent résulter de cette démarche et ne pas chercher à les réduire serait tout autant contraire à l'éthique[57].

---

[56] En 2014, Axa s'était alliée à la startup *Withings*, depuis rachetée par *Nokia Health*, pour proposer un bracelet connecté à ses assurés : l'assureur offrait par exemple des chèques-cadeaux si l'assuré effectuait 7 000 pas par jour. Audiens (assureur des professionnels de la culture) et le constructeur de montres et bracelets intelligents *Garmin* ont proposé une offre destinée aux TPE (très petites entreprises) à l'occasion de la généralisation de la complémentaire santé.

[57] « *The obligations of beneficence affect both individual investigators and society at large, because they extend both to particular research projects and to the entire enterprise of research. In the case of particular projects, Investigators and members of their institutions are obliged to give forethought to the maximization of benefits and the reduction of risk that might occur from the research investigation. In the case of scientific research in general, members of the larger society are obliged to recognize the longer term benefits and risks that may result from the improvement of knowledge and from the development of novel medical, psychotherapeutic, and social procedures* ». (Rapport Belmont : principes éthiques et directives concernant la protection des sujets humains dans le cadre de la recherche - 1974).

Les innovations en matière de santé issues du traitement des données seront probablement multiples, même si elles se heurtent encore à d'importants obstacles techniques (notamment pour les techniques qui recourent à l'apprentissage machine) dans la pratique courante[58]. Parmi les innovations attendues, on peut mentionner la prédiction du risque grâce aux données génomiques, la détection précoce de signaux d'alerte, la surveillance thérapeutique, la classification plus précise de maladies[59] voire l'aide au diagnostic[60], l'aide à la gestion des flux de patients. Les perspectives dans le domaine de la santé publique et l'organisation du système de santé sont tout aussi importantes. Elles peuvent aider à limiter le rythme actuel d'évolution des dépenses de soin qui sera difficilement soutenable dans un contexte marqué par une croissance économique faible et un vieillissement marqué de la population. Sous réserve de pouvoir recouper ces différentes données et en extraire des indicateurs de santé, l'épidémiologie, qui étudie, *via* ces indicateurs, les facteurs de risque et les maladies dans la population, constituerait un puissant instrument pour une politique publique de prévention sanitaire et d'amélioration des décisions thérapeutiques. À ce titre, le développement des objets connectés peut être utile : l'intérêt des français pour le « *quantified self* », nouvelle pratique de « mesure de soi » permise par le numérique, en donnant aux individus la possibilité de mesurer leurs actions, d'en étudier les conséquences et de constater leurs progrès au jour le jour, constitue une réelle opportunité d'améliorer les comportements sanitaires si l'État se donne les moyens d'investir dans une politique de prévention par le numérique. Ces applications seront illustrées au chapitre 3.

Toutefois, deux facteurs freinent actuellement cette évolution :

- la collecte et l'échange des données ne sont pas suffisamment développés en France en raison de la dispersion des entrepôts de données, des registres et cohortes de patients, des lacunes de l'interopérabilité des systèmes d'information et d'une mutualisation insuffisante des sources ; pour pallier cette difficulté, deux projets de plateformes mutualisées regroupant à l'échelon national des données massives dans les domaines de la génomique (plan France médecine génomique 2025[61]) ou des données de santé collectées dans

[58] Yu Kh, Kohane IS. Framing the challenges of artificial intelligence in medicine. *BMJ Quality & Safety* 2019, 28 : 238-41. Weintraub WS, *et al*. Translational medicine in the era of Big Data and machine learning. *Circulation Research* 2018 ; 123 : 1202-4.

[59] Rumsfeld JS, *et al.* Big data analytics to improve cardiovascular care: promise and challenges. *Nature Reviews Cardiology* 2016 ; 13 : 350-9.

[60] Citons les algorithmes d'intelligence artificielle pour le diagnostic du cancer, de la dépression, la prise en charge de de la douleur chronique, la prédiction du suicide, ou encore l'aide aux prescriptions diététiques chez les patients diabétiques. *Nature Medicine* 2018, 24, 1304-5.

[61] https://www.gouvernement.fr/sites/default/files/document/document/2016/06/22.06.2016_remise_du_rapport_dyves_levy_-_france_medecine_genomique_2025.pdf

le cadre des soins ou de leur remboursement (*Health data hub*[62]) devraient être prochainement mises en place ;

- les difficultés d'accès aux technologies de communication de certaines personnes pour des raisons soit de mauvaise couverture géographique, soit de conditions socio-économiques défavorables, qui créent les conditions de ce que l'on a appelé une « fracture numérique » (*73 % de la population possède un smartphone [99% des 18-24 ans], et 94 % un téléphone mobile ; la moitié des 6 % qui n'en possèdent pas sont des personnes âgées de 70 ans ou plus, dont un tiers avec des bas revenus* ; *10 % de la population âgée de 12 ans et plus n'a ni ordinateur, ni smartphone, ni tablette*[63]*) ;* cette difficulté d'accès est d'autant plus sensible qu'elle touche les populations fragilisées qui auraient le plus besoin d'un accompagnement régulier. L'observatoire des inégalités notait ainsi en 2016 « *Reste que 15 % de la population n'utilise pas Internet et que la moitié ne fréquente pas les réseaux sociaux. Ce taux atteint 25 % pour les plus démunis (foyers dont les revenus médians sont d'environ 1 200 euros mensuels) et 43 % pour les non-diplômés. […] La fracture qui persiste est surtout générationnelle : les plus anciens restent à l'écart d'un univers dont ils ne voient pas vraiment l'utilité et qu'ils ne comprennent pas toujours. Pourtant, la référence permanente à Internet et aux réseaux sociaux en particulier, comme s'il allait de soi qu'ils sont fréquentés par tous, constitue une violence symbolique pour ceux qui n'ont pas les moyens d'y participer.*[64]».

Le CCNE insiste sur la nécessité de veiller à ce que les personnes qui n'ont pas accès aux technologies du numérique, pour des raisons économiques ou de difficulté à comprendre leur mode de fonctionnement, bénéficient, comme les autres, des avancées dans le domaine de la santé et ne subissent ni pénalisation ni discrimination dans leur accès aux soins (<u>RECOMMANDATION N° 9</u>).

[62] Le *Health data hub* – ou plateforme d'exploitation des données de santé - pourrait être « *l'instrument de l'état au service d'une ambition : mettre le patrimoine des données de santé financées par la solidarité nationale au service du patient et du système de santé dans le respect de l'éthique des droits fondamentaux de nos concitoyens. […] Il constituerait un guichet unique afin de faciliter l'accès aux données. […] Il serait garant de la qualité des données partagées et représenterait un tiers de confiance pour le partage des données dans le respect des droits des patients* ». (voir rapport https://solidarites-sante.gouv.fr/ministere/documentation-et-publications-officielles/rapports/sante/article/rapport-health-data-hub-mission-de-prefiguration). 12 octobre 2018.

[63] La totalité des 18-24 ans et 98 % des 25-39 ans déclarent posséder un téléphone mobile. Cette proportion chute à 76 % chez les plus âgés, 70 ans et plus. Aujourd'hui, 6 % de la population française âgée de 12 ans et plus n'a pas de téléphone portable personnel. Cette absence d'équipement concerne plus souvent des personnes âgées (24 % des plus de 70 ans n'ont pas de téléphone mobile), des non diplômés (22 %), des retraités (16 %) et des personnes seules (12 %). (Baromètre du numérique, Credoc 2017).

[64] Observatoire des inégalités. Qui a eu son iPhone 7 à Noël ? Décembre 2016.

2.3.2 Une situation nouvelle qui oblige à une autre perception des enjeux : la préservation d'une maîtrise humaine pour assurer la fiabilité des données et des décisions induites du traitement des données massives

Dans l'exercice classique de la médecine, la démarche s'appuie sur l'écoute du patient, l'analyse de données physiques, biologiques et d'imagerie, alliées à l'éducation et à l'expérience antérieure du médecin. La relation médecin-malade se fonde sur « *un contrat de confiance établi sur une information honnête et personnalisée, aboutissant à des décisions prises en commun et véritablement partagées*[65] ». Dans un exercice médical plus innovant, il peut être ajouté le recours à l'usage d'algorithmes d'aide à la décision fondés sur un processus d'apprentissage reposant sur le traitement de données massives.

Celles-ci se définissent classiquement par le sigle 3 V (*volume, variété, vitesse*) auxquels on a ajouté le V de « *véracité* » et le V de « *valeur* ». Comme nous l'avons déjà évoqué lorsque nous avons examiné les enjeux éthiques de la protection de la personne et du respect de la vie privée, l'enjeu est celui de l'intelligibilité et du contrôle, qui doit porter tant sur la qualité des données elles-mêmes que sur celle des conclusions tirées par le système de leur exploitation.

Comment dès lors assurer la personne (ou la collectivité) de la fiabilité des résultats obtenus au regard des connaissances scientifiques connues ? Le risque majeur est celui d'un biais ou d'une erreur (involontaire ou par malveillance) dans la collecte, l'annotation ou le traitement des données en amont[66] (voir chapitre 3, paragraphe 3.3 sur les données génomiques). Il aboutirait à une information inexacte en aval, entraînant une décision erronée ou une discrimination (diagnostic faussé, comportement ou protocole de soin inadapté, organisme de recherche ou administratif incité à des conclusions inappropriées).

Pour prévenir ce risque, il est essentiel – particulièrement dans le domaine de la santé – qu'il y ait une « garantie humaine »[67,68] pour répondre de la rigueur méthodologique

[65] CCNE, avis 58. « Consentement éclairé et information des personnes qui se prêtent à des actes de soin ou de recherche » (12 juin 1988).
[66] The future of biocuration. *Nature* 2008 ; 455 : 47.
[67] Le principe d'une « garantie humaine » a été proposé par le groupe de travail à l'origine du rapport « *numérique et santé : quels enjeux éthiques pour quelles régulations* » commandé par le CCNE et rendu public en novembre 2018 ; le terme a été repris dans l'avis 129 du CCNE. Chapitre « numérique et santé », pp. 94-106.
[68] Le Conseil constitutionnel s'est prononcé sur le seul recours à un algorithme pour traiter les données de santé dans sa décision précitée N° 2018-765 DC du 12 juin 2018 : « *Enfin, le recours exclusif à un algorithme est exclu si ce traitement porte sur l'une des données sensibles mentionnées au paragraphe I de l'article 8 de la loi du 6 janvier 1978, c'est-à-dire des données à caractère personnel « qui révèlent la prétendue origine raciale ou l'origine ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale d'une personne physique », des données génétiques, des données biométriques, des données de santé ou des données relatives à la vie sexuelle ou l'orientation sexuelle d'une personne physique* ».

de ces trois étapes que sont la qualité des données, l'adéquation des traitements algorithmiques à la question posée, et la vérification sur un jeu de données indépendantes de la robustesse et de l'exactitude du résultat donné par l'algorithme[69]. Cette garantie est d'autant plus importante que le préjugé d'infaillibilité et d'objectivité que l'on accorde généralement aux analyses fondées sur des modèles informatiques et mathématiques induit une confiance excessive qui peut accroître le risque d'erreur.

Cette garantie humaine doit aussi s'exercer de façon répétée, car, outre les biais mentionnés, se pose la question de la validité et de la pérennité de décisions déduites d'algorithmes entraînés avec des données rétrospectives, souvent sélectionnées à dessein. Les résultats obtenus pourraient être fragilisés par le flux constant de nouvelles données intégrées automatiquement aux processus d'apprentissage[70].

Un défaut de garantie humaine du processus pourrait avoir deux conséquences majeures :

- le risque qu'un biais ou une erreur méthodologique, fondés sur le traitement des données personnelles, aboutissent à une conclusion ou une décision erronée ayant des conséquences directes pour la personne ;
- le risque d'une inconséquence et d'une erreur de jugement si on ne peut pas vérifier le résultat obtenu ou contrôler le cheminement qui y a conduit, alors même que sont en cause des données personnelles ; ce risque serait d'autant plus inacceptable qu'il porterait sur des décisions sur lesquelles les patients doivent bénéficier d'un droit d'information, auxquelles ils doivent pouvoir consentir ou qu'ils doivent pouvoir refuser, et auxquelles ils doivent être associés en toute connaissance de cause. (Voir **RECOMMANDATION N° 4**).

2.3.3 Comment assurer maîtrise et contrôle ?

Ces risques peuvent être plus efficacement maitrisés en matière de recherche clinique et de soin (ou d'objets connectés reconnus comme « dispositifs médicaux »), malgré l'hétérogénéité des situations, parce que ceux qui appliquent les informations déduites du traitement des données sont des acteurs de santé (médecins, personnel soignant, chercheurs) qui agissent dans un cadre professionnel. Ils sont encadrés par des règles déontologiques strictes, telles que le secret médical, la validation des protocoles thérapeutiques ou celles qui régissent la responsabilité médicale. Les données

[69] Beam AL, Kohane IS. Translating artificial intelligence into clinical care. *JAMA* 2016 ; 316 : 2368-9 ; Big Data and machine learning in health care. *JAMA* 2018 ; 319 : 1317-8.
[70] KH Yu, IS Kohane. Framing the challenges of artificial intelligence in medicine. *BMJ Quality & Safety* 2019 ; 28 : 238-41.

elles-mêmes sont des données de santé au sens de l'article L 1111-8 du code de la santé publique, et nous avons vu dans la première partie qu'elles bénéficient d'une protection particulière. Il ne faut toutefois pas exclure qu'une conclusion erronée puisse être déduite de l'analyse de données, par exemple si celles-ci n'intègrent pas les données de certaines populations minoritaires (voir chapitre 3.3 données génomiques). Pourrait aussi apparaître la tentation de biaiser l'apprentissage algorithmique de façon à privilégier l'évaluation favorable de certains établissements de santé.

Cette question de la véracité des mesures et des données est encore plus aiguë dans le cas des nouvelles applications proposées aux consommateurs par les opérateurs privés. Leur développement n'est pas encadré et leur efficacité n'est évaluée scientifiquement, ni pour les algorithmes utilisés, ni *a fortiori* pour les conclusions déduites et délivrées à distance, qui peuvent être partisanes ou sponsorisées.

Face aux risques ainsi mis en évidence, le danger existe d'une logique de refus de tout protocole ne reposant pas sur des techniques éprouvées, ce qui exclurait dans la plupart des cas l'exploitation des données massives. C'est au contraire au prix d'une ouverture et d'un partage des connaissances – condition d'une démarche scientifique véritablement fructueuse (voir ci-dessous : chapitre 3.2) – que l'on peut espérer faire progresser la médecine et multiplier les perspectives. C'est pour répondre à cette exigence que le rapport Villani[71] a préconisé la création d'une plateforme d'accès et de mutualisation des données pertinentes pour la recherche et l'innovation en santé (*Health data hub*[72]), dont la mise en œuvre devrait être actée au premier semestre 2019[73].

Le contrôle de la qualité du traitement algorithmique des données numériques impose des actions dans plusieurs domaines :

- *Une ambitieuse action de formation*. À cet effet, le CCNE estime que les professionnels de santé doivent bénéficier, lors de leur formation initiale et tout au long de leur carrière, d'une formation adaptée aux technologies numériques, aux principes éthiques et juridiques qui régissent le recueil et le traitement des données, aux moyens notamment techniques à mettre en œuvre pour les respecter, et aux risques de biais et d'atteinte à la confidentialité et au respect des droits des personnes qui résulteraient de leur méconnaissance.

[71] « Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne », rapport du 28 mars 2018. https://www.ladocumentationfrancaise.fr/rapports-publics/184000159/index.shtml

[72] Rapport téléchargeable sur https://solidarites-sante.gouv.fr/IMG/pdf/181012_-_rapport_health_data_hub.pdf

[73] Elle est inscrite dans le projet de loi relatif à l'organisation et à la transformation du système de santé, titre III : Développer l'ambition numérique en santé. (Adopté en première lecture à l'Assemblée nationale le 26 mars 2019, en cours de discussions parlementaires, mai 2019).

Les experts de la gestion et de l'analyse des données massives (*data scientists*), ainsi que les chercheurs, doivent être suffisamment avertis des questionnements éthiques que soulèvent ces technologies pour pouvoir assurer efficacement la protection des droits fondamentaux et des libertés individuelles. (RECOMMANDATION N° 5).

- *Une évaluation qualitative*. En effet, la multiplication de sites et d'applications qui donnent, hors parcours de soin, des conseils pour améliorer l'hygiène de vie et le bien-être, pose la question de la rigueur avec laquelle ils recueillent, interprètent et traitent des données relatives à la santé. Le CCNE considère que ces sites et applications doivent pouvoir être évalués ainsi que la qualité de l'information délivrée aux utilisateurs, afin d'éviter que certaines démarches insuffisamment rigoureuses ne puissent avoir des conséquences négatives sur le comportement et la santé des personnes. (RECOMMANDATION N° 6).

  Dans cette perspective, la HAS (haute autorité de santé) a récemment publié un référentiel destiné à fixer les bonnes pratiques concernant les applications mobiles[74]. En sa qualité d'autorité de contrôle nationale, la CNIL devrait jouer un rôle déterminant de coordination dans ce domaine.

- *Une recherche scientifique de haut niveau* en informatique et mathématique fondamentales afin de relever le défi de l'« explicabilité ». Il faut en effet permettre à un humain de comprendre les étapes fondamentales d'algorithmes fondés notamment sur l'apprentissage algorithmique, et donc les performances de la machine pour parvenir à un résultat. Ce processus s'apparente encore aujourd'hui à une boîte noire, ce qui handicape les applications dans le domaine de la santé.

3. QUELS PRINCIPES D'ACTION EN FONCTION DES DIFFÉRENTS CONTEXTES ?

3.1 Exploitation des données massives pour l'innovation en santé dans le cadre du soin

Qu'il s'agisse du soin, de la recherche, de la prévention ou des politiques de santé publique, le but de la médecine est de soulager les souffrances et de tout mettre en œuvre pour favoriser la santé et le bien-être de la population. Cette relation de soin, fondée sur une relation humaine directe, basée sur la confiance et un ensemble de décisions véritablement partagées, est soumise à une déontologie stricte, incluant le secret médical ; elle est contrôlée par les ordres professionnels.

[74] Référentiel de bonnes pratiques sur les applications et les objets connectés en santé. Haute autorité de santé, octobre 2016. https://www.has-sante.fr/portail/jcms/c_2681915/fr/referentiel-de-bonnes-pratiques-sur-les-applications-et-les-objets-connectes-en-sante-mobile-health-ou-mhealth

Le numérique a fait évoluer la relation soignant-soigné depuis longtemps. C'est ainsi que l'informatisation des lieux de soin a donné lieu à la constitution de bases de données pour l'exercice quotidien. Le médecin accède en ligne à un savoir devenu accessible et mis à jour à chaque instant, à des systèmes experts et à des logiciels d'aide à la prescription, qui sont certifiés par le référentiel de la HAS. La loi du 16 janvier 2016 de modernisation de notre système de santé a introduit le principe du partage de l'information de santé entre les professionnels qui interviennent à différents stades du parcours de soin du patient, sous le contrôle de celui-ci, de plus en plus souvent par le truchement de son dossier médical partagé (DMP), et, demain, de son espace numérique de santé[75]. Les données personnelles recueillies dans le cadre de cette relation de soin, ou *via* les données médico-administratives, sont considérées comme très sensibles, et bénéficient à ce titre d'une protection particulièrement importante[76] (voir supra paragraphe 1.3.2).

3.1.1 Qu'est ce qui change avec les données massives ?

La capacité de traiter des données et d'en tirer de nouvelles informations à une échelle que seule la machine rend possible – un des éléments de la rupture induite par les données massives que nous avons examinée au chapitre 1.1 - entraîne une nouvelle démarche d'acquisition du savoir scientifique. Elle ne repose plus uniquement sur le test de théories et de modèles qu'il s'agit de vérifier, mais sur la production de nouvelles hypothèses et connaissances à partir de données accumulées sans qu'elles aient été sélectionnées pour répondre à une interrogation (ou une intention) explicitement énoncée (démarche inductive). C'est déjà ce que faisaient les études comme celle de Framingham[77], mais il s'agissait alors exclusivement de « *human learning* » et non de « *machine learning* » ; apprentissage où la part de l'humain est plus ou moins réduite, et dont le cheminement est actuellement souvent opaque. Ce dernier élément – le passage d'un traitement humain des données à un traitement par la machine – rend inévitable l'intervention d'opérateurs privés de l'internet pour le recueil, le stockage et l'utilisation des données intéressant la santé.

---

[75] La création de l'espace numérique de santé est une disposition du projet de loi relatif à l'organisation et à la transformation du système de santé, titre III : Développer l'ambition numérique en santé. (En cours de discussions parlementaires, mai 2019).

[76] Les conditions de partage des données de santé, le droit d'opposition, d'information dont bénéficie le patient ainsi que le recueil de son consentement varient selon que les professionnels qui partagent ces données font partie ou non de la même équipe de soins, et sont précisées par les règles d'échange et de partage fixées à l'article L.1110-4 et à l'article R1110-3 du code de la santé publique. .

[77] *L'objectif principal de l'enquête de Framingham qui débute en 1949 était de mieux comprendre et observer le développement des maladies cardio-vasculaires. La cohorte - une population géographique non sélectionnée – était constituée initialement d'un tiers des habitants - entre 30 et 60 ans - de la ville de Framingham aux États-Unis. Elle se poursuit et intègre aujourd'hui la troisième génération. Mahmood SS,* et al. Lancet *2014 ; 383 : 999-1008.*

De cette rupture, on peut induire plusieurs innovations prévisibles même si la pratique quotidienne des professionnels de santé n'a pas encore réellement changé et que certains tempèrent la promesse de bénéfices majeurs rapides pour la santé émanant de ces technologies[78]. Des innovations sont attendues essentiellement dans quatre domaines[79] :

- la prédiction des risques de mortalité ou d'accidents grâce à l'amélioration de la connaissance des facteurs de risque et à la détection précoce de signaux d'alerte[80] ;
- l'amélioration de la pharmacovigilance grâce au dépistage en vie réelle des effets indésirables des médicaments[81] ;
- une classification plus précise des maladies[82] ;
- un affinement des traitements ;
- une automatisation et un affinement de l'interprétation, notamment en matière d'imagerie et de radiologie.

C'est de la combinaison de multiples sources de données que dépendent beaucoup de ces innovations : données collectées par des professionnels de santé pendant le parcours de soin, notamment les données génomiques (voir infra 3.3), les données médico-administratives, celles collectées lors d'un accès à internet par smartphone ou ordinateur, *via* des applications qui deviennent des accessoires de suivi de santé et des outils de recherche. Si certains de ces dispositifs participent au parcours de soin et sont encadrés par les professionnels de santé et soumis à une réglementation spécifique française ou européenne, ce n'est pas le cas des dispositifs ou applications contrôlés par des opérateurs commerciaux extraterritoriaux.

Deux exemples illustrent ces innovations :

- L'évolution vers une médecine de précision et une médecine préventive pourrait être une conséquence majeure de l'exploitation des données massives.

  Dans sa pratique actuelle, le médecin est généralement sollicité à son cabinet par une personne qui exprime une demande ou une souffrance cau-

[78] *Chen JH,* et al. *Machine learning and prediction in medicine — Beyond the peak of inflated expectations.* New England Journal of Medicine *2017 ; 376 : 2507-9.*

[79] Rajkomar A, *et al.* Machine learning in medicine. *New England Journal of Medicine* 2019 ; 380 : 1347-58.

[80] Torous J, *et al*. Smartphones, sensors and machines learning to advance real-time prediction and interventions for suicide prevention: a review of current progress and next steps. *Current Psychiatry Reports* 2018 ; 20 : 51 ; De Fauw J, *et al*. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine* 2018 ; 24 : 1342-50.

[81] Les données de vie réelle, un enjeu majeur pour la qualité des soins et la régulation du système de santé. L'exemple du médicament. Bernard Bégaud, D. Polton, F. von Lennep. Rapport réalisé à la demande de la ministre de la santé. Mai 2017.

[82] Rumsfeld JS, *et al.* Big data analytics to improve cardiovascular care: promise and challenges. *Nature Reviews Cardiology* 2016 ; 13 : 350-9.

sée par une maladie. Or, cette maladie est l'aboutissement d'une série de dysfonctionnements de l'organisme qui se sont accumulés et qui auraient pu être détectés plus tôt, si une surveillance de certains paramètres liés à des facteurs de risques personnels avait été instituée[83].

L'exploitation de données massives issues de multiples sources et mesurées en permanence hors de toute prise en charge médicale pourrait aboutir à l'amélioration de la prédiction du risque[84]. Ce passage « *d'une médecine diagnostique symptomatique à une médecine prévisionnelle asymptomatique* » (CCNE avis 77) se fonde non pas sur la prise en charge d'un symptôme déclaré, contrairement à l'exercice actuel de la médecine, mais sur l'anticipation de l'émergence d'un dysfonctionnement. Parce qu'elle intervient trop tardivement par rapport aux manifestations latentes ou insidieuses de la maladie, la médecine clinique classique n'est souvent pas adaptée pour combattre à temps les maladies. Pour améliorer significativement les choses, il est efficace d'instituer une surveillance pratiquement en temps réel de multiples paramètres biologiques, physiques et environnementaux. Cette surveillance établit pour chacun la cartographie individuelle d'un « état de base » dont la déviation, mesurée en continu et en situation, doit alerter et permettre d'adapter une stratégie ou d'instituer des mesures thérapeutiques. Il s'agit donc d'une démarche préventive basée sur la mesure de paramètres et non plus sur l'expression formulée de la demande d'une personne. L'autre innovation est que l'état « pathologique » ne se définit plus comme la déviation par rapport à une zone de normalité établie sur une large population témoin incluant des phénotypes[85] très divers, mais plutôt comme une déviation par rapport à son propre état de base personnel. Chacun devient sa propre référence statistique. Les premières études récemment publiées à partir de l'exploitation de la base de données cliniques, génomiques et d'imagerie cérébrale de la *UK Biobank* (500 000 participants) sont à cet égard convaincantes[86]. À l'extrême, on

[83] Atul J Butte. Big data opens a window onto Wellness. *Nature Biotechnology* 2017; 35: 702. Price ND. *et al.* A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nature Biotechnology* 2017; 35 : 747-56.

[84] Citons le projet Hu-PreciMED (*Human Precision MEDicine*), une initiative d'origine industrielle, a pour objectif de rassembler tous les acteurs publics et privés travaillant dans le domaine de la médecine de précision et de s'articuler *in fine* avec les données de santé des patients du *Health data hub* pour valoriser les données cliniques afin d'améliorer les thérapies et outils de diagnostic disponibles, mais aussi développer de nouvelles approches de médecine prédictive et préventive ; le tout en s'appuyant sur les dernières avancées du Big Data et de l'intelligence artificielle. Quarante-cinq organisations publiques et privées ont rejoint le projet.

[85] Ensemble des caractères observables, apparents, d'un individu. Le phénotype est la résultante de facteurs génétiques (génotype), comportementaux et environnementaux.

[86] Depuis 2006, la UK Biobank au Royaume-Uni a accumulé les données cliniques, génomiques, comportementales de 500 000 personnes. L'ensemble des données génomiques ont été publiées en 2017, et 7 000 chercheurs ont demandé un accès à ces données. Voir : UK Biobank debuts as a powerful resource for genomic research. *Nature Medicine* 2018 ; 24 : 1792-4.

pourrait imaginer construire un « jumeau numérique » de la personne, sur le modèle de ce que l'on conçoit déjà dans le domaine de la robotique[87].

- D'autres innovations sont attendues notamment pour l'interprétation des images radiologiques ou d'imagerie[88], ainsi que pour la gestion de la prise en charge des patients. Sur ce dernier point, on peut citer la détection des priorités dans un service d'urgence[89] et l'amélioration de la réaction à un accident vasculaire ou un choc septique[90], deux situations pour lesquelles la rapidité de mise en œuvre de la thérapeutique a une incidence majeure sur le pronostic. On peut aussi s'attendre à une amélioration de la prédiction des réadmissions hospitalières potentielles [91].

3.1.2 Le respect des principes éthiques lors de l'utilisation des données massives dans la relation de soin aujourd'hui

Schématiquement, dans le cadre du soin, le recueil et l'exploitation des données se font avec un objectif défini, concernant une personne identifiée – le patient - qui en attend un bénéfice, au cours d'un échange avec un professionnel de santé qui assure sa prise en charge et prend des décisions avec lui. Dans le contexte nouveau que nous examinons, si le corps médical s'appuie aujourd'hui sur le traitement de données massives pour améliorer la prise en charge des patients, d'autres acteurs interviennent, souvent sans lien avec les professionnels de santé. Ils traitent la santé non pas dans le cadre du soin et de la déontologie médicale, mais comme un marché. Par le biais des applications ou des objets connectés qu'ils sont les seuls à produire et des réseaux sociaux qu'ils maitrisent, ces opérateurs peuvent concourir au soin et à la recherche de façon crédible. Mais le stockage et leur possibilité d'utilisation de données relatives à la santé posent problème[92]. L'exigence éthique porte sur la clarification des rôles.

---

[87] C'est Michael Grieves qui a introduit cette notion. Le jumeau numérique est « *l'expression numérique des informations d'un système physique. Ces informations numériques sont liées à ce système physique et le seront tout au long de son cycle de vie. L'objectif est de pouvoir accéder à toutes les informations concernant un produit, ses besoins, son comportement, etc., sans le posséder physiquement. Ainsi, le concept du « jumeau numérique » est scindé en deux espaces : un réel, l'autre virtuel. Un flux de données transite de l'espace réel vers l'espace virtuel, et un flux d'informations de l'espace virtuel vers l'espace réel et les sous-espaces virtuels.* ». (Le Monde 26 février 2018 - https://abonnes.lemonde.fr/les-cles-de-demain/article/2018/02/26/le-jumeau-numerique-est-un-interessant-moteur-de-l-innovation_5262662_4758288.html)

[88] Lancement du projet d'un « écosystème français de l'IA dédié à l'imagerie médicale ». Ce système sera indépendant, notamment des GAFA américains et BATX chinois. Hosny A, *et al.* Artificial Intelligence in radiology. *Nature Reviews Cancer* 2018 ; 18 : 500-10.

[89] *Hong WS,* et al. *Predicting hospital admission at emergency department triage using machine learning.* Plos One *2018 ; 13 : e0201016.*

[90] Liu VX, Walkey AJ. Machine Learning and Sepsis: On the Road to Revolution. *Critical Care Medicine* 2017 ; 45 : 1946-7.

[91] Lynch CJ, Liston C. New machine-learning technologies for computer-aided diagnosis. *Nature Medicine* 2018 ; 24 : 1304-5.

[92] Un exemple emblématique est apporté par les manquements à l'exigence éthique lors de la collaboration entre *Deep mind* et le National Health Service (NHS) britannique. En 2016, *Deepmind*, le système d'intelligence artifi-

3.1.2.1 Du point de vue de la personne : l'information et le recueil du consentement

Le consentement est un élément essentiel de la relation entre le médecin et le patient[93]. Le patient doit être informé de la démarche utilisée par le médecin pour établir son diagnostic, déterminer le traitement ou assurer le suivi, et il doit y consentir. L'information sur la démarche suivie implique qu'il soit informé d'un recours au traitement de données massives, soit pour l'interprétation d'examens radiologiques ou d'imagerie, soit pour une aide au diagnostic. Mais il n'est pas tenu de donner un consentement spécifique à l'utilisation des données personnelles collectées dans le cadre de cette prise en charge médicale, le médecin étant soumis au secret médical et donc à la confidentialité des données. En revanche, la question se pose lors de l'utilisation ultérieure de ces données pour la constitution d'une collection ou d'un entrepôt pour une recherche clinique, situation très fréquente à la suite d'une prise en charge hospitalière. Si son consentement n'est pas requis, le patient doit en revanche être informé que ses données pourront être utilisées à des fins de recherches. Il peut s'opposer à cette utilisation et demander communication de ses données. C'est pour celles collectées et conservées hors d'une prise en charge par des professionnels de santé que se posent avec le plus d'acuité les questions relatives à l'information et au consentement, à l'intelligibilité du traitement et à la maîtrise du devenir des données personnelles, sachant que l'utilisation d'objets connectés peut être très profitable dans un parcours de soin. La question de la fiabilité technique et de la protection des données personnelles dans ce contexte sera examinée à la rubrique 3.4.3 qui porte sur l'exploitation des données massives hors relation de soin.

Enfin, on peut s'interroger, dans ce contexte où tout se sait, se prévoit, peut et doit être mesuré et annoncé, sur la place à accorder au droit de ne pas savoir.

3.1.2.2 Du point de vue du professionnel de santé

Le respect du secret médical. Il assure la non-identification de la personne par des tiers. Classiquement, la relation personnelle, directe et exclusive instaurée entre le patient et le corps médical permet à celui-ci d'assurer la protection de ce secret, sous

---

cielle d'Alphabet Inc. (Google) – qui a battu le joueur de Go - annonçait un projet de collaboration avec le *Royal Free London NHS Foundation Trust*, pour concevoir un algorithme et une application mobile de prise en charge des patients développant une insuffisance rénale aiguë. Mais les données de près de 1,5 million de patients – nombre disproportionné par rapport à la finalité déclarée - ont été transférées à *Deepmind* sans information et sans le consentement des patients concernés, témoignant d'une atteinte manifeste aux principes éthiques, révélée par des journalistes. Deux autres projets sont en cours entre *DeepMind* et le NHS.
Powles J, et al. *Google DeepMind and healthcare in an age of algorithms*. Health Technology *2017 ; 7 : 351–67 ; Hodson H. Google's new NHS deal is start of machine learning market place.* New Scientist *6 Jul 2016 ; le partenariat entre Google DeepMind et les hôpitaux londoniens jugé non conforme à la loi.* Le Monde *du 4 juillet 2017 (Pixels).*

[93] « *Toute personne prend, avec le professionnel de santé et compte tenu des informations et des préconisations qu'il lui fournit, les décisions concernant sa santé.* » (Article L.1111-4 du code de la santé publique).

la seule réserve qu'il respecte sa déontologie propre. Mais à l'heure de l'informatisation des cabinets médicaux et d'une nécessaire transmission d'informations médicales, le secret est déjà fragilisé aujourd'hui par la notion « d'informations partagées » entre les personnels soignants (*via* le DMP, dossier médical partagé, ou encore le dossier pharmaceutique[94], ou, bientôt, l'espace numérique de santé du patient[95])[96] . Mais il est encore davantage fragilisé par le recours à des opérateurs ne relevant pas du milieu médical ; il l'est aussi par la divulgation sur internet de données pouvant devenir secondairement des données de santé ; il l'est enfin par la pratique de professionnels de santé utilisant des services numériques grand public.

La responsabilité des décisions qui concernent le patient. Qu'elle porte sur le diagnostic ou le traitement du patient, l'utilisation d'algorithmes doit être considérée comme une aide à la décision humaine, excluant toute automatisation de la décision médicale. Mais ce nouveau type d'intervention pose en premier lieu la question de la vérification de la qualité des prestations des entreprises qui interviennent sur ce marché. Les résultats fournis par leurs logiciels doivent pouvoir être attestés utiles et fiables, ce qui renvoie à la nécessité de contrôle évoquée ci-dessus. C'est ainsi que la FDA (*Food and drug administration*) américaine a récemment validé et autorisé deux logiciels d'intelligence artificielle d'aide au diagnostic pour la rétinopathie diabétique et l'interprétation de mammographies[97]. Mais ces résultats doivent rester une aide confortant le diagnostic élaboré par un professionnel de santé. Son énoncé et les décisions qui en découlent ne peuvent être que de la responsabilité du médecin. (Voir RECOMMANDATION N° 4)

La relation de soin préservée. Une relation personnelle directe doit subsister entre les professionnels de santé et les patients[98]. Elle est indispensable à la confiance qui

---

[94] « *Le dossier pharmaceutique est un dossier informatique, créé et consulté par votre pharmacien, avec votre accord. Il recense les médicaments qui vous ont été délivrés au cours des 4 derniers mois, ainsi que les traitements et prises en cours. Les médicaments figurant sur le dossier peuvent avoir été prescrits par un médecin ou avoir été achetés librement.* » https://www.service-public.fr/particuliers/vosdroits/F16033.

Voir en particulier la délibération de la CNIL (n° 2017-285 du 26 octobre 2017) autorisant la société *OpenHealth Company* à mettre en œuvre un traitement de données à caractère personnel ayant pour finalité la constitution d'un entrepôt de données à caractère personnel à des fins de recherche, d'étude ou d'évaluation dans le domaine de la santé, issues des données des pharmacies d'officines.

[95] Voir note 75 ainsi que le rapport final sur « virage du numérique en santé ». https://solidarites sante.gouv.fr/IMG/pdf/masante2022_rapport_virage_numerique.pdf

[96] Loi de modernisation de notre système de santé, promulguée le 26 janvier 2016, art. L1110-4 et R1110-3.

[97] https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm604357.htm. L'homologation permettra à un médecin généraliste disposant de la caméra appropriée de poser ce diagnostic, facilitant la surveillance des patients diabétiques, et un diagnostic précoce. La FDA a également validé un autre logiciel d'intelligence artificielle pour l'interprétation de mammographies (https://www.fdanews.com/articles/189314-fda-clears-screenpoint-medicals-ai-system-for-reading-mammograms). D'autres devraient suivre (par exemple pour le diagnostic de lésions cutanées).

[98] L'importance de cette relation est rappelée dans le Livre blanc publié en janvier 2018 par le Conseil national de l'ordre des médecins « médecins et patients dans le monde des data, des algorithmes et de l'intelligence artificielle.

est au cœur même de la relation de soin. Parce qu'ils sont en contact direct avec les patients, seuls les personnels médicaux peuvent acquérir une intuition suffisamment fine de la personnalité de chacun d'eux pour comprendre ce que requiert une véritable compréhension des questions intéressant leur santé. Il est intéressant de rappeler combien le monde de la médecine et celui des nouvelles technologies diffèrent par leur vision du corps. C'est traditionnellement par l'interrogatoire, l'observation, l'auscultation et la palpation que le personnel soignant cherche à orienter un diagnostic, et que s'établit la relation. Ce que rend possible le numérique c'est de pouvoir accéder à toutes ces informations hors de la présence réelle du corps. La pérennité du savoir-faire de la pratique médicale classique pourrait être remise en cause si l'aide à la décision fournie par les nouvelles technologies fait apparaître moins nécessaires, voire inutiles, l'enseignement et la pratique de l'observation corporelle. « *La santé du futur, c'est le désenchantement des corps* » exprimait Jean-Michel Besnier[99].
Le risque serait que le traitement des données massives détermine un résultat en puisant dans les données du patient confrontées aux autres données comparables et aux acquis de la science, sans accorder aucune place au patient lui-même. Cela consacrerait également une dépendance face à la machine qui viderait d'une part appréciable de sa substance la garantie humaine objet de la recommandation n° 4 visée ci-dessus. Or, le patient ne saurait être réduit à un ensemble de données à interpréter, rendant inutile son écoute et la prise en compte de son vécu. Aussi utile qu'elle puisse être pour aider au diagnostic et guider le traitement, la donnée ne saurait remplacer le dialogue. Bien au contraire, l'utilisation par le professionnel de santé des technologies récentes doit aussi avoir pour but de libérer du temps pour l'écoute et l'échange en simplifiant le recueil des informations pertinentes. Elle devrait permettre au patient de devenir davantage l'acteur de son parcours de soin en lui permettant une appropriation de ses données, condition d'une attitude pleinement responsable. (Voir **RECOMMANDATION N° 7**).

3.1.2.3 Du point de vue de la puissance publique et de la collectivité (ou du point de vue du système de santé)

Nous avons vu que le développement d'une médecine de précision née de l'exploitation des données pouvait porter atteinte à la mutualisation des risques. Une telle évolution vers la médecine de précision n'est encore qu'à l'état hypothétique, car elle requiert une organisation et des infrastructures qui ne sont pas opérationnelles. Si elle devait advenir, elle poserait des questions éthiques redoutables.

Citons la recommandation N°10 : le CNOM recommande « *[…] que le développement des dispositifs techniques ayant recours à l'intelligence artificielle soit incité à aller dans le sens d'un marché industriel d'aide à la décision médicale et non pas vers celui qui dicterait au médecin comme au patient une décision rendue par l'algorithme qui s'imposerait à eux sans être susceptible de critique ou de transgression.* »

[99] Interview de J.M. Besnier par Hugo Jalinière. *Science et Avenir*, 6 septembre 2015.

La principale, à l'image de ce que nous venons de voir, serait de réduire le citoyen à un ensemble de données dont l'accumulation aurait la prétention de l'appréhender dans sa globalité, comme l'a déjà souligné le CCNE dans son Avis 98 sur la biométrie[100].

Mais cette évolution accroîtrait en outre fortement la tension qui est toujours sous-jacente entre les attentes individuelles et celles de la société, guidée par des préoccupations économiques. Si l'analyse et la surveillance systématique des données de santé, environnementales et comportementales de tous, malades ou bien portants, devenaient la règle, il serait porté atteinte à deux principes éthiques majeurs : d'une part, celui de la liberté des choix de vie des individus, et leur droit de savoir ou de ne pas savoir, notamment lorsqu'il n'existe pas de stratégie curative ou préventive efficace pour lutter contre une vulnérabilité détectée ; d'autre part, le risque d'une perte de mutualisation et de solidarité dans la prise en charge de la maladie pour les personnes qui ne se conformeraient pas aux mesures qui leur seraient préconisées. La prédiction du risque médical en viendrait à déterminer les politiques de santé et à commander leur application.

Un autre risque, commun à tous les contextes d'utilisation des données massives dans le champ de la santé, est celui d'une perte d'autonomie et de souveraineté de notre pays (voir aussi 2.2.2 *in fine*, et 3.4.2). Cette perte est due à un retard technologique dans les domaines de l'hébergement et du traitement de données dont le volume augmente de manière particulièrement importante avec les nouvelles techniques d'analyse génomique et avec la politique de numérisation systématique des examens radiologiques ou d'imagerie pratiquée par les établissements[101]. Les lieux actuels de stockage des données de santé sont encadrés par la loi 2016-41 du 26 janvier 2016 de modernisation de notre système de santé. Celle-ci a institué une procédure de certification encadrée par l'État et confiée à des organismes publics ou privés. L'évolution se fait donc vers la mutualisation et la migration des données vers des plateformes gérées par des prestataires privés ou publics. Les géants américains (Amazon, Microsoft) s'installent en France et Microsoft a obtenu en novembre 2018 la certification d'hébergeur de santé pour ses quatre centres situés en France.

Face au défi technologique que posent, pour la souveraineté nationale et européenne, le stockage, le partage et le traitement des données massives dans le domaine de la

[100] « *Leur conjugaison [des données] confère à l'ensemble un caractère de quasi-infaillibilité et enferme chacun d'entre nous dans un cadre bien défini, la société tendant à s'accommoder de cet enfermement de la personne en une série de données ainsi rassemblées. [..] Ne réduisent-elles pas l'homme à une accumulation de données et de critères cartographiques, ceci paradoxalement à l'heure où la biologie, délaissant quelque peu une approche analytique et réductionniste, s'attache à appréhender un système dans sa globalité, en cherchant à intégrer l'ensemble des propriétés d'un organisme ou d'un être vivant (biologie intégrative)*. » CCNE Avis n° 98 : Biométrie, données identifiantes et droits de l'homme.

[101] *Par exemple, le volume d'imagerie médicale augmente de 20 % à 40 % par an.*

santé, le CCNE préconise le développement de plateformes nationales mutualisées et interconnectées. Ouvertes selon des modalités qu'il faudra définir aux acteurs publiques et privés, elles doivent permettre à notre pays et à l'Europe de préserver leur autonomie stratégique et de ne pas perdre la maîtrise de la richesse que constituent les données, tout en privilégiant un partage contrôlé qui est indispensable à l'efficacité du soin et de la recherche médicale. **(RECOMMANDATION N° 10).**

C'est le sens de la plateforme des données de santé qui devrait être créée dans le cadre du projet de loi relatif à l'organisation et à la transformation du système de santé[102], suite aux préconisations du rapport Villani[103].

3.2 Exploitation des données de santé dans le cadre des protocoles de recherche

L'accumulation de données recueillies dans des contextes très variés, parfois en temps réel, et qu'il est possible de réutiliser, rend plus ténue la frontière entre soin et recherche.

Celle-ci n'en conserve pas moins une spécificité qui tient à son objectif propre : elle vise un progrès des connaissances en matière médicale, dans l'intérêt de tous, et non les soins préventifs ou curatifs que nécessite l'état d'un patient déterminé.

L'exploitation des données massives (regroupées dans d'immenses bases locales, nationales ou internationales) pour la recherche sert à identifier des groupes de personnes qui partagent les mêmes caractéristiques, à partir du traitement croisé de données génomiques, cliniques, d'imagerie, etc. De ces corrélations – obtenues dans certains cas sans hypothèse préalable – on tire plusieurs éléments : d'une part de nouvelles pistes de recherche sur les mécanismes d'une maladie (projet de recherche de causalité), d'autre part la définition de marqueurs diagnostics précoces, d'indicateurs pronostiques ou thérapeutiques.

Dans ces bases de données, le plus souvent, l'individu n'est pas nominalement identifié*, il est en quelque sorte « effacé* ». S'il y a toujours une personne qui fournit les données primaires, elle n'a aucun lien avec ceux qui vont les exploiter et elle reste généralement anonyme. À la différence du clinicien qui est face à son patient, le chercheur utilisant une grande base de données, ou le curateur ou le scientifique qui la gèrent,

[102] Voir le titre III du projet de loi « Développer l'ambition numérique en santé » du projet de loi déposé en février 2019. https://solidarites-sante.gouv.fr/actualites/actualites-du-ministere/article/presentation-du-projet-de-loi-relatif-a-l-organisation-et-a-la-transformation. Voir note 73.

[103] Voir Rapport *Health Data Hub,* mission de préfiguration. 12 octobre 2018. https://solidarites-sante.gouv.fr/ministere/documentation-et-publications-officielles/rapports/sante/article/rapport-health-data-hub-mission-de-prefiguration

ignorent tout du titulaire des données. Il en résulte deux risques : ne pas voir dans le titulaire des données une personne humaine et ne pas assurer une réelle anonymisation (ou pseudonymisation) des données.

L'objectif recherché est l'intérêt général. Il s'agit de faire progresser la connaissance ou l'établissement de mesures de santé publique, ce dont l'individu pourra bénéficier, souvent de manière indirecte et incertaine. Comment dès lors se déclinent les principes éthiques énoncés au chapitre 2, selon qu'on se place du côté de la personne qui est à la source des données, du spécialiste des données (*data scientist*), du chercheur, et de la collectivité qui en tirera des décisions ? C'est dans ce contexte que s'opposent plusieurs courants de pensée et que se discutent les nouvelles formes de consentement et d'accès aux données.

3.2.1 L'enjeu éthique du partage des données

Le partage des données (ou l'accès aux données) est au fondement même de l'exploitation des données massives. Celles-ci n'ont d'intérêt que si elles sont accessibles au plus grand nombre de chercheurs ou de cliniciens et si elles peuvent être croisées avec des données cliniques, environnementales, souvent stockées dans le même « nuage (*cloud)* » ou le même « gisement » ou « entrepôt ».

Ce partage suscite une tension inévitable entre le risque d'une sous-exploitation des données pouvant mettre en péril des recherches menées dans l'intérêt général et celui d'un partage trop large pour que les droits fondamentaux de la personne soient respectés. De cette difficulté d'assurer à la fois un accès large aux données et une protection efficace des droits individuels[104] sont nées les réflexions sur un *« droit collectif »* sur les données de santé [105] ou encore sur un « *droit à la science*[106] » que nous évoquerons au § 3.4. Cette difficulté est accrue par l'absence de standards internationaux de bonnes pratiques et de gouvernance[107], même si le règlement européen apporte un progrès important[108].

Il faut considérer les principes éthiques tels qu'ils sont déclinés par les trois entités qui interviennent dans un processus de recherche : la personne titulaire des données et concernée par le traitement de ses données (la question du consentement et de la sécurité des données), la gouvernance des bases de données (qui en contrôle la sécu-

[104] Joly Y, *et al*. Are data sharing and privacy protection mutually exclusive? *Cell* 2016 ; 167 : 1150.
[105] Bourcier D, Filippi P. Vers un droit collectif sur les données de santé. *Revue de droit sanitaire et social* (Dalloz revues) 2018 : pp.444-50.
[106] Knoppers BM, Thorogood AM. Ethics and Big Data in health. *Current Opinion in Systems Biology* 2017, 4 : 53-7.
[107] Il est à cet égard intéressant de relever la récente condamnation en Chine d'entreprises ayant enfreint la législation sur le partage des données (*Nature* 15 novembre 2018, page 301)
[108] Stein L, *et al.* Data analysis : Create a cloud commons. *Nature* 2015 ; 523 : 149.

rité et l'accès ?) et le chercheur, qui doit faire preuve de vigilance dans l'utilisation des données pour sa recherche.

3.2.1.1 Le point de vue de la personne : quelle place pour le consentement /nouvelles formes de consentement ?

Les caractéristiques mêmes des données décrites au paragraphe 1.3 créent deux difficultés dans le contexte de la recherche ; l'une vient de l'absence de définition claire de ce qu'est une activité de « recherche », terme qui n'est pas défini par le RGPD alors que l'objectif de recherche scientifique peut être un critère de licéité du traitement des données de santé ou de la réutilisation des données, indépendamment du consentement. La seconde vient de l'imbrication entre ce qui relève du soin ou de la recherche. Une situation fréquente (que reconnaît le RGPD) est celle de l'imprécision de la finalité du traitement des données, lorsqu'elle sera énoncée à la personne lors de son information et – s'il est requis – de son consentement initial au recueil et au traitement des données[109]. C'est dans ce contexte d'imprécision de la spécificité du consentement que prennent toute leur importance la qualité de l'information donnée, et donc la relation de confiance avec l'interlocuteur et que doivent être garantis la transparence, le respect des exigences éthiques applicables à la recherche scientifique en particulier médicale, et un engagement de communication régulière sur l'utilisation des données.

Il est difficile de transposer le modèle traditionnel de consentement, qui a été conçu dans le cadre d'une relation entre un individu (parfois un patient) et un acteur (chercheur/médecin), œuvrant pour un projet précis (finalité), limité dans le temps, collectant des données spécifiques en adéquation avec le projet, régi par une réglementation stable. Ce modèle n'est plus approprié pour un tel flux de données ni pour un partage généralisé, alors que les possibilités d'exploitation sont multiples et qu'aucune information précise sur le devenir de ces données n'est disponible au moment de leur collecte. Il n'est d'autre part pas toujours réaliste d'envisager de recontacter les personnes qui ont initialement fourni des données pour les faire consentir à une utilisation pour un autre projet. Le RGPD dispense de cette obligation si la finalité n'est pas incompatible avec celle du projet initial (art. 6.4 du RGPD).

Dès lors, d'autres modes de consentement sont discutés dans le contexte de la recherche :

- une possibilité est celle du *consentement large (ou broad consent),* qui ne précise pas une finalité précise, mais seulement un champ d'applications, qui peuvent être

[109] Groupe de travail « article 29 » : lignes directrices sur le consentement au sens du règlement 2016-679, version finale du 10 avril 2018. ttps://www.cnil.fr/sites/default/files/atoms/files/ldconsentement_wp259_rev_0.1_fr.pdf.

variées. Les participants peuvent consentir à l'utilisation de leurs données par une biobanque, interlocuteur qui, en retour, garantit une sécurité et un contrôle des accès *via* un comité de gouvernance (voir à cet égard, l'organisation du type de consentement de la *UK Biobank*[110]) ;
- un second mode est celui du *consentement à options*, la personne pouvant sélectionner les domaines de recherche pour lesquels elle autorise l'utilisation de ses données (par exemple : recherche cardio-vasculaire, mais pas recherche sur le cancer, ou consentement pour l'utilisation des données génétiques distinct de celui accepté pour les données biologiques classiques) ;
- le *consentement dynamique* (souvent lié au consentement large) tire parti des technologies numériques ; les titulaires des données sont considérés comme des participants à la recherche. Ils modifient ou actualisent leur consentement en fonction des nouvelles finalités dont ils sont informés *via* un site dédié assurant ces échanges d'informations ;
- 'une autre modalité est le *consentement présumé* (*opt-out*) dans lequel les données peuvent être utilisées, sauf en cas de refus, et c'est le refus (et non l'approbation) qui est signifié par la personne.

Si les données sont complètement anonymisées, elles ne peuvent être qualifiées de données personnelles et ne relèvent pas de la protection du RGPD ni de celle de la version actuelle de la loi informatique et liberté. L'accès à ces données cliniques agrégées peut ainsi être complètement libre. Mais se pose alors la question du caractère irréversible de l'anonymisation des données ; non seulement une réidentification ultérieure ne peut plus, aujourd'hui, être exclue, mais l'anonymisation ampute les données d'une grande partie de leur utilité, car elle oblige à effacer ou brouiller une partie de l'information utile[111]. Peut-on dès lors éthiquement se dispenser d'un consentement ?

Lorsque le consentement est exigé, il ne peut être unique et homogène, quel que soit le contexte de recherche dans lequel il est requis. L'impératif éthique est que ce consentement soit adapté à chaque situation particulière, que l'information donnée en retour sur les recherches soit disponible, évolutive, claire et loyale, de manière à établir et justifier une relation de confiance entre titulaires des données et ceux (curateurs et chercheurs) qui y ont accès et qui les traitent. Cette information devrait inclure à la fois des publications scientifiques en accès libre et des lettres d'information rédigées pour le public non scientifique.

---

[110] https://www.ukbiobank.ac.uk/the-ethics-and-governance-council/. Voir aussi le rapport annuel du comité de gouvernance. https://egcukbiobank.org.uk/sites/default/files/UKBEGC_Review2016_2017.pdf

[111] « *Dé-identification et pseudonymisation, ces expressions un peu lourdes sont des quasi-synonymes en ce sens que dans l'un et l'autre cas, la vraie identité de la personne (nom-prénoms, NIR…) est absente ou masquée. L'emploi d'un pseudonyme signifie en outre qu'on a remplacé la vraie identité par un identifiant conventionnel (souvent un « numéro d'anonymat ») qui dans un contexte donné désigne toujours la même personne afin de permettre un suivi longitudinal (suivi du parcours). L'attribution d'un pseudonyme par un procédé qui interdit au gestionnaire des données de remonter lui-même au nom de la personne concernée (un chiffrement irréversible par exemple) était souvent appelée anonymisation et l'est parfois encore mais on sait mieux aujourd'hui que les jeux de données ainsi modifiés ne sont pas nécessairement anonymes ; c'est pourquoi il vaut mieux parler en ce cas de pseudonymisation* » (André Loth, DRESS, Solidarité Santé, Juillet 2015).

Rappelons que le consentement n'est pas l'unique fondement de la licéité du traitement des données personnelles relatives à la santé, et que, dans les situations nombreuses où le consentement n'est pas exigé, une information « *concise, transparente, compréhensible et aisément accessible* » est en revanche requise. On peut dès lors s'interroger sur la gestion de cette information complexe par le responsable du traitement : priorisation des informations, canaux qui seront utilisés pour la délivrer – interactions humaines ou outils numériques –, possibilité pour la personne concernée de pouvoir vraiment déterminer à l'avance la portée et les conséquences du traitement de ses données, ce d'autant qu'elle peut être en situation de vulnérabilité. Si les caractéristiques de cette information ont été détaillées par le G29[112], dans quelle mesure seront-elles réellement suivies et comment s'assurer qu'elles ont été délivrées et qu'elles ont été comprises par la personne ?

Une autre difficulté, déjà évoquée (chapitre 2.1.3), est celle du retrait par la personne de ses données, qui ne serait pas justifié par un motif légitime. Outre qu'un retrait discrétionnaire pourrait n'être pas techniquement possible, il pourrait aussi être contraire à l'éthique. La qualité de la recherche serait en effet affectée par un risque de biais et de conclusions faussés si certaines catégories de données étaient soustraites du projet de recherche.

Compte tenu de la diversité des usages et des contextes utilisant les données massives, notamment dans le domaine de la recherche, il est nécessaire de mener une réflexion sur la notion de consentement au traitement de données massives. Cette réflexion devrait porter sur l'objet du consentement, ainsi que sur les modalités de son recueil, afin d'assurer durablement l'équilibre entre le respect des droits des personnes et la dynamique des usages.

Cette réflexion devra nourrir le débat public sur les recommandations éthiques et permettra l'actualisation périodique de la loi.
(RECOMMANDATION N° 3)

Les modalités recherchées supposent une relation de confiance entre titulaires des données et ceux, techniciens, ingénieurs et chercheurs qui ont accès à ces données et qui les traitent.
Il est essentiel que le titulaire des données soit informé des modalités par lesquelles l'autorité de contrôle assure sa fonction de tiers de confiance. Doit ainsi être assurée une triple exigence éthique :

[112] Groupe de travail « Article 29 ». Lignes directrices sur la transparence au sens du règlement (UE) 2016/679 – version 11 avril 2018. https://www.cnil.fr/sites/default/files/atoms/files/wp260_guidelines-transparence-fr.pdf

- une évaluation rigoureuse et transparente de l'intérêt des recherches, qui doivent contribuer au bénéfice de tous à un enrichissement des connaissances dans le domaine de la santé (notion de bien commun) ;
- un partage des informations sur la progression des recherches avec les participants, selon des modalités diverses ;
- l'assurance d'une sécurité des données, de leur traçabilité et de l'absence d'usage malveillant.

(**RECOMMANDATION N° 11**)

3.2.1.2 La gouvernance : l'importance d'une garantie de l'institution pour l'accès aux données

La confiance accordée par le titulaire des données repose aussi sur la manière dont est sauvegardé le contrôle de l'accès aux bases de données. Il est assuré par un comité tiers, souvent un comité de gouvernance, qui s'assure d'une mise en conformité avec les principes éthiques et la réglementation de l'institution, publique ou privée, dépositaire des données, ce qui limite les risques de mésusage. L'accès aux données, pour des projets de recherche ou de santé publique, est plus souvent dit *contrôlé ou restreint* ; un comité évalue la pertinence du projet de recherche, son adaptation au consentement des participants, la conformité aux règles éthiques et de sécurité, ainsi que l'engagement de diffuser les résultats à la communauté ; une procédure intermédiaire est celle d'un *accès déclaré*, où le chercheur ou le clinicien déclinent seulement leur identité et acceptent de se conformer aux termes d'utilisation des données (par exemple conformité aux méthodologies de référence publiées par la CNIL pour l'accès à l'INDS [Institut national des données de santé] en France[113]). Ces différents modes de contrôle sont plus ou moins adaptés selon l'usage de la base de données (recherche fondamentale ou clinique), le type de données, et le statut du demandeur. Aujourd'hui, le modèle de l'accès contrôlé reste le plus sûr. Le respect des principes éthiques exige que la gouvernance des données sensibles soit exempte d'ambiguïté et rigoureuse dans la mesure où certaines données des bases internationales, notamment les bases génomiques, peuvent parfois être hébergées et donc accessibles *via* un Cloud géré par un partenaire tiers (par exemple Amazon). Le développement de techniques dites *privacy enhancing technics* pourrait aussi apporter une réponse[114].

[113] La CNIL a publié, en date du 13 juillet 2018, 3 nouvelles méthodologies de référence destinées à encadrer les traitements de données à caractère personnel à des fins d'études, évaluations ou recherches n'impliquant pas la personne humaine. Il s'agit de procédures d'accès simplifié aux données à caractère personnel de santé permettant d'éviter une demande d'autorisation (INDS-CEREES-CNIL) sans pour autant contourner la nécessité pour l'étude, l'évaluation ou la recherche de présenter un intérêt public. En contrepartie de cet allègement des formalités, le responsable de traitement s'engage à respecter plusieurs obligations.

[114] Erlich Y, Narayanan A. Routes for breaching and protecting genetic privacy. *Nature Reviews Genetics* 2014 ; 15 : 409-21.

En France, des exemples d'accès contrôlé sont ceux de l'accès à l'INDS, aux entrepôts hospitaliers ou aux bases de données génomiques (voir projet France-médecine génomique). L'exemple de la UK Biobank cité plus haut, en est un autre. Mentionnons la création en France, inscrite dans le projet de loi de transformation du système de santé déposé en février 2019, d'une plateforme des données de santé (*Health data hub)*, projet engagé par la mission Villani, qui associera les grands organismes de recherche français, l'Inserm, le CNRS et l'Inria, ainsi que de nombreux partenaires publics et privés et dont l'objectif est de permettre, dans un cadre fortement sécurisé, d'avoir, entre autres, un espace de travail pour l'apprentissage algorithmique. Le projet propose une gouvernance de la plateforme, des principes d'intervention et des modalités juridiques et opérationnelles pour régir le partage des données.

3.2.1.3 La chaîne transdisciplinaire des intervenants-chercheurs : réflexion éthique

Le devenir des données échappe à la maîtrise de leur titulaire initial en raison de l'intervention de multiples processus et acteurs et d'une technologie qu'il ne peut le plus souvent comprendre. C'est donc sur une relation de confiance et de réciprocité que reposera son consentement, et non pas sur une définition restrictive des finalités, établie *a priori*. Les chercheurs doivent mériter cette confiance et il ne faut pas sous-estimer la résistance de la société à certains aspects du partage des données en matière de recherche[115].

- *Une personne est à l'origine des données.* À la différence du médecin qui est face à son patient, le chercheur utilisant des données massives n'a aucun contact avec leurs titulaires et il ne sait rien d'eux. Le danger est donc qu'il considère les données comme un simple instrument de sa recherche, en oubliant qu'elles sont liées à des personnes qui sont exposées à un risque de nuisance. Cela rend d'autant plus importante sa connaissance de la qualité de l'information qu'ont reçue les titulaires des données et les termes du consentement qu'elles ont donné. L'attention apportée au risque d'atteinte au respect de la vie privée, à la sécurité des données et aux failles permettant une réidentification doit être d'autant plus grande que le consentement est large. Il est essentiel aussi que les chercheurs s'assurent que les données qu'ils exploitent n'ont pas été obtenues dans des conditions non éthiques. On peut rappeler à cet égard l'exemple de la société *23&me*, vendant les données génétiques de ses cinq millions de clients au géant britannique de l'industrie pharmaceutique GlaxoSmithKline. La société a fait valoir que 80 % de ses clients avaient donné leur accord pour que leurs données soient utilisées à des fins de recherche médicale[116].

[115] Majumder MA, *et al.* Beyond our borders? Public resistance to global genomic data sharing. *PLoS Biology* 2016 ; 14 : e2000206.

[116] *23andMe's Pharma deals have been the plan all along.* Wired *8 march 2018. (https://www.wired.com/story/23andme-glaxosmithkline-pharma-deal/)*

- Comme « *producteur* » *de nouvelles données*, qui pourront contribuer à l'élaboration d'une décision dans le domaine de la santé, le chercheur est responsable de la qualité de l'enchaînement des étapes nécessaires à l'obtention des données qu'il utilise, de l'absence de biais pouvant affecter leur collecte, des questions sur lesquelles repose le traitement, de l'évaluation de l'algorithme. Les données massives apportent une *transdisciplinarité nouvelle* : c'est ainsi que le chercheur en biologie voulant utiliser les données massives doit se former aux analyses algorithmiques ou collaborer avec des *data scientists*, dans le cadre établi des règles qui gouvernent ces bases de données.

- Les *mathématiciens/informaticiens* constituent un tiers entre la personne qui fournit les données et le chercheur ou le médecin qui les utilise. Au même titre que le chercheur, ces intervenants doivent connaître les enjeux éthiques liés à l'exploitation des données massives et être formés à la protection de la vie privée des personnes. Ils doivent être informés des conséquences d'erreurs et d'imprécisions dans la collecte des données, dans la prise en considération de leur nombre, de leur hétérogénéité, de la diversité des contextes de recueil (patient, chercheur, etc.). Celles-ci ne sont en effet pas négligeables, et peuvent avoir des effets néfastes, voire délétères, sur les systèmes d'apprentissage machine. (Voir **RECOMMANDATION n° 5**)

- *Organisation et gouvernance de données, responsabilité des acteurs.* La transdisciplinarité accroît l'importance de l'organisation et de la gouvernance des données (recueil, annotation, hébergement) pour qu'elles favorisent les découvertes médicales et qu'elles puissent être utilement réutilisées. Cette gouvernance doit être assurée sous la responsabilité (*accountability*) des acteurs, qui doivent « *mettre en œuvre les mécanismes et les procédures internes permettant de démontrer le respect des règles relatives à la protection des données* ». C'est en quelque sorte un contrat de confiance entre les personnes qui acceptent de confier leurs données et l'organisation qui en assure le devenir et l'accès. Une des réponses possibles envisagées est le concept de *privacy by design*, ou « *protection de la vie privée dès la conception du projet* », concept créé aux États-Unis dans les années 1990 et introduit dans le RGPD[117]. Chaque programme traitant des données personnelles doit garantir dès sa conception et lors de chaque utilisation, même si elle n'a pas été prévue à l'origine, le plus haut niveau possible de protection des données.

3.2.2 Le partage des données dans le domaine de la recherche avec les entreprises privées du médicament et les opérateurs d'internet

- *L'industrie pharmaceutique* considère que l'exploitation des données massives peut aider substantiellement la recherche et le développement et soutenir un modèle que mettent en difficulté la hausse des coûts et l'absence de molé-

[117] Article 25 - Protection des données dès la conception et protection des données par défaut.

cules phares[118]. Cette aide peut porter sur des modèles prédictifs de l'action de nouveaux médicaments, sur l'exploitation des données des essais cliniques et sur une sélection plus adaptée et plus rapide des patients participant à ces essais, ainsi que sur une détection plus rapide et précise des effets secondaires des nouvelles molécules, détectés en vie réelle[119]. Ce dernier point est très important puisque les essais cliniques sont habituellement réalisés sur une population restreinte et très ciblée, ce qui ne permet pas d'anticiper des effets néfastes inattendus qui peuvent résulter de la diffusion plus large du produit. Lorsque des opérateurs privés sont associés à la recherche, ils doivent, pour accéder aux données de santé et aux données personnelles collectées et conservées par l'autorité publique, prendre des engagements de non-divulgation et de non-utilisation à d'autres fins. Un contrôle doit être effectué pour vérifier que le projet commercial ou le projet d'entreprises privées (*big pharma*) est conforme à ces exigences. Cette base contractuelle est d'autant plus importante que la recherche est aujourd'hui mondialisée et qu'elle associe donc des acteurs soumis à des législations inégalement contraignantes.

- *Les données issues des plateformes internet dans le domaine de la recherche :* Les réseaux sociaux (*Facebook*, *Twitter*) et les plateformes internet de partage d'informations de santé destinées aux patients (par ex. *Patientslikeme* créé en 2004 aux États-Unis, ou en France Carenity[120], construit sur le même modèle) sont devenus une source très importante d'informations concernant la santé : pour le soin et les alertes sur les effets indésirables des médicaments, mais aussi pour la recherche clinique (ciblage à des fins de recrutement de patients), ou des enquêtes dans le cadre des politiques de prévention ou de veille sanitaire. L'intérêt de ces données vient en partie de ce qu'elles se rapportent à « la vie réelle » - parce qu'elles sont collectées hors des modalités usuelles de prise en charge. L'importance de ces données en vie réelle est aujourd'hui reconnue[121]. De plus en plus de programmes de recherche sollicitent directement les titulaires des comptes de réseaux sociaux pour leur participation – avec leur consentement - à des études de recherche clinique. Citons le *World Diabetes Distress Study*[122], la reconnaissance d'épisodes dé-

[118] AI-powered drug discovery pharma invest. *Nature Biotechnology* 2017 ; 35 : 605.

[119] Les données de vie réelle, un enjeu majeur pour la qualité des soins et la régulation du système de santé. L'exemple du médicament. Bernard Bégaud, Dominique Polton, Franck von Lennep. Rapport réalisé à la demande de la ministre de la Santé. Mai 2017.

[120] Carenity (créé en 2011, plus de 15 000 utilisateurs) « facilite la mise en relation d'hommes et de femmes concernés par les mêmes maladies en mettant gratuitement à disposition un réseau social : groupes d'amis, fil d'actualité, forums de discussion, messagerie privée, etc. Les données peuvent être utilisées dans le cadre de programmes de recherche publics ou privés. Toutes les données sont agrégées ». https://www.carenity.com/

[121] Bégaud B, Polton D, von Lennep F. Les données de vie réelle, un enjeu majeur pour la qualité des soins et la régulation du système de santé. Rapport réalisé à la demande de Madame la ministre de la Santé, mai 2017. https://solidarites-sante.gouv.fr/IMG/ pdf/rapport_donnees_de_vie_reelle_medicaments_mai_2017vf.pdf

[122] Fagherazzi G, *et al.* Étude mondiale de la détresse liée au diabète : le potentiel du réseau social Twitter pour la recherche médicale - *Revue d'épidémiologie et de santé publique*. Doi : 10.1016/j.respe.2018.04.002. *World Diabetes Distress Study* (WDDS) est un projet de recherche international (auquel participe l'Inserm) sur le diabète (de type 1 et de type 2) qui a pour but d'identifier des marqueurs de la détresse liée au diabète, d'une mauvaise qualité de vie et du risque de complications. La détresse liée au diabète est définie par le fardeau que représente

pressifs *via* les échanges sur *Facebook*[123], le suivi de maladies mentales sur *Twitter*[124], le suivi thérapeutique de maladies de Parkinson[125], ou la prévention du suicide[126].

Plusieurs études plaident en faveur d'une ouverture plus large de ces données. Le rapport récent de la CERNA sur la souveraineté numérique préconise que « *les plateformes collectant massivement des données (par exemple les GAFAMI et BATX) soient tenues d'ouvrir ces données à des fins de science ouverte dans des conditions strictes d'éthique, d'intégrité et de déontologie scientifique* ». C'est aussi la conclusion d'un livre blanc du *Healthcare Data Institute :* « *Permettre un accès simplifié et gratuit aux bases de données comprenant les données rendues publiques sur les réseaux sociaux par leurs utilisateurs, pour les acteurs de la recherche publique* »[127]. Ces réseaux induisent un dialogue entre chercheurs et patients, permettant à ces derniers de susciter des thèmes de recherche en santé.

Toutefois, la diffusion et l'utilisation des données hors d'un encadrement sécurisé institutionnel posent des questions éthiques importantes pour la protection des patients, notamment le respect des limites de leur consentement à la diffusion, à l'hébergement et à la réutilisation de ces données. On peut aussi s'interroger sur une possible remise en cause de leur libre arbitre pour prendre les décisions médicales qui les concernent. Au-delà, il est possible, en analysant un comportement sur les réseaux sociaux, de prédire certaines particularités sociales[128], ce qui pourrait aboutir à d'inacceptables stigmatisations.

Le CCNE considère qu'il est nécessaire de faciliter le partage des données de santé pour les besoins de la recherche. Il est notamment d'avis de permettre, pour des protocoles de recherche aux finalités strictement définies et dans le respect des droits des personnes dont les données ont été mises à disposition et avec leur consentement, l'accès des chercheurs à des données collectées sur internet ou les réseaux sociaux par des plateformes dont la gouvernance est contrôlée. (**RECOMMANDATION N° 12**).

---

le stress, les craintes ou encore les émotions liées à la gestion du diabète au quotidien. Elle est considérée comme le facteur de santé psychosocial le plus important dans la gestion d'un diabète.

[123] Eichstaedt JC, *et al*. Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences of the USA* 2018 : 115 : 11203-8.

[124] Reece AG, *et al.* Forecasting the onset and course of mental illness with Twitter data. *Sci Rep* 2017 ; 7 : 13006 ; Reece AG, Danforth CM. Instagram photos reveal predictive markers of depression. *EPJ Data Science* 2017 ; 6 : 15.

[125] *Gravitz L. Technology : Monitoring gets personal.* Nature *2016 ; 538 : S8–S10.*

[126] Rous J, *et al.* Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: a review of current progress and next steps. *Current Psychiatry Reports* 2018 ; 20 : 51.

[127] Livre blanc : les réseaux sociaux et la santé : un enjeu pour le suivi des patients et la recherche scientifique. https://healthcaredatainstitute.com/wp-content/uploads/2015/02/livre-blanc-hdi-2018-print_bd-bd.pdf

[128] *Kosinski M,* et al. *Private traits and attributes are predictable from digital records of human behavior.* Proceedings of the National Academy of Sciences of the USA *2013 ; 110 : 5802-5.*

3.3 Un exemple emblématique à la frontière du soin et de la recherche : les données génomiques

Parmi les données de santé, les « *données à caractère personnel relatives aux caractéristiques génétiques héréditaires ou acquises* », selon la définition du RGPD[129], sont symptomatiques à plusieurs titres : elles posent toutes les questions évoquées ci-dessus : volume et stockage des données, risque de diffusion incontrôlée, possibilité d'identification, de perte de confidentialité, et donc de sécurité. Elles illustrent aussi les questions soulevées par un partage des données au bénéfice de la santé et leur croisement, souvent transnational, qui seul permet d'acquérir du sens, mais qui soulève des questions légales et éthiques aussi bien que techniques très difficiles. Les données génomiques mettent par ailleurs en évidence la tendance, qui se développe surtout aux États-Unis et en Asie, qui veut que la génétique ne soit plus aujourd'hui l'exclusivité du monde médical[130] et des cercles académiques. Les données, surtout en Amérique du Nord, acquièrent une valeur marchande et intéressent le secteur économique : leur exploitation est essentiellement le fait des *biotechs*, avec des objectifs de généalogie familiale, mais aussi de santé et de recherche.

D'un point de vue juridique, plusieurs textes de lois encadrent les données génétiques : le code civil, les lois de bioéthique (pour ce qui concerne spécifiquement les conditions de la réalisation des tests génétiques), la loi informatique et liberté, notamment dans sa version issue de la loi du 20 juin 2018[131]. Cette loi identifie les données génétiques comme des données sensibles (ce qui est le cas général des données concernant la santé), ce qui entraîne une interdiction de les traiter (article 8 de la LIL actualisée). Cette interdiction est toutefois assortie d'exceptions[132], notamment en

[129] « *les données à caractère personnel relatives aux caractéristiques génétiques héréditaires ou acquises d'une personne physique qui donnent des informations uniques sur la physiologie ou l'état de santé de cette personne physique et qui résultent, notamment, d'une analyse d'un échantillon biologique de la personne physique en question » (RGPD, article 4) notamment une analyse des chromosomes, de l'acide désoxyribonucléique (ADN) ou de l'acide ribonucléique (ARN), ou de l'analyse d'un autre élément permettant d'obtenir des informations équivalentes* » (RGPD, considérant 34).

[130] Sharon T. The googlization of health research : from disruptive innovation to disruptive ethics. *Personalized Medicine* 2016 ; 13 : 563-74.

[131] Voir notamment l'article 75 de l'ordonnance n° 2018-1125 du 12 décembre 2018 prise en application de l'article 32 de la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles.

[132] Extrait de l'article 8 de la loi informatique et liberté (LIL) modifiée pour être en conformité avec le RGPD (juin 2018).

I-Il est interdit de traiter des données traiter des données à caractère personnel qui révèlent la prétendue origine raciale ou l'origine ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale d'une personne physique ou de traiter 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.

II. - Dans la mesure où la finalité du traitement l'exige pour certaines catégories de données, ne sont pas soumis à l'interdiction [] prévue au I :

« 4° Les traitements portant sur des données à caractère personnel rendues publiques par la personne concernée ;

ce qui concerne la recherche et la santé, si la personne a donné un consentement libre et éclairé.

3.3.1 Du génotype au phénotype[133] : un changement de logique dû à l'avancée des technologies d'analyse du génome

Le séquençage du génome humain en 2001 a joué un rôle majeur dans la compréhension des déterminants génétiques des maladies. Il est à l'origine d'une stratégie qui consiste à corréler variants génétiques et vulnérabilité à certaines maladies, stratégie engagée par l'analyse de quelques familles et poursuivie par des études d'association génétique à grande échelle sur des milliers de participants (patients, apparentés et témoins). Ces analyses dites GWAS (pour étude d'association globale du génome)[134] ont rendu nécessaire la mise en commun des données *via* des consortiums et des bases de données[135]. Avant même que la terminologie ne les désigne comme données massives, ces études portaient déjà sur des cohortes de milliers de personnes. Elles visaient à déterminer des régions du génome (gènes ou séquences régulatrices) influençant la vulnérabilité des personnes à une maladie. La révolution technologique du séquençage de nouvelle génération (NGS) a rendu presque routinière l'analyse du génome complet, ou plus fréquemment de l'exome complet (le séquençage uniquement des parties codantes des gènes), qui apparaissait chimérique encore récemment. Le séquençage à large échelle est rendu possible par son coût de plus en plus faible (moins de 800 euros début 2019) et par la considérable augmentation des capacités de stockage et de traitement des données[136]. Il en résulte une utilisation élargie aujourd'hui en oncologie, pour l'identification de mutations tumorales pouvant constituer des cibles thérapeutiques chez un patient déterminé. Un autre domaine d'intervention porte sur l'exploration clinique de nouveaux gènes, siège de

---

6° Les traitements nécessaires aux fins de la médecine préventive, des diagnostics médicaux, de l'administration de soins ou de traitements, ou de la gestion de services de santé et mis en œuvre par un membre d'une profession de santé, ou par une autre personne à laquelle s'impose en raison de ses fonctions l'obligation de secret professionnel

11° Les traitements nécessaires à la recherche publique au sens de l'article L. 112-1 du code de la recherche, mis en œuvre dans les conditions prévues au 2 de l'article 9 du règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 précité, après avis motivé et publié de la Commission nationale de l'informatique et des libertés rendu selon les modalités prévues à l'article 28 de la présente loi. »

[133] Le génotype désigne les caractéristiques génétiques d'un individu, déterminées par l'analyse de son génome. Le phénotype désigne l'état physiologique ou pathologique d'un individu, par exemple l'expression d'une maladie. Par exemple, une mutation de l'ADN (génotype) aura pour conséquence le développement d'une maladie, qui représente le phénotype.

[134] Elles utilisaient la technique des « puces » à ADN et l'analyse à très grande échelle des polymorphismes génétiques humains identifiés - les SNP, *single nucleotide polymorphisms*.

[135] Une base de données est un ensemble structuré et organisé permettant la conservation de grandes quantités d'informations afin d'en faciliter l'utilisation. Il existe de multiples bases selon la nature des données (brutes ou annotées), selon qu'elles fournissent ou non directement les données et sont donc responsables de leur qualité.

[136] Les données produites par les technologies de séquençage à haut débit seront plus importantes que tout ce qui a jamais été produit dans le passé. Par exemple, le plan France Médecine Génomique 2025 prévoit de produire plusieurs dizaines de petaoctets (Po) de données par an d'ici 5 ans. Rappelons que l'ensemble des fichiers pour un génome humain (3 milliards de lettres A, T, G, C) représentent 300 gigaoctets (et 20 gigaoctets pour un exome).

mutations rares, dans la période périnatale, afin de limiter l'errance diagnostique. Enfin, l'utilisation des données massives a récemment démontré qu'en complément de l'analyse des variants rares, le risque résultant de l'accumulation de variants fréquents dans la population générale pouvait être estimé par un « score polygénique[137] ». Ce score peut représenter un marqueur clinique majeur et il est donc très probable que l'analyse du génome fasse partie, dans quelques années, du parcours de soin de toute maladie.

On assiste surtout à un changement de logique de la démarche. Elle va désormais d'un génotype à un phénotype, d'un génotype à la prédiction d'une maladie, et non plus seulement de la maladie au génotype. Cet accent mis sur l'évaluation du risque est ce qui fonde la constitution de grandes banques génomiques, souvent soutenues par les États et les institutions publiques[138].

Un fait majeur est que les données génomiques ne sont plus circonscrites à la sphère médicale classique, gérée par la déontologie médicale ou les organismes de recherche. Elles circulent aujourd'hui sur internet et les réseaux sociaux, avec peu de protection depuis l'avènement des tests dits *direct to consumer*. Des entreprises proposent depuis une dizaine d'années une analyse globale de l'ADN de leur client, supposée révéler l'ensemble des vulnérabilités génétiques dont il est porteur. Elles s'adressent directement à lui, sans faire intervenir médecins ou institutions médicales. En 2018, une compagnie proposait le séquençage du génome pour 199 dollars.

Le CCNE, dans son avis 124 et, plus récemment, son avis 129, a abordé les questions éthiques que pose l'élargissement des tests génétiques dans le domaine médical. Ont été évoquées spécifiquement les questions éthiques liées au recueil, à la conservation et au traitement des données génomiques. Celles-ci sont associées depuis quelques années par de très grands projets à d'autres données de santé, essentiellement en Amérique du Nord, au Royaume-Uni et, plus récemment, en Asie[139]. En France, où les cohortes étaient de taille réduite, Aviesan[140] a dévoilé en 2017 le plan France médecine génomique 2025, dont l'objectif est que « *la France se dote d'une filière médicale et industrielle en vue d'introduire la médecine de précision dans le parcours de soin et de développer une filière nationale en ce domaine* ».

[137] Le score polygénique représente l'ensemble des milliers de variants détectés dans certaines régions d'un génome et dont la combinaison peut donner une indication fiable du risque de développer une maladie.

[138] Citons la UK Biobank (500 000 personnes), mais aussi aux États-Unis le 100K *Wellness Project* et le *All of Us Research Program* (un million d'individus), en Asie la biobanque Kadoorie financée par le Wellcome Trust–funded China (515 000 chinois), et d'autres encore. Le NHS (*National Health Service* du Royaume-Uni) et le ministère de la Santé encouragent aussi les individus à payer pour faire l'analyse de leur génome (Genomics England), avec des garanties de fiabilité. Leurs données pourraient être mises à disposition des chercheurs *(Life Sciences Sector Deal 2, decembre 2018, UK government).*

[139] Citons aussi les *Precision Medicine Catapult Centre* (1 milliard de £) au Royaume-Uni, ou encore la *Precision medicine initiative* (250 millions de $) aux États-Unis.

[140] Créée en avril 2009, l'Alliance pour les sciences de la vie et de la santé regroupe les grands acteurs en sciences de la vie et de la santé en France (Inserm, CNRS, Inra, Inria, IRD, Institut Pasteur, conférence des présidents d'université).

3.3.2 Données massives en génomique : un enjeu pour la santé publique et la recherche

C'est probablement dans le domaine de la génomique que l'exploitation des données a permis les plus spectaculaires avancées des connaissances et la plus notable amélioration de la prise en charge des patients. L'identification des causes des maladies, l'analyse de la diversité génétique d'une catégorie de malades pour établir de nouvelles sous-classifications et l'exploration des mécanismes sont des étapes déterminantes pour découvrir de nouveaux traitements efficaces. Si le domaine de l'oncologie a été pionnier dans cette démarche, elle est aujourd'hui appliquée à d'autres maladies ou déficiences.

L'exploitation des données massives dans le domaine génomique a des conséquences majeures, présentes et futures. Elles s'étendent à la santé publique, à la prise en charge des patients et à la recherche, illustrant le lien étroit qui existe entre recherche fondamentale « pour comprendre la nature » et recherche appliquée ayant un objectif d'amélioration de la santé publique. Comme le disait déjà Louis Pasteur « *Il n'y a pas de recherche appliquée mais des applications de la recherche* ». Mais l'exploration de ces données génomiques souligne aussi le rapprochement des secteurs public et marchand, puisque son développement dans le domaine de la santé dépend des outils de séquençage, généralement apportés par les entreprises privées à but lucratif, et que les résultats, potentiellement valorisables, représentent une part importante du marché de la santé.

Les bases de données : un croisement indispensable

L'analyse du génome complet ou de l'exome identifie plusieurs milliers ou millions de variants. Seul un petit nombre – mutations causales de maladies monogéniques - a une signification connue et validée. La contribution des autres variants est actuellement inconnue ; elle peut être nulle, faible, ou associée à un risque. Certaines associations ne se détectent que par l'analyse d'une grande population de patients et de témoins. Il existe néanmoins des méthodes statistiques qui agrègent le risque cumulé par la présence de ces variants et qui donnent un « score polygénique » pour une maladie donnée et pour un individu donné. Ces méthodes sont en pleine expansion. Pour certaines maladies, un risque polygénique élevé constitue un risque aussi significatif que la présence d'une mutation rare[141].

Si ces méthodes génétiques sont de plus en plus performantes, nous devons rappeler que : « *Les informations qu'apporte l'analyse du « phénotype » sont au moins aussi importantes que celles du génome et c'est le croisement des deux types de données*

[141] Khera AV, *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics* 2018 ; 50 : 1219-24.

– génomiques et phénotypiques - qui caractérise au mieux l'état physiologique d'un « individu » (CCNE avis 124).

C'est l'analyse croisée d'au moins trois types de données, cliniques, génomiques et environnementales qui est la plus pertinente pour fournir le meilleur diagnostic et calculer la trajectoire clinique du patient.

Ces données peuvent être collectées chez un très grand nombre d'individus et conservées dans des grandes bases de données, contenant les données de plusieurs centaines de milliers de personnes. Cela permet d'identifier des corrélations entre les variants et des caractéristiques cliniques ou biologiques. De ces corrélations sont déduits des biomarqueurs utiles pour la prise en charge diagnostique de groupes de personnes/patients, des indications sur les causes des maladies, enfin des perspectives thérapeutiques ciblées.

C'est ce qui se fait aux États-Unis et en Chine avec des millions de participants, au Royaume-Uni avec des centaines de milliers de participants) et que propose, chez nous, à plus petite échelle, le plan France médecine génomique.

Il existe de multiples bases de données génomiques, nationales et internationales[142] : établissements hospitaliers (par ex. Institut Curie, Institut Gustave Roussy), plateformes de données à l'échelon national (UK Biobank, *Genomics England* et le *1000 genomes* émanant du secteur public *National Health Service* [NHS], projet *All of Us*, *Kadoorie Bank China*, et en France, celle que prévoit le plan France médecine génomique 2025) ; ou à l'échelon international (par ex. l'initiative GENIE[143] [sous l'égide de *l'American Association for Cancer Research*], la *Global alliance for genomics and health*[144], ICGC [*International cancer genome consortium*] ou encore, banques de données de *start-up* de biotechnologies[145] ou de firmes privées (par ex. *23&me*, Helix, etc.). Les questions juridiques et éthiques que posent l'organisation des bases de données internationales et le partage des données font l'objet d'importants débats. L'hétérogénéité des différents acteurs partageant parfois ces données, et notamment ceux venant des *biotechs*, effaçant la frontière entre les secteurs public et marchand, constitue un autre défi.

[142] Voir aussi note 140. Cook-Deegan R, *et al*. Sharing data to build a medical information commons: from Bermuda to the Global alliance. *Annual Review Genomics Human Genetics* 2017 ; 18 : 389-415.

[143] GENIE développe un registre qui permet de relier des données génomiques (déidentifiées) en oncologie avec des données cliniques de 50 000 patients traités en France (IGR), mais aussi dans les établissements d'autres pays.

[144] La *Global alliance* comprend 800 personnes appartenant à 400 organisations (dont l'Inserm) dans 70 pays, ce qui donne une idée du défi de la gouvernance acceptable par toutes les parties.

[145] Citons Arivale, Human Longevity, Verily, Amgen/deCode Genetics, Regeneron et iCarbonX.

Ces approches requièrent des moyens informatiques, humains et financiers considérables, et elles mettent en évidence des enjeux économiques et de souveraineté. Le partage est indispensable compte tenu des coûts d'acquisition.

L'objectif du plan France médecine génomique est, « *à l'horizon 2025, la couverture par la médecine génomique de l'ensemble des patients [atteints de cancer] concernés sur notre territoire. Cela implique de prendre en charge, à l'horizon 2020, environ 235 000 séquences de génomes par an.* »

3.3.3 Les données génétiques sont-elles des données de santé comme les autres ?

La très forte diminution de la durée d'analyse et du coût d'un séquençage (un génome humain peut être séquencé en 40 heures pour moins de 800 dollars), ainsi que les progrès dans la puissance de calcul et l'interprétation des séquences, font aujourd'hui de la séquence génomique (qu'elle porte sur le génome ou l'exome entier ou sur des panels de gènes) une donnée banale, dont l'obtention sera aussi facile qu'une donnée biologique[146], comme le dosage de cholestérol, ou d'imagerie comme une échographie. Dans quelques années, la séquence génomique sera probablement intégrée au DMP (désormais dossier médical « partagé » et non plus « personnel ») et bientôt l'espace numérique de santé, véritable carnet de santé numérique et instrument de la coordination des soins.

Les données de séquençage à haut débit seront les plus importantes, en termes de volume, de l'ensemble des données produites dans le domaine de la santé.

La question porte sur ce que l'on a appelé « *l'exceptionnalisme des données génomiques* », autrement dit ce qui les distingue d'autres données de santé et pourrait à ce titre justifier un traitement spécifique.

- Avant tout, c'est le *caractère unique* de la séquence génomique ; chacun des quelque sept milliards d'individus a un génome non seulement unique, mais invariant. La spécificité tient à l'existence de multiples variants génétiques, dont la combinaison est spécifique d'une personne donnée (environ 3 millions de variants distinguent deux individus). Ce qui n'est pas le cas pour d'autres données biologiques (plusieurs individus peuvent avoir un même taux de globules rouges, ou de cholestérol), ni même pour des résultats d'imagerie. Cette séquence d'ADN est donc « identifiante », au même titre que les empreintes digitales, et elle le reste tout au long de la vie puisqu'elle est invariante. C'est cette caractéristique qui en explique l'intérêt en matière judiciaire (notamment

[146] Plus de 20 % des données de séquençage du génome et de l'exome seront utilisées dans le contexte de la médecine génomique. En 2030, le nombre de génomes de maladies rares séquencés atteindra les 83 millions et 250 millions dans le cadre d'une démarche diagnostique pour un cancer.

via les empreintes génétiques[147]). Enfin, dans la mesure où le génome se transmet à la descendance, toute information déduite de la séquence concerne non seulement la personne, mais également son entourage familial. Se pose alors – plus que pour d'autres données - la question de l'information de la personne sur les résultats obtenus. C'est cette caractéristique identifiante qui crée les risques les plus importants dans l'exploitation des données génomiques, surtout l'exploitation des données massives : la séquence seule ne peut faire le rapprochement avec une personne donnée, mais le croisement avec d'autres données facilite grandement ce lien, ce qui explique l'impérieuse nécessité de respecter une grande confidentialité. La question se pose dès lors de la légitimité d'une réglementation particulière. L'ordonnance n° 2018-1125 du 12 décembre 2018 a ajouté à la loi informatique et libertés du 6 janvier 1978 un article 75, en vigueur à compter du 1er juin 2019, et qui réitère la nécessité du consentement exprès de la personne concernée pour l'examen de ses caractéristiques génétiques[148].

- Une deuxième caractéristique concerne *les étapes nécessaires pour obtenir la séquence d'ADN d'un individu.* Contrairement à d'autres données de santé, la production d'informations médicales à partir de la séquence génomique reste un défi et requiert de nombreuses étapes entre le séquençage de l'ADN et l'interprétation : Il est nécessaire dans un premier temps de séquencer l'ADN (détermination de la succession des nucléotides), puis d'assembler le génome (aligner les fragments de séquence) et d'annoter l'ensemble des variations génétiques identifiées. Ces étapes techniques sont effectuées en utilisant des algorithmes standardisés établis par des bio-informaticiens. Comme les méthodologies diffèrent et sont en constante évolution, il est important que ces différentes étapes soient transparentes car c'est à partir de ces données que le généticien va faire ses déductions sur la santé de l'individu.

- Une troisième caractéristique concerne *les étapes nécessaires pour déduire de la séquence d'ADN une information signifiante pour la santé de l'individu.* Actuellement, le généticien peut comparer les données de séquençage aux données de la littérature afin d'identifier la mutation responsable de la maladie chez un patient. Il peut s'aider d'un algorithme simple qui filtre les variations génétiques afin d'identifier les plus signifiantes. Avec les données massives et l'analyse algorithmique, il sera probablement possible de préciser le risque d'une maladie en prenant en compte non seulement les gènes déjà caractérisés comme responsables de la maladie mais aussi l'ensemble du génome qui contient des variants fréquents dans la population. Chacun de ces variants pris isolément a un effet mineur mais pris collectivement ils peuvent influer très for-

[147] Celles-ci sont basées sur la mesure de la longueur d'un nombre de séquence répétées de l'ADN, le nombre de répétitions étant spécifique d'un individu.

[148] *« Dans le cas où la recherche nécessite l'examen des caractéristiques génétiques, le consentement éclairé et exprès des personnes concernées doit être obtenu préalablement à la mise en œuvre du traitement de données. Le présent article n'est pas applicable aux recherches réalisées en application de l'article L. 1131-1-1 du code de la santé publique. »*

tement sur le risque[149]. C'est ainsi qu'en matière d'hypercholestérolémie, certaines personnes qui accumulent de nombreux variants fréquents ont un risque aussi élevé que des porteurs d'une mutation rare de type monogénique. Alors que l'identification d'une mutation causale pouvait être réalisée grâce à un panel de gènes et une analyse simple de comparaison avec la littérature, l'analyse algorithmique de l'ensemble du génome va permettre un diagnostic plus précis, mais en utilisant une telle quantité d'informations qu'il sera difficile pour le généticien de comprendre comment la prédiction a été faite. Il est aussi important de rappeler que les scores de prédictions peuvent être différents d'une population à l'autre. Actuellement la très grande majorité des données sur lesquelles se fondent les algorithmes proviennent de populations européennes. Ils sont beaucoup moins informatifs sur d'autres populations non européennes[150].

- Un point important est la *possibilité d'identifier des informations qui n'étaient pas sollicitées initialement*. Par exemple une personne vient consulter pour un diabète et grâce à l'analyse génétique, le généticien identifie que la personne est aussi porteuse d'un gène de vulnérabilité au cancer du sein. On parle alors de *données incidentes*, et leur découverte va devenir de plus en plus fréquente, avec le séquençage du génome entier. Cette information peut entraîner une démarche préventive empêchant l'apparition d'une maladie mais aussi inquiéter inutilement la personne si aucune prise en charge ne lui est proposée. C'est pour cela que le collège des généticiens cliniciens américains a établi une liste très limitée de gènes dit « actionnables[151] » pour lesquels une prise en charge adaptée est disponible (en janvier 2019, il existait 66 gènes actionnables[152]). Avec la généralisation des analyses génétiques et la restitution des résultats aux patients, la prise en charge des données incidentes deviendra probablement une étape inévitable de tous les parcours de soin.

- La puissance statistique nécessaire pour analyser certaines questions de génétique requiert un nombre de participants qui dépasse souvent les données disponibles dans un seul pays. L'échange de données internationales est rendu difficile du fait de l'hétérogénéité des législations et réglementations, mais aussi de la diversité des modes d'accès aux « gisements » de données.

3.3.4 Les principes éthiques confrontés aux avancées de l'exploitation des données génomiques

[149] C'est ce que l'on appelle le « score polygénique » : c'est l'analyse de millions de variants à de multiples endroits du génome (polygénique) qui donnera une information très robuste et précise en termes de prédiction d'une maladie commune. Voir aussi M. Waren. Polygenic scores : The power of many. *Nature* 2018 ; 562 : 181.

[150] *Martin AR,* et al. *Hidden 'risk' in polygenic scores: clinical use today could exacerbate health disparities.* BioRxiv. *https://doi.org/10.1101/441261. 11 Octobre 2018.*

[151] Il s'agit d'un anglicisme. Le terme « actionnable » est défini ici comme signifiant que des mesures préventives ou thérapeutiques peuvent être prises pour éviter la survenue des conséquences pathogènes de ce variant ou pour en modifier la progression naturelle d'une façon ou d 'une autre.

[152] https://www.ncbi.nlm.nih.gov/clinvar/docs/acmg/

Peut-être plus que dans d'autres domaines, l'exploitation des données génétiques crée une *tension forte* entre deux positions éthiques, l'une fondée sur les principes de bienfaisance et de non-nuisance, l'autre – ici en opposition – fondée sur les principes d'autonomie et de respect des personnes. Les perspectives de bénéfice médical, qui sont réelles et majeures, requièrent une exploitation accrue des données et un partage élargi. Mais une telle diffusion des données génomiques induit une crainte que le respect des droits fondamentaux de la personne soit mis en danger. Le risque est donc réel que les obstacles et les défis soulevés par l'accès à l'information génétique représentent un frein à la recherche avec pour conséquence un retard à l'acquisition d'un bénéfice médical.

La question est dès lors d'examiner comment développer l'accès aux données génétiques sans affaiblir les principes éthiques dont le respect est indispensable à la confiance. Nous en examinerons trois : le consentement et la transparence sur l'utilisation des données, la réponse aux risques de discrimination, l'information de la personne et la gestion des données incidentes, en insistant sur ce qui est spécifique aux données génomiques.

3.3.4.1 La protection de la personne et le respect de son identité.

Dès lors que le génome peut « identifier » non seulement la personne, mais aussi ses ascendants et ses descendants et touche à l'intime en ce qu'il dit quelque chose des maladies avérées ou possibles, il importe particulièrement de respecter la protection de la personne et son identité. Cela repose sur :

- *Le respect de la personne et le recueil de son consentement*. Conformément aux dispositions du code civil, du code de la santé publique, du RGPD et de la loi informatique et libertés (voir supra 1.4 et 2.1[153]), un *consentement écrit* des personnes concernées ou de leurs représentants légaux doit être recueilli si un examen des caractéristiques génétiques héréditaires ou acquises est envisagé. La liste des informations à donner à la personne est définie par le RGPD (art. 13)[154]. En outre, une analyse d'impact relative à la protection des données est

[153] CNIL : délibération n° 2018-153 du 3 mai 2018 portant homologation d'une méthodologie de référence relative aux traitements de données à caractère personnel mis en œuvre dans le cadre des recherches dans le domaine de la santé avec recueil du consentement de la personne concernée (MR-001).

[154] L'identité et les coordonnées du responsable de traitement ; celles du délégué à la protection des données du responsable de traitement ; la finalité du traitement de données (présentation du projet de recherche) ; la base juridique du traitement (article 6 du RGPD) ; la nature des informations qui seront utilisées dans la recherche ; les destinataires ou les catégories de destinataires des données ; les droits d'accès, de rectification, d'opposition, à l'effacement, à la limitation du traitement ; les modalités d'exercice de ces droits ; le caractère facultatif de la participation ; le cas échéant, le transfert de données à caractère personnel hors de l'Union européenne et la référence aux garanties appropriées ; la durée de conservation des données ; les informations prévues par l'article L.1122-1

requise pour la mise en œuvre de recherches portant sur des patients et incluant leurs données génétiques. Comme pour toute analyse génétique indépendamment des données massives, la personne doit faire un double choix : d'une part, accepter ou non l'étude (pour son bénéfice s'il s'agit d'un patient, ou pour celui de la collectivité s'il s'agit de participer à une cohorte ou une base de données) et, d'autre part, en connaître ou non les résultats, s'agissant de sa maladie ou de données incidentes (voir ci-dessous).
Mais les données génomiques massives soulèvent plusieurs questions : l'inadaptation du consentement au fur et à mesure que progresse la recherche, la difficile délimitation des contours de sa « finalité » dans ce contexte marqué par une forte évolution et un partage des données, notamment si un partage avec des firmes commerciales peut être envisagé[155], la décision de recontacter les participants à une recherche, le caractère fictif de l'anonymisation alors que l'on cherche à corréler données génomiques, cliniques, environnementales et comportementales pour une médecine de précision. Le RGPD prend en compte toutes ces interrogations (art. 4(11) et considérant 33). Il offre une certaine marge d'interprétation (art. 5(1)(b) et 6(4), considérant 50). On peut dès lors s'interroger sur l'efficacité du consentement comme mode de protection de la personne.

- Nous avons vu *le changement d'une logique* qui ne repose plus sur une garantie de sécurité, mais sur une information loyale sur le devenir des données, sur un consentement plus large et l'assurance que les risques seront minimisés *via* une gouvernance attentive, incluant les progrès de la bio-informatique[156], l'utilisation d'une pseudo-anonymisation[157], et des sanctions à l'égard des contrevenants aux bonnes pratiques.

- Cette information loyale doit se faire par l'établissement d'une relation de confiance entre la personne et le responsable du traitement de ses données, institué par le RGPD. La qualité de *l'information délivrée à la personne* quant au cheminement et à l'utilisation de ses données génomiques (finalité, type de traitement, conservation, non-exploitation commerciale) et celle des moyens

---

du code de la santé publique ; l'inscription du patient au fichier national des personnes qui se prêtent à des recherches.

[155] Le rapport d'activité 2016/2017 de la *UK Biobank ethics and governance council* mentionne par exemple l'autorisation donnée à deux firmes pharmaceutiques de séquencer l'exome des participants de la *UKBiobank*, soit 500 000 personnes. L'accord stipulait que les résultats du séquençage soient inclus dans la base de la *UK Biobank* et mis à la disposition des chercheurs. Une exclusivité d'utilisation des résultats pendant 9 mois était consentie aux firmes. La gouvernance de la *Biobank* a considéré qu'il n'était pas nécessaire de recontacter les participants puisque ceux-ci avaient accepté l'utilisation des échantillons et des données par des compagnies commerciales dans leur *consentement* initial. Une information a toutefois été adressée à tous les participants.

[156] Il est aujourd'hui possible d'interroger les bases de données des différents sites internationaux sans avoir à faire transiter les données.

[157] Traitement effectué « *de telle façon que les données ne puissent plus être attribuées à une personne concernée précise sans des informations complémentaires, pour autant que ces informations soient conservées séparément et soumises à des mesures techniques et opérationnelles* » pour empêcher leur attribution à une personne identifiée.

mis en œuvre pour assurer leur confidentialité sont fondamentales. Dans le contexte des données génomiques, se pose aussi la question de la restitution à *la personne de l'information résultant du traitement de ses données.* Ce point est particulièrement important s'il existe un manque de précision sur la finalité initiale de la recherche ou si celle-ci est évolutive.

Il faut distinguer deux types distincts d'informations, sur lesquelles la personne doit pouvoir exprimer son choix de savoir ou de ne pas savoir *:*

- l'information régulière des participants sur l'avancement de la recherche clinique ou fondamentale utilisant des données génétiques. Il s'agit d'une information générale qui ne concerne pas directement ni personnellement les participants ;
- l'information personnelle déduite de l'interprétation des variations identifiées dans le génome d'une personne déterminée, et notamment celle portant sur d'autres problèmes de santé indépendants de sa maladie d'origine. L'avis 124 a traité de cette question dans le cadre du soin, et nous évoquons au paragraphe 3.3.4.3 la gestion des données incidentes. Dans le cadre de la recherche, la restitution de cette information est un sujet très débattu : validité, signification clinique et « actionnabilité » sont en général requises pour décider d'une restitution à la personne, et seul un généticien est capable aujourd'hui d'expliciter cette information.

- Ces modalités d'information sont celles appliquées par les établissements de santé ou les institutions de recherche, dans un contexte sécurisé. Il ne faut toutefois pas sous-estimer la divulgation, par les personnes elles-mêmes, de leurs données génétiques, en réponse à des offres commerciales concernant la généalogie, ou pour créer des liens entre personnes, voire dans un objectif d'accélération de la recherche[158].

3.3.4.2 Les risques de discrimination.

L'interprétation des données génomiques, expose à plusieurs risques.

- *La stigmatisation de groupes à risque.* L'objectif de l'exploitation des bases de données génomiques et des études génétiques de la population est de trouver des corrélations entre l'existence de variants de séquence et des caractéristiques cliniques ou biologiques. Ces corrélations identifient donc des groupes d'individus ou de patients ayant certaines particularités – facteurs de risque de maladie, origine géographique – et pouvant être ciblés en raison de ces caractéristiques.

[158] Citons Helix, deCODEme, *personal genome project*, 23&me.

- *Une inégalité de prédiction génétique, en raison de biais dans la constitution des bases de données génomiques.* Une interprétation peu rigoureuse des données de séquençage ou fondée sur un jeu de données biaisé pourrait entraîner une perte de chance médicale. Or, il existe un biais de constitution des bases de données sur lesquelles sont entraînés les algorithmes, parce que certaines populations – en particulier non européennes – n'y sont pas suffisamment représentées[159]. Il en résulte que si les scores de risque sont fiables et de bonne qualité pour les populations européennes, ce n'est pas le cas pour les autres (asiatiques, africaines), avec une possible perte de chance[160]. Il faut rapprocher ce fait de l'absence d'égalité dans l'accès aux tests génétiques – et, plus généralement, dans l'accès aux avancées technologiques du système de santé – qui a été soulignée dans l'avis 129 du CCNE[161] et dans son rapport de synthèse des États généraux.

- *Le risque de recoupement* – qui existe aux États-Unis –des données génétiques conservées à des fins de généalogie dans des banques publiques avec celles qui ont été recueillies en matière de procédures judiciaires.

3.3.4.3 La gestion des données incidentes

L'analyse du génome permet l'identification de données pertinentes, mais sans lien avec la maladie sur laquelle portait la consultation ou le projet de recherche initial. Se pose dès lors la question éthique de savoir s'il faut ou non rechercher systématiquement les variants pathogènes.

Ces données incidentes deviendront extrêmement fréquentes, soit dans le cadre du soin, soit lors de l'exécution d'un projet de recherche, et l'atténuation de la distinction entre ces deux domaines ne fait que rendre la question plus complexe, la recherche clinique visant une analyse très large du génome afin d'identifier l'ensemble des variants à risque. Qui faut-il informer, quand et comment ? Comment respecter le droit de la personne de ne pas savoir ? Les données incidentes peuvent être cliniquement utiles et déboucher sur une action de prévention ou thérapeutique (variants dit « actionnables »). Elles peuvent aussi être signifiantes mais ne pas conduire, en l'état des connaissances, à une stratégie pouvant bénéficier à la personne. Elles peuvent enfin rester de signification incertaine. Tout ceci est susceptible d'évolution, ce qui pose la question du réexamen ultérieur des variants. Selon les bonnes pratiques actuellement admises en France, la personne est informée des découvertes incidentes s'il existe

[159] Parmi les participants de toutes les études génétiques, 79 % sont d'ascendance européenne, alors qu'ils ne représentent que 16 % de la population (GWAS catalogue). Popejoy, AB, Fullerton SM. Genomics is failing on diversity. *Nature* 2016 ; 538 : 161-4.

[160] *Martin AR,* et al. *Hidden 'risk' in polygenic scores: clinical use today could exacerbate health disparities.* BioRxiv *https://doi.org/10.1101/441261. 11 Octobre 2018.*

[161] CCNE avis 129, pp.75 ; Rapport de synthèse des États généraux, pp. 36-50.

une stratégie préventive ou thérapeutique (découvertes dites « actionnables »)[162]. La fréquence de découverte de mutations délétères touchant des gènes « actionnables » est actuellement de un à trois pour cent, mais cette fréquence ne fera que s'accroître. Il ne faut pas sous-estimer le désarroi que la découverte de ces données incidentes peut engendrer non seulement chez les personnes porteuses des variants, mais aussi chez leurs parents qui, potentiellement, peuvent aussi partager le risque détecté. Les cliniciens ont donc besoin d'être formés pour interpréter et gérer ces données incidentes, afin d'être mieux à même qu'aujourd'hui d'anticiper et d'accompagner les personnes.

Il est essentiel – quel que soit le contexte, clinique ou de recherche – que cette question de la restitution des découvertes incidentes soit anticipée avant toute prescription ou établissement d'une plateforme de données, et discutée avec la personne lors de la rédaction initiale du consentement (voir avis 129 du CCNE). L'intervention des CPP (comités de protection des personnes) ou IRB (*institutional review board*) qui valident les projets de recherche est souhaitable.

3.3.5 Risques de perte de souveraineté

Outre qu'elle ne concerne pas seulement une personne déterminée mais aussi sa parentèle, la spécificité de la gestion des données génomiques vient de la masse particulièrement importante qu'elle représente. Ceci requiert des infrastructures de stockage, mais aussi de calcul, capables de traiter des tera, peta et exabytes, et, enfin, de nouvelles compétences pour exploiter et interpréter les données. Or, comme le souligne le plan France génomique « *La France avec une capacité annuelle de 20 000 exomes et 10 000 génomes accuse un retard sensible face à ces pays capables de réaliser des dizaines de milliers d'analyses chaque année* ». Cela pose la question (voir ci-dessous, §3.4.2) de l'implication de la puissance publique dans la gestion de ces données, et du risque d'une perte de souveraineté. Il ne faut pas en effet ignorer les bénéfices que les structures privées (mais aussi les États où elles se trouvent) tirent des informations de grande valeur issues du traitement de la masse des données génétiques qu'elles collectent. « *La génétique s'est aujourd'hui intégrée à ce phénomène global d'économie numérique et participative, en reprenant ses modalités d'organisation et de fonctionnement. [....] À l'ère de l'économie numérique, qu'on le veuille ou non, le partage des données est devenu synonyme d'échange commercial, et la donnée génétique synonyme de capital*[163]. »

[162] Arrêté du 27 mai 2013 définissant les règles de bonnes pratiques applicables à l'examen des caractéristiques génétiques d'une personne à des fins médicales.

[163] Corto-Stoeklé H, *et al*. Le partage des données génétiques : un nouveau capital. *Médecine/sciences* 2018 ; 34 : 735-40.

3.4 Quelles réflexions au vu des problématiques nouvelles que révèlent ces différents contextes ?

Nous avons vu que la donnée – quel que soit le contexte envisagé (voir chapitres 3.1, 3.2, 3.3) – devient une ressource conservée dans une multiplicité de plateformes et de gisements ou entrepôts, parfois disséminés dans le monde entier, et accessibles à différents acteurs, qui les traiteront pour en tirer de nouvelles informations. On ne peut pas toujours connaître, lorsqu'elle est collectée, l'usage qui en sera fait, par qui et quand. Mais recueil et divulgation des données sont souvent à l'initiative des personnes concernées elles-mêmes, hors démarche de soin ou projet de recherche, ce qui a pu faire parler « d'ubérisation » de la santé.

Cette rupture temporelle et géographique entre la donnée et la personne dont elle est issue, déclinée ci-dessus dans différents contextes, suscite une réflexion éthique sur notre rapport à nos données, et ce dans trois aspects dont nous esquissons les perspectives : le consentement, la dimension internationale et notre propre divulgation des données.

3.4.1 Consentement individuel et confiance collective : quelles évolutions ?

Nous avons vu que le consentement est l'une des bases juridiques possibles du traitement des données personnelles relatives à la santé, mais que ce n'est pas la seule, sauf pour les données génomiques pour lesquelles un consentement exprès est toujours nécessaire (sauf le cas très particulier prévu par l'article L 1131-1 du CSP).

Même lorsque ce consentement est requis, il peut prendre diverses formes, essentiellement pour faciliter l'efficacité de la recherche, pour peu que le responsable du traitement ait, dès l'origine de son projet, pris toutes les dispositions utiles pour assurer au mieux l'information de la personne et la protection de ses droits individuels.

Mais face à cette conception individualiste de la relation de la personne à ses données et du consentement, certains proposent une vision différente. En effet, les données personnelles relatives à la santé, même si elles relèvent pour chacun d'entre nous de la sphère privée la plus intime, deviennent aussi – par leur mise en commun - les composantes d'un réseau d'informations utile à l'intérêt général.

Ce réseau constituerait un bien « commun » relevant d'une protection collective de la vie privée[164]. Deux éléments la justifieraient :

[164] Quel que soit le prestataire (public ou privé, voire l'État lui-même) en charge de cette protection, il lui incomberait de prendre les mesures appropriées pour garantir collectivement la sécurité des données et éviter des consé-

- dans un réseau, les données ne sont pas indépendantes les unes des autres – des données personnelles peuvent aussi révéler des informations concernant un autre individu, l'intervention sur l'une modifiant les autres ; elles deviennent « relationnelles » et ne pourraient donc pas être considérées comme relevant d'un enjeu purement individualiste ; leur utilisation devrait ne pas dépendre de la seule volonté de leur titulaire ;
- la protection du réseau serait plus adaptée que celle des données de chaque personne, maillon du réseau. Pour être efficace[165], une protection relevant d'une conception individualiste imposerait en effet des limites sévères au partage des données, sauf anonymisation pouvant provoquer une dégradation qui priverait le traitement d'une partie de son efficacité.

Dans cette approche, les « *données de santé, ne peuvent pas être protégées avec une approche libérale concentrée sur la maximisation des libertés individuelles (par le biais du consentement ou de la patrimonialisation de ces données) mais il serait plutôt nécessaire d'adopter une optique plus communautaire ou collective, qui pourrait exiger une limitation de certaines libertés individuelles, au nom de l'intérêt général et du bien commun.*»[166]

La notion d'intérêt général inspire aussi la revendication d'un « droit à la science[167] » qui, au nom du bénéfice que celle-ci rend à la population et de la nécessité de favoriser le progrès et l'innovation, permettrait de remettre en cause la nécessité même d'un consentement au traitement des données, qui pourrait être considéré comme un obstacle excessif[168]. D'autres plaident pour un contrôle moins exigeant. Ils estiment qu'un consentement individuel pourrait n'être plus exigé s'il existe une forte probabilité que le traitement contribue à l'amélioration de la santé de la personne elle-même et, au-delà, à celle de la collectivité (principe de réciprocité), lorsque le risque de nuisance est faible (principe de proportionnalité).

Une troisième conception a été proposée, intermédiaire entre la vision individuelle et la vision collective : celle d'une autonomie interactive et relationnelle, dans laquelle la personne gère ses données mais est intégrée dans une collectivité qui met en œuvre un projet collectif susceptible d'évolution et qui la protège. Cette conception se fonde

---

quences fâcheuses pour les personnes ou des atteintes à l'éthique. Voir aussi Pierre Bellanger : Les données personnelles : une question de souveraineté. *Le Débat* 2015 n° 183, pp. 14-25.

[165] Joly Y, *et al.* Are data sharing and privacy protection mutually exclusive? *Cell* 2016 ; 167 : 1150.

[166] Bourcier D, de Filippi P. Vers un droit collectif sur les données de santé. *Revue de droit sanitaire et social*, 2018 ; pp. 444-56.

[167] Knoppers BM, Thorogood AM. Ethics and Big Data in health. *Current Opinion in Systems Biology* 2017, 4 : 53-7.

[168] « *we must remember who gets left behind when consent is required*". Taylor P. When consent gets in the way. *Nature* 2008 ; 456 : 6.

aussi sur la nature « *dynamique* » *des données, qui « circulent et ne sont donc ni localisées, ni « propriétaires », mais relationnelles.* » [169]

Ces approches divergentes montrent que ce qui est au cœur du débat, c'est l'évolution du rapport entre l'individuel et le collectif, entre l'autonomie accrue de chacun et la protection que requiert l'utilisation généralisée des technologies recourant au traitement des données massives. Les progrès de l'une appellent le renforcement de la seconde. Leur conciliation ne va pas sans heurts et suscite des tensions car intérêt individuel et intérêt général ne coïncident pas nécessairement, du moins à court terme. Conceptions individuelle, collective et relationnelle peuvent toutes trois contribuer à la recherche d'un point d'équilibre. Celui-ci est toujours à redéfinir car il est constamment remis en cause par les progrès technologiques et par l'évolution des modes de vie.

(Voir RECOMMANDATION N°3)

3.4.2 La dimension internationale et la question de la souveraineté nationale

Le rapport de la CERNA[170] « *La souveraineté à l'ère du numérique. Rester maîtres de nos choix et de nos valeurs* » rappelle que la numérisation modifie profondément le problème que pose la maîtrise des données de santé puisqu'elle s'abstrait potentiellement des frontières géographiques et change les échelles de temps. Elle permet en effet une quasi instantanéité des échanges *via* internet et une conservation des données sans limite de durée. Elle peut ainsi permettre d'établir entre elles des corrélations signifiantes, en s'abstrayant de la souveraineté nationale des États.

Notre pays ne pourra conserver la maîtrise de l'innovation scientifique et médicale que laisse espérer le traitement des données, et donc l'évolution de son système de soin, s'il ne consent pas un effort important pour répondre à plusieurs défis :

- le défi technologique et humain du stockage, de la sécurité et de l'exploitation d'un volume sans cesse croissant des données de santé (voir § 3.3.5 les données génomiques). Le risque existe aussi de perdre la garantie d'une gestion des données respectant les principes éthiques (voir recommandation n°10). En France, la création de la plateforme de données de santé (*health data hub*), qui doit être annoncée dans la

[169] Le modèle d'un « *personal data ecosystem* » *qui serait centré sur la personne met en évidence une approche sociologique très égocentrique, alors même que les réseaux provoquent des circulations de traces, qui sont très partielles, très éphémères et très partagées. Nous avons proposé le concept « d'habitèle », qui permet de traiter la nouvelle enveloppe que les humains se créent, faite de données mais non centrée sur leur ego, mais sur leurs engagements situationnels et sur les affaires (issues) qu'ils doivent résoudre* » (Dominique Boullier, Sociologie du numérique, 2016 ; Armand Collin).

[170] La CERNA est la Commission d'éthique sur la recherche en sciences et technologies du numérique d'Allistene. Le rapport est disponible à https://www.allistene.fr/files/2018/10/55708_AvisSouverainete-CERNA-2018.pdf

prochaine Loi de santé et sera engagée au premier semestre 2019 est une première réponse.
- le défi de la recherche en mathématiques fondamentales, évoqué précédemment, doit permettre d'assurer un haut niveau technologique, garant de l'indépendance nationale et européenne pour l'exploitation des données et leur application dans le domaine de la santé.

Cette question est très prégnante aux États-Unis, et, plus récemment en Chine, deux pays dont les investissements financiers sont sans commune mesure avec ceux des pays européens, ce dont bénéficient leurs entreprises privées (GAFAM [Google, Amazon, Facebook, Apple, Microsoft] et BATX [Baidu, Alibaba, Tencent et Xiaomi]), qui sont dominantes technologiquement. En Chine, une volonté existe de définir des limites dans l'utilisation des données et une loi relative à la protection des données personnelles est entrée en vigueur le 1er juin 2017. Une des dimensions notables de cette loi est la contrainte de stocker les « données importantes » et les données à caractère personnel sur des serveurs localisés sur le territoire chinois. Il en résulte des tensions entre l'expression d'un nationalisme s'appliquant aux données, l'internationalisation des programmes de recherche et les intérêts des grands groupes pharmaceutiques[171].

Dans un monde où « *de grands opérateurs privés prétendent de plus en plus rivaliser avec les États et assumer des fonctions qui faisaient jusqu'à une date récente l'objet d'un monopole régalien* » (rapport précité de la CERNA), la garantie de l'institution pour l'accès aux données de recherche dont nous avons souligné l'importance au paragraphe 3.2.1.2 suppose que cette même institution ait une maîtrise suffisante des opérations concernant ces données pour que sa garantie soit efficace et puisse inspirer confiance. Nous avons évoqué précédemment l'investissement des géants américains dans l'hébergement des données en France, ce qui pourrait les inciter à faire d'autres offres de solutions efficaces pour leur exploitation, posant clairement la question de la souveraineté.

3.4.3 Les nouvelles pratiques de la e-santé, hors parcours de soins et sans réglementation précise

La banalisation des réseaux sociaux et d'internet comme source d'information ou d'accès à des services de santé, ainsi que la pratique du *quantified self,* incitent les personnes à divulguer elles-mêmes leurs données : recherche d'informations sur la santé, fréquentation de sites de communautés de patients, vente en ligne de médicaments ou de conseils médicaux, prises de rendez-vous médicaux, offres directes de

[171] *Cyranosky D. China's crackdown on genetic breaches could deter data sharing.* Nature *15 novembre 2018 ; et Morgane Tual : En Chine, une loi controversée sur les données personnelles et la cybersécurité. Le Monde -Pixels, 1e juin 2017.*

téléconsultations et prestations médicales électroniques. Une tendance qu'un rapport du Conseil national de l'ordre des médecins[172] a qualifiée « *d'ubérisation de la santé* ». Il peut aussi s'agir de la transmission en apparence anodine de données lors de la fréquentation d'internet ou de l'utilisation d'objets connectés (messagerie, réseaux sociaux et consultation de sites), données recueillies par les objets connectés (informations physiologiques, géolocalisation, achats, lieux fréquentés, habitudes de vie). Ces données ne deviennent que secondairement des données relatives à la santé, lorsque leur recoupement renseigne sur les paramètres de santé ou les habitudes de vie d'une personne. C'est ainsi, pour prendre un exemple, que la fréquentation assidue d'un établissement de soins spécialisé (déterminée par géolocalisation) et les achats répétés de certains produits ou des choix de régime alimentaire (repérés par des paiements à l'aide de cartes bancaires) permettent de savoir qu'une personne souffre d'une maladie et qu'elle suit éventuellement un traitement.

C'est pour ces données collectées et conservées hors relation de soin que se posent avec le plus d'acuité les questions relatives à la confidentialité, à la sécurité des données personnelles, à l'information sur leur traitement et leur devenir. Si ces sites sont maintenant tenus de respecter les obligations imparties par le RGPD, on peut s'interroger sur le respect des principes éthiques touchant la vie privée, et en premier lieu sur la valeur de l'information délivrée et du consentement obtenu. C'est d'ailleurs ce qu'illustre la récente sanction prononcée par la CNIL à l'encontre de Google pour manquement à l'exigence de clarté et d'accessibilité des politiques de confidentialité[173].

Les très grands opérateurs privés qui règnent sur ce marché agissent dans un cadre mondial. Ils sont essentiellement implantés aux États-Unis (GAFAM), même si la Chine tend à devenir un acteur majeur dans ce domaine (BATX). Or, si les États-Unis ont un attachement aux droits individuels comparable à celui des européens, ils ont toutefois une conception divergente de la protection des données personnelles et du consen-

[172] « *Le CNOM observe une tendance accélérée vers « l'ubérisation de la santé », par des offres en ligne qui correspondent à du commerce électronique non régulé et qui tendent à réduire la pratique médicale à une simple prestation électronique moyennant rétribution, via des plateformes du secteur marchand.* » Le CNOM posait la question « *L'État peut-il à la fois continuer de produire des textes réglementaires normatifs appliqués à l'exercice de la médecine utilisant des moyens numériques et laisser prospérer des offres numériques non régulées sur le marché de la e-santé ?* » et demandait « *l'instauration d'une régulation des offres numériques en santé, dans le respect de principes éthiques et déontologiques dans le champ sanitaire.* » (Rapport télémédecine et autres prestations médicales électroniques, février 2016).

[173] La formation restreinte de la CNIL prononce une sanction de 50 millions d'euros à l'encontre de la société GOOGLE LLC. « *Le 21 janvier 2019, la formation restreinte de la CNIL a prononcé une sanction de 50 millions d'euros à l'encontre de la société GOOGLE LLC en application du RGPD pour manque de transparence, information insatisfaisante et absence de consentement valable pour la personnalisation de la publicité* ».

tement à leur utilisation[174]. Pour résumer, on peut qualifier leur conception de contractualiste, alors que celle des européens est personnaliste[175].

Même si le règlement européen a prévu des dispositions visant à éviter que les règles protectrices qu'il instaure ne puissent être éludées, la logique des grands opérateurs demeure celle d'une offre sans cesse améliorée de biens et de services personnalisés, qui repose sur l'exploitation d'un nombre toujours croissant de données personnelles. L'utilisateur est avisé de ses droits mais, dans la plupart des cas, lorsqu'il souhaite profiter des fonctionnalités qui lui sont offertes, il donne son consentement sans en mesurer toutes les conséquences.

Mais la confiance du grand public est fragile, comme l'ont montré les réactions lorsqu'a été révélé, il y a quelques mois, le piratage à grande échelle de données collectées par Facebook[176].

C'est pourtant l'intérêt commun du public et des grands opérateurs que leurs relations puissent reposer sur une confiance mutuelle. Les opérateurs en sont conscients car ils savent que leur modèle de développement suppose la coopération du plus grand nombre. Pour préserver cette confiance, ils instaurent en leur sein des comités d'éthique, n'hésitant pas à y faire intervenir des personnalités particulièrement qualifiées, extérieures à l'entreprise[177]. Ces initiatives sont à l'évidence utiles et les réflexions ainsi engagées sont de nature à contribuer au débat qui est nécessaire sur l'utilisation éthique des données personnelles. Elles conduisent à mettre en avant des précautions et des procédures destinées à être portées à la connaissance du public, mais force est tout de même de constater qu'elles s'inscrivent dans l'objectif général d'efficacité de l'entreprise avec l'idée sous-jacente que ces garanties permettront d'obtenir une adhésion suffisante pour continuer d'obtenir le plus grand nombre possible de données personnelles. La multiplication des comités d'éthique dédiés, même s'ils peuvent être dotés de moyens en rapport avec ceux des entreprises qui les portent, ne doit pas nuire au maintien d'une réflexion éthique plus fondamentale, que peut seule mener une autorité indépendante, dégagée de tout objectif utilitariste. Le

[174] Rappelons que le RGPD doit être respecté si ces acteurs traitent des données de ressortissants de l'Union européenne, ou visent les ressortissants de l'UE.

[175] Aux États-Unis, la protection des données personnelles est fondée sur la logique de la liberté contractuelle et elle repose donc sur l'idée que le consentement exprimé par le titulaire des données suffit, même s'il apparaît largement formel. En revanche, en Europe, la protection est conçue comme relevant des libertés publiques et elle s'appuie sur des textes juridiques visant à garantir la protection de la vie privée, partant du principe que le déséquilibre entre les contractants et l'opacité du système ne permettent pas de se satisfaire d'un consentement formel.

[176] *Voir Le Monde-Pixels. D. Leloup et M. Untersinger. Faille Facebook : des données de 29 millions de comptes récupérées par les pirates. 12 octobre 2018.*

[177] Voir par exemple : Facebook lance un centre de recherche consacré à l'éthique de l'intelligence artificielle. Par Morgane Tual - Le Monde (Pixels) 20 janvier 2019. Cet institut, financé par Facebook et développé au sein de l'Université technique de Munich, se veut « indépendant », souligne l'entreprise.
https://www.lemonde.fr/pixels/article/2019/01/20/facebook-lance-un-centre-de-recherche-consacre-a-l-ethique-de-l-intelligence-artificielle_5411861_4408996.html

très grand intérêt des comités d'éthique dédiés est d'être un moyen privilégié pour les opérateurs de démontrer leur loyauté dans la manière dont ils doivent appliquer à leurs clients, notamment européens, les règles protectrices du RGPD. Au-delà du champ d'application de celui-ci, il revient aux opérateurs de démontrer de manière concrète et vérifiable l'importance qu'ils accordent à la protection de la confidentialité des données.

La loyauté, si elle est à l'évidence nécessaire, ne suffit cependant pas à fonder un rapport durable de confiance. Il apparaît aussi nécessaire d'informer et de sensibiliser le public sur tout ce qui relève de la divulgation de ses données personnelles de santé. Il doit notamment savoir configurer et utiliser ses objets connectés de manière à ne laisser accéder qu'aux seules informations nécessaires à ses besoins. Les associations de patients peuvent être à cet égard des intermédiaires très efficaces et les grands opérateurs d'internet qui interviennent dans le domaine de la santé auraient tout intérêt à les associer aux actions par lesquelles ils cherchent à s'assurer la confiance des utilisateurs. (Voir RECOMMANDATION N° 1).

CONCLUSION GÉNÉRALE

Face aux enjeux fondamentaux qui s'expriment aujourd'hui au cœur des sciences et technologies du numérique dont l'impact sur le futur de l'humanité sera probablement considérable, le CCNE a souhaité exprimer un soutien fort au développement de l'innovation dans ce domaine et affirmer l'importance de la vigilance sur la protection des droits fondamentaux et des libertés individuelles des personnes lors de l'utilisation de ces technologies dans le champ de la santé. Il lui est apparu nécessaire de rappeler l'invariance de certains repères éthiques sur lesquels se fonde le respect de la personne humaine. Celle-ci, qui est aussi représentée par les données numériques de santé qui la concernent, ne saurait être instrumentalisée. Le CCNE a souhaité analyser la manière dont ce socle de repères éthiques peut aider à répondre au défi que posent la complexité et la dimension internationale des applications numériques des données massives dans le domaine de la santé.

Plus que toutes autres données massives, celles relatives à la santé sont à la croisée de l'individuel et du collectif, de l'intime et du général, du public et du privé. Si, dans cette perspective, ce qui fonde la réflexion éthique doit être à nouveau énoncé et réaffirmé, cela n'exclut pas de faire évoluer le regard que nous portons sur elle, à la lumière de l'usage de ces données massives et de son appropriation par les personnes. C'est ainsi que la notion de vie privée répondait à une définition *a priori*, conçue comme devant résister à toute intrusion, alors que les modes de vie actuels évoluent vers une société en réseau où la sphère intime se redéfinit en permanence en fonction des relations que l'individu tisse avec son environnement. Il ne s'agit pas de la « fin de la vie privée » évoquée par certains, mais plutôt d'une sortie « *de la logique de la personnalisation de la vie privée, pour entrer dans celle d'une « négociation collective » inscrite dans un cadre dans lequel les autonomies et les libertés sont respectées by design* » comme le souligne Antonio Casilli[178].

Le traitement des données sera très certainement un facteur d'amélioration diagnostique et thérapeutique et, sans doute, d'amélioration de la santé. Ces progrès pourront-ils profiter à tous, alors que l'égalité d'accès aux soins n'est que très imparfaitement assurée ? Peut-il se créer, à travers l'utilisation des données massives, une autre forme de relation et de solidarité, au bénéfice de la communauté, de la société ? La révolution numérique ouvre des perspectives immenses mais la rapidité des évolu-

[178] Antonio Casilli. « Quatre thèses sur la surveillance numérique de masse et la négociation de la vie privée ». Jacky Richard et Laurent Cytermann. Étude annuelle 2014 du Conseil d'État "Le numérique et les droits fondamentaux", La Documentation Franc ¸aise, pp.423-434, 2014, études et documents, Conseil d' État.

tions technologiques bouscule nos repères. Elle nous fait percevoir plusieurs risques nouveaux, particulièrement exprimés lors des États généraux de la bioéthique, qui sont autant d'invitations à nous saisir des opportunités qu'offre cette mutation technologique :

- une exploitation automatisée des données dans un parcours de soin pourrait fragiliser l'écoute et le dialogue qui fondent la relation entre patient et personnel soignant ; ce constat nous invite à une réflexion sur la protection des personnes et sur le renforcement du lien de confiance qui pourrait être obtenu si l'automatisation de certaines tâches servait à dégager du temps pour privilégier la relation humaine ;

- le traitement indifférencié de données massives risque d'effacer les particularités de celles qui sont relatives à la santé et de leur appliquer une pure logique de marché, mais une utilisation pertinente des applications développées pour les besoins de cette activité économique peut aider à mesurer en vie réelle les paramètres de santé : si elle interroge sur la place à accorder, dans ce contexte, au droit de ne pas savoir, elle peut contribuer aussi à une responsabilisation des patients et à une meilleure information des personnels soignants ;

- le déficit de compréhension du grand public sur les processus de traitement des données personnelles relatives à la santé et sur les risques d'une divulgation indésirée[179] fait peser une menace sur les droits individuels, mais il est une puissante incitation à développer une information de qualité et à promouvoir un comportement éthique de tous les acteurs ;

- la possibilité de profilage, rendue possible par le croisement de données personnelles les plus diverses, fait peser une menace sur le principe de solidarité qui anime la mutualisation des risques, fondement du financement de notre système de santé, et elle comporte aussi un risque de discrimination au préjudice des personnes les plus vulnérables ; mais elle donne aussi les moyens de repérer plus précisément les facteurs d'inégalité, dans le but d'y remédier ;

- le contexte de la mondialisation, auquel sont confrontés les chercheurs et tous les acteurs publics et privés qui interviennent dans le domaine de la santé, nous confronte à des interrogations majeures sur les conditions de préservation de notre souveraineté, ce qui nous incite à consentir en ce domaine les efforts nécessaires pour développer nos savoir-faire, ainsi que les infrastructures que requiert le stockage et le traitement d'un nombre toujours croissant de données.

[179] Particulièrement exprimée lors des États généraux de la bioéthique. Voir CCNE rapport de synthèse pp 38-50.

LISTE DES RECOMMANDATIONS[180]

Les recommandations que propose le CCNE relayent et prolongent de très nombreuses initiatives privées ou publiques récentes. Elles expriment la certitude que de nouveaux équilibres doivent être trouvés dans l'exercice d'une démarche éthique qui doit accompagner les conséquences qu'introduisent les sciences et technologies du numérique, sans en freiner les bénéfices attendus, mais sans affaiblir les principes qui fondent la qualité d'être humain et la relation humaine. Elles énoncent une possible réponse éthique aux principales interrogations et tensions que suscite l'utilisation des données massives dans le domaine de la santé. Ces recommandations sont ordonnancées ci-dessous selon les principes qu'il nous paraît essentiel de pérenniser face à cette évolution, et que les États généraux et l'avis 129 du CCNE ont rappelés également : assurer l'autonomie de la personne, sa protection, et lui permettre d'élaborer les choix et les décisions qui la concernent ; respecter la liberté individuelle sans compromettre la solidarité et l'intérêt collectif ; permettre, dans le domaine de la recherche, l'acquisition de nouvelles connaissances au bénéfice de la santé de tous, sans céder aux risques de dérive.

- **Assurer l'autonomie de la personne et lui permettre d'élaborer les choix et les décisions qui la concernent**

RECOMMANDATION N° 1 (*2.1.3 et 3.4.3*)

Toute personne a droit à une information compréhensible, précise et loyale sur le traitement et le devenir de ses données, que son consentement soit ou non requis. Cette information, dont l'application effective doit pouvoir être évaluée, est adaptée à chaque contexte et actualisée. Elle doit en outre être aisément accessible, particulièrement lorsqu'elle s'adresse aux personnes les plus vulnérables, qui doivent recevoir, directement ou par l'intermédiaire de leurs représentants légaux le cas échéant, les incitations appropriées pour leur permettre de faire usage de leurs droits.

Pour que cette exigence d'information soit respectée et efficace, le CCNE estime que tous nos concitoyens doivent être sensibilisés aux spécificités des technologies numériques et aux avancées et risques qui leur sont associés, tant pour eux-mêmes que pour la société, afin qu'ils puissent faire un usage responsable de leurs données personnelles.

[180] Les chiffres entre parenthèses renvoient aux paragraphes du texte qui développent l'argumentaire ayant conduit à l'énoncé des différentes recommandations.

RECOMMANDATION N° 2 (2.1.3)

Compte tenu du rythme particulièrement important des innovations scientifiques et technologiques et des évolutions qu'elles déterminent dans le recueil et l'exploitation des données relatives à la santé, le CCNE estime qu'il est nécessaire d'évaluer périodiquement la mise en œuvre effective des dispositifs juridiques, afin de vérifier le maintien dans le temps de l'efficacité du système de protection des données personnelles qu'ils instaurent.

RECOMMANDATION N° 3 (3.2.1.1)

Dans ce contexte très évolutif, il est nécessaire de mener une réflexion sur la notion de consentement au recueil et au traitement de données massives dans un champ où prévaut l'hétérogénéité des acteurs et l'évolutivité des pratiques. Cette réflexion devrait porter sur l'objet du consentement, ainsi que sur les modalités de son recueil, afin d'assurer durablement l'équilibre entre le respect des droits des personnes et la dynamique des usages.

Cette réflexion devra nourrir le débat public sur les recommandations éthiques et permettra l'actualisation périodique de la loi.

RECOMMANDATION N° 4 (2.3.2 et 3.1.2.2)

La pertinence des décisions qui sont prises dans un parcours de soin à l'aide du traitement algorithmique de données massives repose sur la qualité de ces données, sur l'absence de biais dans leur sélection et sur la rigueur de la méthodologie utilisée pour leur traitement. Le CCNE estime que les résultats obtenus ne peuvent être évalués et validés que par une garantie humaine, condition d'une responsabilisation des acteurs. Cette garantie devrait être assurée par des instances indépendantes de contrôle.

En matière de recherche, il est nécessaire de préserver la diversité génétique des titulaires des données lors de la sélection des données traitées afin que les résultats obtenus ne soient pas faussés par une insuffisante représentation des populations non européennes.

RECOMMANDATION N° 5 (2.3.3 et 3.2.1)

Le CCNE estime que les professionnels de santé doivent bénéficier, lors de leur formation initiale et tout au long de leur carrière, d'une formation adaptée aux technologies numériques, aux principes éthiques qui régissent le recueil et le traitement des données, aux moyens à mettre en œuvre pour les respecter, et aux risques de biais qui résultent de leur non-respect.

Les experts de la gestion et de l'analyse des données massives (*data scientists*), ainsi que les chercheurs, doivent être suffisamment avertis des questionnements éthiques que soulèvent ces technologies pour pouvoir assurer efficacement la protection des droits fondamentaux et des libertés individuelles.

RECOMMANDATION N° 6 (2.3.3)

La multiplication de sites et d'applications qui donnent, hors parcours de soin, des conseils pour améliorer l'hygiène de vie et le bien-être, pose la question de la rigueur avec laquelle ils recueillent, interprètent et traitent des données relatives à la santé. Le CCNE considère que ces sites et applications doivent pouvoir être évalués ainsi que la qualité de l'information délivrée aux utilisateurs, afin d'éviter que certaines démarches insuffisamment rigoureuses ne puissent avoir des conséquences négatives sur le comportement et la santé des personnes.

RECOMMANDATION N° 7 (3.1.2.2)

Le CCNE rappelle que la confiance qui est au cœur de la relation de soin nécessite la préservation d'une relation personnelle directe entre le professionnel de santé et le patient. Celui-ci ne saurait être réduit à un ensemble de données à interpréter, rendant inutile son écoute et la prise en compte de son vécu. Aussi utile qu'elle puisse être pour aider au diagnostic et guider le traitement, la « donnée » ne saurait remplacer le dialogue.

Bien au contraire, l'utilisation par le professionnel de santé des technologies récentes doit aussi avoir pour but de libérer du temps pour l'écoute et l'échange en simplifiant le recueil des informations pertinentes. Elle devrait permettre au patient de devenir davantage l'acteur de son parcours de soin en lui permettant une appropriation de ses données, condition d'une attitude pleinement responsable.

- **Respecter la liberté individuelle sans compromettre la solidarité et l'intérêt collectif**

RECOMMANDATION N° 8 (2.2.3)

L'avènement d'une médecine de précision est de nature à favoriser d'importants progrès pour la prévention, le diagnostic et le traitement des maladies. Mais l'individuation du risque qu'il implique peut porter atteinte à la mutualisation dans les mécanismes de financement de la santé. Il contient ainsi le risque d'une dérive vers un profilage de nature discriminatoire, notamment pour des raisons économiques. Le

CCNE estime nécessaire que les acteurs de santé se montrent particulièrement vigilants pour la préservation de nos valeurs d'égalité, d'équité et de solidarité.

RECOMMANDATION N° 9 (2.3.1)
Le CCNE insiste sur la nécessité de veiller à ce que les personnes qui n'ont pas accès aux technologies du numérique, par exemple pour des raisons économiques ou de difficulté à comprendre leur mode de fonctionnement, bénéficient, comme les autres, des avancées dans le domaine de la santé et ne subissent ni pénalisation ni discrimination dans leur accès aux soins.

- **Permettre l'acquisition dans le domaine de la recherche de nouvelles connaissances au bénéfice de la santé de tous, sans céder aux risques de dérives.**

RECOMMANDATION N° 10 (3.1.2.3)
Face au défi technologique que posent, pour la souveraineté nationale et européenne, le stockage, le partage et le traitement des données massives dans le domaine de la santé, le CCNE préconise le développement de plateformes nationales mutualisées et interconnectées. Ouvertes selon des modalités qu'il faudra définir aux acteurs publics et privés, elles doivent permettre à notre pays et à l'Europe de préserver leur autonomie stratégique et de ne pas perdre la maîtrise de la richesse que constituent les données, tout en privilégiant un partage contrôlé qui est indispensable à l'efficacité du soin et de la recherche médicale.

Le CCNE souligne la nécessité d'un important effort de recherche scientifique afin de pouvoir assurer, avec un haut niveau de compétence, une évolution technologique du traitement des données dans le respect des principes éthiques.

RECOMMANDATION N° 11 (3.2.1.1)
Le CCNE estime qu'en matière de recherche l'impératif éthique doit être adapté à chaque situation particulière, de manière à justifier une relation de confiance entre les titulaires des données et ceux qui ont accès à ces données et qui les traitent. Il est essentiel que le titulaire des données soit informé des modalités par lesquelles l'autorité de contrôle assure sa fonction de tiers de confiance. Doit ainsi être assurée une triple exigence éthique :
(1) une évaluation rigoureuse et transparente de la pertinence des recherches justifiant l'accès aux données, qui doivent contribuer, au bénéfice de tous, à un enrichissement des connaissances dans le domaine de la santé ;

(2) un partage des informations relatives à la progression des recherches avec les participants, selon des modalités adaptées aux contextes ;
(3) l'assurance d'une sécurité des données, de leur traçabilité, et de l'absence d'usage malveillant.

RECOMMANDATION N° 12 (3.2.2)

Le CCNE considère qu'il est nécessaire de faciliter le partage des données de santé pour les besoins de la recherche. Il est notamment d'avis de permettre, pour des protocoles de recherche aux finalités strictement définies et dans le respect des droits des personnes dont les données ont été mises à disposition et avec leur consentement, l'accès des chercheurs à des données collectées sur internet ou les réseaux sociaux par des plateformes dont la gouvernance est contrôlée.

ANNEXES

ANNEXE 1

Membres du CCNE ayant participé au groupe de travail

Gilles Adda
Thomas Bourgeron
Laure Coulombel (rapporteur)
Pierre Delmas-Goyon (rapporteur)
Jean-Marie Delarue (membre du CCNE jusqu'en décembre 2017)
Claude Delpuech
Pierre-Henri Duée
Claude Kirchner
Martine Le Friant
Jean-Pierre Mignard
Francis Puech
Dominique Thouvenin (membre du CCNE jusqu'en décembre 2016)
Bertrand Weil (membre du CCNE jusqu'en décembre 2017)

ANNEXE 2

Personnalités auditionnées

Charles Auffray (European Institute for systems biology)
Thomas Bourgeron (Professeur de génétique, université Paris 6, Institut Pasteur)
Mathieu Cunche (Inria - Institut national de recherche dédié aux sciences du numérique)
Victor Demiaux (conseiller de la présidente de la Commission nationale de l'informatique et des libertés)
Mathieu Galtier (start up Owkin)
Jean-Gabriel Ganascia (Professeur d'université, Université Paris 6, laboratoire d'informatique de Paris 6)
(xxx confidentiel) (ancien conseiller juridique Google France)
Erwann Le Pennec (Professeur associé au département de mathématiques appliquées de l'École polytechnique)
Frédérique Le Saulnier (juriste, déléguée à la protection des données de l'Institut national de la santé et de la recherche médicale)
Alain Livartovski (Institut Curie, département de l'information médicale)
Sophie Narbonne (Commission nationale de l'informatique et des libertés)
Gilles Wainrib (start up Owkin et Assistance Publique-Hôpitaux de Paris)

AVIS 130

?

AVIS130

COMITÉ CONSULTATIF NATIONAL D'ÉTHIQUE
POUR LES SCIENCES DE LA VIE ET DE LA SANTÉ