

Valorisation de la recherche en santé humaine et protection des données à l'ère du numérique

Par Frédérique LESAULNIER

Docteure en droit, déléguée à la protection des données de l'Institut du cerveau

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

Qu'il s'agisse de données scientifiques issues du soin et du système de santé, de données médico-administratives recueillies initialement à des fins gestionnaires ou encore de celles issues de l'utilisation d'objets connectés et d'Internet, les données personnelles de santé sont un enjeu de premier plan pour la recherche en santé humaine. Leur exploitation permet chaque jour de nouvelles avancées dans la compréhension des maladies, des facteurs de risque, mais aussi une amélioration de la prise en charge et de la prévention.

Les chercheurs soulignent la nécessité de favoriser l'utilisation optimale de ressources précieuses produites dans l'intérêt général, qui ont nécessité une mise en qualité indispensable à leur exploitation pertinente. Ils relèvent qu'en l'absence de partage ou d'utilisation de ces données et/ou des échantillons biologiques, de nouvelles collectes devront être effectuées, engendrant de nouveaux risques pour les personnes, mais aussi des délais et des surcoûts. Ils insistent enfin sur l'importance de garantir le respect de la confiance accordée par les personnes concernées, en tirant de leur effort de participation tout le parti qu'elles pourraient souhaiter. C'est particulièrement vrai dans le cas de maladies rares, où la recherche scientifique fait naître de grands espoirs d'identifier les causes de ces maladies, ce qui répond à une attente des patients, notamment pour prévenir leur transmission, identifier les facteurs de sévérité et, surtout, trouver des traitements⁽¹⁾.

⁽¹⁾ Comme souligné dans l'avis 129 du Comité consultatif national d'éthique, « 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. »

Cela suppose que ces données sensibles soient exploitées avec la plus grande rigueur, l'expertise et l'esprit critique nécessaires, le tout dans le respect du cadre éthique, légal et réglementaire. Et c'est encore plus vrai pour des informations de nature à renseigner sur l'état de santé futur des personnes, et de toute la lignée de leurs apparentées génétiques.

Les données personnelles de santé, parce qu'elles relèvent de l'intimité de la vie privée des personnes, sont des données qui doivent faire l'objet d'une protection particulière. À ce titre, le droit reconnaît un statut particulier et impose le respect de règles ayant pour objet de garantir leur confidentialité. Elles sont ainsi soumises à un principe d'interdiction de traitement sauf pour un certain nombre d'exceptions prévues par la loi et assorties de garanties (de fond et de procédure) au respect desquelles la Commission nationale de l'informatique et des libertés (CNIL) veille.

Les garanties de fond sont les suivantes : une finalité de traitement déterminée, explicite et légitime ; des données adéquates, pertinentes et proportionnées au regard de l'objectif poursuivi (principe de minimisation des données) ; une durée de conservation déterminée à l'avance et dont la pertinence est appréciée au regard de la finalité poursuivie (droit à l'oubli) ; le respect des droits des personnes qui passe en premier lieu par le principe de loyauté et de transparence à leur égard ; et, enfin, la mise en place de mesures de sécurité de nature à garantir la confidentialité des données, qui est une composante majeure de la conformité s'agissant du traitement de données de santé.

Les spécificités du *Big Data* et de l'intelligence artificielle sont souvent présentées comme susceptibles de questionner certaines notions clés ou principes

cardinaux de la protection des données⁽²⁾. Ainsi, dans le contexte du *Big Data* et compte tenu des moyens techniques disponibles, l'anonymat devient relatif et des données réputées neutres peuvent devenir sensibles par recoupement. On est donc tenté de considérer la donnée comme personnelle à défaut de preuve contraire⁽³⁾ et de raisonner en termes de genre plutôt que d'espèce. Avec la médecine personnalisée, la ligne de partage entre le soin et la recherche tend à perdre de son évidence : ainsi, l'objectif des instituts hospitalo-universitaires est de favoriser le développement d'une recherche translationnelle qui bénéficie directement au patient ; les législations « Informatique et libertés » raisonnent, quant à elles, par type de finalités. Par ailleurs, le contrôle de la pertinence des données est présenté comme difficilement compatible avec l'approche exploratoire inhérente au *Big Data*. De même, la limitation de la durée de conservation des données est difficilement compatible avec la constitution de vastes bases de données ayant vocation à être ouvertes à la communauté scientifique. Enfin, les formalités administratives préalables et l'information individuelle dont doivent bénéficier les personnes avant chaque étude, consacrées dans l'optique de recherches ponctuelles, sont difficilement compatibles avec les extractions répétées et standardisées de lots de données très indirectement identifiantes. Il est donc nécessaire que les principes puissent être interprétés avec pragmatisme et souplesse.

Plusieurs dispositions du Règlement général relatif à la protection des données (RGPD) témoignent d'une prise en considération des enjeux de la recherche scientifique. Ce Règlement est devenu le texte de référence depuis le 25 mai 2018 dans l'ensemble de l'Union européenne et même au-delà, dès lors que l'on cible des résidents européens. Il maintient une dérogation au principe d'interdiction de traiter certaines catégories particulières de données, instaurée au bénéfice de la recherche scientifique, moyennant un certain nombre de garanties mentionnées à l'article 89 (art. 9-2-j). Il pose, notamment, le principe d'une présomption de compatibilité de la finalité de la recherche scientifique avec une finalité initiale différente et prévoit la possibilité de la conservation des données au-delà de la réalisation de la finalité initiale à des fins de recherche scientifique.

À l'heure des grands défis de la science ouverte, le RGPD modifie l'approche de la protection des données personnelles en renversant la charge de la preuve. Il opère un recentrage de l'organisation de la protection des données sur les organismes qui les traitent et/ou les mettent à disposition. C'est à eux qu'il appartient de prendre les mesures techniques et organisationnelles garantissant le respect des grands principes de protection des données et ils doivent être en mesure de démontrer que tel est bien le cas (*accountability*).

⁽²⁾ VULLIET-TAVERNIER S. (2014), « *Big Data* et protection des données personnelles : quels enjeux ? », *Statistique et société*, vol. 2, n°4, p. 27.

⁽³⁾ Avis sur les techniques d'anonymisation émis par le groupe de travail du G29 sur la protection des personnes à l'égard du traitement des données à caractère personnel, https://www.cnil.fr/sites/default/files/atoms/files/wp216_fr.pdf

Le respect de cette réglementation n'est pas uniquement une question de conformité à la législation en vigueur, même si cette dimension est essentielle. Les bonnes pratiques induites dans la collecte, le traitement, le stockage et la diffusion des données peuvent concourir à l'amélioration de la science elle-même et favoriser la confiance des personnes qui concourent à son progrès. C'est pourquoi les organismes de recherche développent des écosystèmes facilitant l'accès et l'exploitation des données dans le respect de la vie privée et de la confidentialité des données personnelles et des libertés individuelles, avec toute la rigueur, le bon sens et l'éthique indispensables à la confiance dans le numérique.

Nous interrogerons ici les valeurs à l'origine de cette protection et présenterons quelques pistes de réflexion et d'action développées par les acteurs de la recherche pour une utilisation éthique et responsable des données personnelles de santé.

La protection de la valeur éminente de l'être humain

Il existe un cadre juridique riche et ancien qui définit les conditions d'accès et d'utilisation des données personnelles de santé et qui traduit le caractère sensible de celles-ci.

Il repose en Europe sur le principe selon lequel la protection des données à caractère personnel est un droit fondamental inscrit dans la loi (Charte des droits fondamentaux de l'Union européenne, art. 8 §1).

L'application de ce cadre juridique est subordonnée à l'existence de données susceptibles de permettre d'identifier la personne, que cette identification soit directe ou indirecte par référence à un identifiant ou tout élément qui soit propre à cette personne et qui, seul ou croisé avec d'autres (un faisceau de données), permet son identification.

L'information personnelle est une information dont le trait est de porter une empreinte individuelle et de la conserver. C'est dans le rattachement par un lien d'attribution à une personne individualisée que réside la notion de donnée à caractère personnel. Chacune porte l'empreinte d'un individu singulier et demeure une émanation de cette personne qu'elle singularise et caractérise. Toutes, elles forment une masse, un corpus dont la personne apparaît à la fois comme le sujet et l'objet, de sorte que sans englober toute la personne, le corpus des informations personnelles contribue à la plénitude de son *animus*.

Les données personnelles sont des attributs de la personne humaine⁽⁴⁾. Révélatrice est à cet égard la référence à l'identité humaine dans la formulation

⁽⁴⁾ « Cette vie juridique, vouée à traduire, conformément à la justice, les besoins de l'homme, en y adaptant, dans leur continuuel mouvement, les structures de fait, ne peut, sans se tarir, méconnaître les constantes nécessaires, soit de l'homme, soit du droit » – SAVATIER René (1950), *Réalisme et irréalisme en droit civil d'aujourd'hui. Structures matérielles et structures juridiques*, Mélanges Ripert, Tome 1, p. 75.

d'objectif de l'article 1^{er} inchangé et toujours actuel de la loi du 6 janvier 1978 modifiée⁽⁵⁾.

Il faut attendre les années 1970 et l'émergence des nouvelles technologies de l'information pour voir la notion d'information à caractère personnel se détacher avec netteté et qu'une définition juridique en soit donnée ; on parlait alors d'informations nominatives. À l'heure où les développements technologiques et le large phénomène de décloisonnement qui les accompagne conduisent à l'écrasement des spécificités et à la circulation sans entrave d'informations sans ancrage avec leur contenu substantiel, il faut promouvoir, avec force, l'approche personnaliste de la protection des données personnelles, laquelle est justifiée par l'unité du genre⁽⁶⁾. C'est également la position du Conseil d'État qui considère que « s'il convient de renforcer la dimension de l'individu acteur dans le droit de la protection des données, c'est en envisageant celui-ci comme un droit à l'autodétermination plutôt que comme un droit de propriété »⁽⁷⁾.

Il faut donc, tout en favorisant l'usage des données personnelles dans l'intérêt de la santé publique, maintenir un régime de protection qui les considère fondamentalement dans leur lien avec la personne.

Développement d'une dimension participative de la recherche

Les droits des personnes concernées sur leurs données sont réaffirmés et renforcés par le RGPD. Ils participent des principes de transparence et de loyauté qui visent à leur conférer une meilleure maîtrise de leurs données personnelles (RGPD, art. 12 et s.).

Les chercheurs sont conscients de travailler avec des « données confiées » ; les volontaires entendent pouvoir décider de participer ou non aux différents projets en fonction de l'intérêt qu'ils en comprennent. Cela dépasse le seul respect du droit au consentement éclairé et relève bien plus d'un véritable « partenariat » de recherche, en particulier pour la mise en place de bases de données longitudinales, dont les données et/ou échantillons biologiques associés ont vocation à être réutilisés à des fins d'études, de recherches ou d'évaluations dans le domaine de la santé⁽⁸⁾.

⁽⁵⁾ « L'informatique – on parlerait aujourd'hui davantage des technologies numériques – doit être au service de chaque citoyen. [...] Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques. »

⁽⁶⁾ LESAULNIER F. (2005), *L'information nominative*, thèse, Paris II.

⁽⁷⁾ Conseil d'État, « Le numérique et les droits fondamentaux », Études 2014, p. 264.

⁽⁸⁾ Comité d'éthique de l'Inserm (2022), « Note d'étape sur le Health Data Hub, les entrepôts de données de santé et les questions éthiques posées par la collecte et le traitement de données de santé dites "massives" ».

Nécessité d'un accord et d'une information honnête et accessible réaffirmée par les textes

Il est rappelé, à titre liminaire, que le consentement des personnes n'est pas systématiquement requis pour les traitements de données personnelles réalisés à des fins de recherche scientifique. Le RGPD prévoit que chaque traitement de données personnelles repose sur une ou plusieurs conditions qui en fondent la licéité, des conditions dont la liste est précisée à l'article 6. Le consentement est l'une de ces bases légales au même titre que l'exécution d'une mission d'intérêt public dont est investi le responsable du traitement, tout comme le sont le respect d'une obligation légale ou l'intérêt légitime du responsable du traitement. S'agissant du traitement des données de santé, l'objectif de recherche scientifique est un critère de licéité indépendant du consentement (art. 9.2.j).

Le point de savoir si le consentement de la personne est requis dépend donc de la qualification réglementaire de la recherche. Mais le consentement à la participation à la recherche lorsqu'il est requis ne constitue pas pour autant nécessairement la base légale du traitement.

Rappelons qu'un consentement à la participation à la recherche (*opt-in*) est requis pour les recherches impliquant la personne humaine (RIPH)⁽⁹⁾ relevant des catégories 1^{o(10)} ou 2^{o(11)}, des recherches qui présentent par définition des risques ou des contraintes, et pour les recherches nécessitant l'analyse des caractéristiques génétiques. Les études non interventionnelles (RIPH relevant de la catégorie 3⁽¹²⁾), de même que les recherches portant exclusivement sur des données ou des échantillons biologiques déjà acquis (NRIPH) sont soumises, quant à elles, à un régime d'opposition de la personne (*opt-out*).

Enfin, si à l'occasion d'une recherche, il est envisagé de constituer un « entrepôt de données »⁽¹³⁾, c'est-à-dire une base de données ouverte à des fins de recherches

⁽⁹⁾ Cette expression introduite par la loi n°2012-300 du 5 mars 2012 désigne les recherches organisées et pratiquées sur l'être humain en vue du développement des connaissances biologiques ou médicales (art. L. 1121-1 du CSP).

⁽¹⁰⁾ La catégorie 1 concerne les recherches interventionnelles qui comportent une intervention sur la personne non justifiée par sa prise en charge habituelle.

⁽¹¹⁾ La catégorie 2 concerne les recherches interventionnelles ne comportant que des risques et des contraintes minimales (liste fixée par arrêté du 12 avril 2018). Exemples de ce type de recherches : prélèvements de sang ou d'échantillons biologiques effectués pour les besoins de la recherche, réalisation de scanners, entretiens ou questionnaires dont les résultats peuvent conduire à la modification de la prise en charge médicale habituelle du patient.

⁽¹²⁾ La catégorie 3 se rapporte aux recherches non interventionnelles, c'est-à-dire celles qui ne comportent aucun risque ni contrainte et dans lesquelles tous les actes sont pratiqués et les produits utilisés de manière habituelle (liste fixée par arrêté du 12 avril 2018). Quelques exemples : recueil d'éléments ou produits du corps humain ne présentant aucun caractère invasif, enregistrements audio et/ou vidéo, questionnaires ne pouvant conduire à la modification de la prise en charge des personnes.

⁽¹³⁾ C'est l'expression utilisée par la CNIL qui a publié, en 2021, un référentiel relatif aux traitements de données à caractère personnel mis en œuvre à des fins de création d'entrepôts de données dans le domaine de la santé.

menées ultérieurement, il est recommandé de recueillir le consentement exprès des personnes concernées par cette réutilisation secondaire des données et/ou échantillons biologiques. En effet, si l'entrepôt n'est pas fondé sur le recueil du consentement, une autorisation de la CNIL sera nécessaire (LIL, art. 65.2°), à défaut de conformité au référentiel homologué par cette autorité. Le consentement, s'il est recueilli, ne couvrira pas les recherches ultérieures qui nécessiteront une information spécifique préalable assortie d'un droit d'opposition, mais il sera de nature à fonder la constitution de l'« entrepôt ».

Quels que soient la qualification réglementaire de la recherche et le régime du recueil de l'accord des personnes, l'information de ces dernières est essentielle. En matière de recherche en santé humaine, cette transparence est une garantie reconnue aux participants en contrepartie de la levée du secret professionnel. Elle leur permet de comprendre les objectifs de la recherche, les modalités de leur participation, la portée de l'accord qu'ils donnent et de maîtriser l'utilisation qui sera faite de leurs données. Elle conditionne et facilite l'exercice effectif de leurs droits. Elle permet de développer la compréhension et d'instaurer une relation de confiance avec les chercheurs et l'organisme dont ils dépendent.

Par principe, l'information doit être délivrée individuellement à chaque personne participant à la recherche, que les données soient recueillies directement auprès d'elle ou par l'intermédiaire de tiers (LIL, art. 58). L'information individuelle doit être doublée d'une information générale dans les structures de soins qui transmettent des données personnelles de leurs patients pour permettre des activités de recherche (affichage dans leurs locaux, mention dans le livret d'accueil, etc.).

Si le RGPD favorise l'utilisation de ces données à des fins de recherches scientifiques en posant le principe d'une présomption de compatibilité des traitements ultérieurs (RGPD, art. 5 b), il impose cependant que les personnes concernées en soient individuellement informées⁽¹⁴⁾. L'information doit être réalisée pour chaque projet auquel le patient participe ou pour lequel les données du patient feront l'objet d'un traitement.

Le rôle des patients ou des personnes à l'origine des données est majeur et leurs droits à être informés et à donner leur accord à la réutilisation de ces données et échantillons doivent être respectés. Pourtant, l'exigence d'une information individuelle et spécifique à chaque projet issu de ces données et/ou échantillons est difficilement compatible avec le fonctionnement des biobanques, des cohortes, notamment celles financées par le programme d'investissements d'avenir, qui ont vocation à créer une ressource pour la communauté des chercheurs.

À la faveur du numérique, des modalités d'exercice des droits plus souples et plus dynamiques

Comme le souligne Georges Dagher au sujet des biobanques, « [...] Il est temps de repenser le rôle des personnes-sources plus largement en termes de participation et de contribution à la recherche. Le nouveau paradigme développé par l'utilisation des collections biologiques et visant à créer une ressource pour la recherche invite à une évolution du cadre réglementaire et éthique qui régit la question de la participation des patients aux projets de recherche »⁽¹⁵⁾.

De plus en plus, les patients ou leurs représentants au travers des associations intègrent les organes de gouvernance des infrastructures qui permettent l'utilisation secondaire des données (cohortes, biobanques, entrepôts de données...). Ils contribuent ainsi à la définition des orientations stratégiques, des principes directeurs et des procédures d'accès aux données qui seront mises en place. Ils participent également à la relecture des documents remis aux participants et sont informés régulièrement des résultats globaux des recherches. Les responsables de cohortes soulignent l'importance de cette dimension participative dans le fonctionnement en routine des cohortes, une dimension qui résulte du contrat de confiance qui lie le scientifique aux volontaires.

En outre, il n'est pas toujours réaliste ni éthique de recontacter trop fréquemment les personnes concernées, en particulier quand elles sont atteintes de maladies graves ou incurables, pour les informer de chaque utilisation de leurs données et leur permettre, à cette occasion, de s'y opposer. Une telle information ne permet pas à une personne constamment sollicitée de disposer d'une vue d'ensemble de l'utilisation des données qui la concernent.

À la faveur du numérique, d'autres modes d'information et de recueil du consentement voient le jour et sont développés, tels que la mise en place dans la durée de « portails de transparence », qui sont assortis d'une communication régulière sur l'utilisation des données et d'un droit de retrait dynamique, permettant d'adapter, le cas échéant, le consentement initial de façon flexible.

Il faut saluer le pragmatisme de la CNIL sur ce point, illustré par la rédaction de la méthodologie de référence MR004⁽¹⁶⁾. Celle-ci admet que des données et/ou des échantillons biologiques puissent faire l'objet d'une réutilisation et que l'information puisse être considérée comme valablement délivrée sans qu'il soit besoin de procéder à une nouvelle information individuelle des personnes : cela nécessite que l'information délivrée lors de la collecte initiale des données mentionne clairement la possibilité de les réutiliser et renvoie à un support d'information dynamique tel qu'un site Internet,

⁽¹⁴⁾ CEPD (Comité européen de la protection des données), lignes directrices 5/2020 sur le consentement au sens du règlement (UE) 2016/679, p. 35. « La transparence [ultérieure] constitue une garantie complémentaire lorsque les circonstances de la recherche ne permettent pas un consentement spécifique. »

⁽¹⁵⁾ Article paru dans le supplément « Sciences & Santé » du journal *Le Monde* du mercredi 8 juillet 2015.

⁽¹⁶⁾ Délibération n°2018-155 du 3 mai 2018 portant homologation de la méthodologie de référence relative aux traitements de données à caractère personnel mis en œuvre dans le cadre des recherches n'impliquant pas la personne humaine, des études et évaluations dans le domaine de la santé (MR-004).

centralisant les informations relatives à l'ensemble des projets menés et leurs caractéristiques et auquel les personnes pourront se reporter pour, notamment, exercer leur droit de retrait s'ils le souhaitent. Ce type d'approche, conforme au rapport du 15 septembre 2017 du Comité international de bioéthique (CIB)⁽¹⁷⁾, apparaît de nature à favoriser l'utilisation des données en matière de recherche, tout en préservant l'autonomie des patients qui deviennent des parties prenantes au projet.

Cette souplesse suppose toutefois que le devenir des données personnelles et des échantillons biologiques collectés dans le cadre d'un projet initial ait été anticipé et que les personnes aient été renvoyées vers un « portail de transparence », dont elles ont été informées de la localisation de façon claire et précise.

À défaut, il faudra recontacter les personnes pour les informer de l'utilisation secondaire des données. Une dérogation à l'obligation d'information sera possible, mais une absence d'information des personnes rendra les recherches subséquentes non éligibles aux mesures facilitatrices concernant les traitements jugés les plus courants, que sont les méthodologies de référence. Une demande d'autorisation, spécialement motivée, devra être effectuée auprès de la CNIL, ce qui sera de nature à retarder la mise en œuvre à l'échelle nationale ou internationale de projets de collaboration qui présentent des enjeux de santé publique.

Dans un contexte évolutif et longitudinal où prévaut l'hétérogénéité des acteurs, il est également nécessaire d'envisager d'autres modes de consentement pour assurer durablement l'équilibre entre le respect des droits des personnes et la dynamique des usages des données personnelles.

Une réflexion en lien avec les chercheurs et les partenaires doit être menée sur l'objet du consentement à la réutilisation des données/échantillons et les modalités de son recueil et de son maintien dans le temps⁽¹⁸⁾. Le RGPD permet le recueil d'un consentement « pour une ou plusieurs finalités spécifiques » (art. 6.1.a), ce qui suppose que les finalités aient été déterminées et que la personne concernée en ait été informée. Il résulte de la lecture du considérant 33 du Règlement qu'une finalité spécifique est compatible avec un consentement global. Ce considérant prévoit en effet que les personnes concernées devraient pouvoir donner leur consentement « pour ce qui concerne certains domaines de la recherche scientifique dans le respect des normes éthiques reconnues en matière de recherche scientifique. »

Les attributs de la personne humaine étant indisponibles, une personne ne devrait pas pouvoir consentir à l'utilisation de ses données à des fins de recherche scientifique sans autre précision⁽¹⁹⁾. En revanche, la

personne concernée pourrait être en mesure d'accepter que ses données soient utilisées dans le cadre de différents projets de recherche susceptibles d'être menés dans une branche ou un domaine particuliers pour lesquels elle autorise l'utilisation de ses données (par exemple, la recherche en oncologie) ou encore dans le cadre d'une infrastructure qu'elle aura choisie selon des modalités de partage définies⁽²⁰⁾.

On pourrait également imaginer un consentement à options, la personne pouvant, par exemple, accepter un partage de ses données avec des organismes académiques, mais pas avec des industriels ; ou un partage de ses données quand leur utilisation reste au sein de l'Union européenne, mais pas quand elles circulent en dehors de cet espace.

Les algorithmes sont susceptibles de constituer de précieux outils d'aide à la décision⁽²¹⁾. Ils constituent également un sujet de réflexion pour la recherche. Outre les questions des données personnelles et de la qualité et de la fiabilité des sources des données issues d'un traitement algorithmique, se pose la question de la clarté et de l'intelligibilité nécessaires à l'explicabilité qui pèse sur les concepteurs, afin que les professionnels de santé utilisateurs puissent effectivement exercer un contrôle sur les résultats du traitement dans le respect de la garantie qu'apporte une intervention humaine (CSP, art. L. 4001-3).

Le développement d'une gouvernance des données

L'implication des personnes à l'origine des données est essentielle. Mais la protection des données dépend également de la capacité à développer des écosystèmes facilitant l'accès et l'exploitation de données pseudonymisées avec toute la rigueur et la qualité scientifiques qui s'imposent, ainsi que dans le respect de la vie privée et de la confidentialité des données.

La protection des données est un sujet de gouvernance transversale qui demande aux organismes de recherche, en lien avec leurs partenaires, d'assurer une meilleure coordination de l'action des communautés de recherche et de leur proposer un accompagnement. Comme le souligne Romain Boidin, « les contraintes réglementaires sont l'occasion de se prémunir contre les risques liés aux fuites ou aux pertes de données,

⁽²⁰⁾ Voir, en ce sens : le CCNE (2018), *Numérique & santé : quels enjeux éthiques pour quelles régulations ?*, en ligne sur Internet : « Il serait donc de bonne pratique, dans toute la mesure du possible, de recueillir un consentement qui permette aux personnes d'autoriser le partage de leurs données en sachant comment elles vont être partagées (plan de partage), plutôt que pourquoi (par qui et pour quelle recherche). »

⁽²¹⁾ Pilotée par Stanley Durrleman, Thomas Nedelec, chercheur à l'Institut du cerveau, et Carole Dufouil, directrice de recherche à l'Inserm, une étude permettant d'identifier les facteurs de risque de démence due à la maladie d'Alzheimer grâce à l'analyse de données massives vient d'être publiée dans la revue *The Lancet Digital Health*, le but étant de détecter le plus tôt possible les signes précoces et facteurs de risque pour retarder le début de la maladie.

⁽¹⁷⁾ UNESCO, rapport du CIB sur les mégadonnées et la santé, Bibliothèque numérique, https://unesdoc.unesco.org/ark:/48223/pf0000248724_fre

⁽¹⁸⁾ CCNE, avis n°130, 28 mai 2019.

⁽¹⁹⁾ Voir, en ce sens : le G29 indique que mentionner « des fins de recherche » n'est pas suffisamment clair, car le type de recherches visées n'est pas précisé. Lignes directrices sur la transparence au sens du règlement (UE) 2016/679.

et d'enclencher, dès aujourd'hui, la transition numérique du soin et de la recherche »⁽²²⁾.

Le développement de solutions d'hébergement et de mise à disposition de données sécurisées et mutualisées permettant de garder leur maîtrise et présentant des garanties de conformité technique et réglementaire certifiées est indispensable. Il nécessite le développement d'architectures techniques appropriées et régulièrement réévaluées. De nombreux organismes sont actuellement confrontés sur ce point à la question du choix de solutions techniquement robustes non soumises exclusivement aux juridictions européennes en l'absence d'alternatives françaises ou européennes.

De même, doivent être mises en place en concertation avec l'ensemble des partenaires, des procédures et des chartes d'accès aux données permettant de faciliter l'évaluation scientifique et éthique des projets de recherche, selon des procédures de partage respectueuses du cadre légal et réglementaire.

Renforcer l'acculturation et l'appui aux chercheurs

Il n'est pas simple pour les chercheurs et les institutions qui les accompagnent de comprendre et de respecter les exigences de la réglementation.

Un environnement juridique complexe

La lisibilité du droit pour les acteurs de la recherche n'est pas facile. Le RGPD doit être articulé avec la loi « Informatique et libertés » française⁽²³⁾ qui maintient un régime d'autorisation pour les recherches, études ou évaluations réalisées dans le domaine de la santé, dès lors qu'elles ne sont pas conformes à une méthodologie de référence (art. 66 III). Cette même loi doit aussi être articulée avec d'autres dispositions de l'UE et nationales applicables à la recherche⁽²⁴⁾, notamment les dispositions du Code de la santé publique applicables aux recherches impliquant la personne humaine et au système national des données de santé (SNDS), et les dispositions du Code civil. De nouveaux règlements européens sont en cours d'élaboration qui devront, eux aussi, s'articuler avec le cadre juridique existant⁽²⁵⁾.

La préparation et l'instruction réglementaire des dossiers de demandes d'autorisation pour des études sont devenues plus complexes ; elles requièrent une

expertise juridique et technique toujours plus importante. Les procédures administratives demeurent complexes, faisant intervenir une multiplicité d'organismes selon que la recherche implique ou non la personne humaine.

Dans ce contexte, les chercheurs doivent pouvoir bénéficier d'un accompagnement expert qui leur assure, en amont de la soumission, une conformité de leurs dossiers à la réglementation et aux attentes de la CNIL et qui, finalement, facilite l'obtention des autorisations dans des délais compatibles avec ceux parfois très contraints des recherches (projets sur l'étude de la pandémie de Covid-19...). En parallèle de cette évolution, le respect des contraintes réglementaires évoquées précédemment devient un élément déterminant pour la participation à des projets internationaux et européens.

Mise en place d'un dispositif de formation des chercheurs et d'autorégulation adapté aux spécificités du RGPD

Les organismes de recherche doivent se doter d'équipes pluridisciplinaires dédiées, à même de relever les défis juridiques et techniques posés par la réglementation. La fonction de délégué à la protection des données (DPD, ou DPO pour *Data Protection officer*) a été rendue obligatoire dans tous les organismes dont l'activité est d'effectuer des traitements à grande échelle de données sensibles. Son rôle est de contribuer, en lien avec l'ensemble des acteurs présents dans l'organisme et les partenaires de celui-ci, à l'élaboration et à la déclinaison de politiques institutionnelles favorisant la protection des données. Il est également de faciliter la tâche des scientifiques pour leur permettre de se consacrer à la production de connaissances utiles à la santé de tous. L'objectif est de diffuser une culture et de bons réflexes, l'enjeu étant la déclinaison opérationnelle de la protection des données personnelles au sein de l'organisme.

Pour ce faire, des campagnes d'information et de formation au RGPD et à ses contraintes doivent être réalisées régulièrement afin de sensibiliser les personnels, chercheurs comme administratifs.

Par ailleurs, la délégation à la protection des données de chaque organisme doit proposer des supports pédagogiques permettant aux scientifiques de s'approprier le cadre juridique et de les guider de façon didactique dans leur démarche de mise en conformité, tels que des guides de bonnes pratiques assortis de modèles permettant une auto-régulation par la communauté scientifique.

Enfin, les dossiers complexes, stratégiques et innovants, qui laissent plus difficilement place à une standardisation des procédures, doivent être évalués en recourant à des compétences scientifiques et juridiques croisées. Il faut capitaliser sur l'expérience acquise pendant la pandémie de Covid-19 et associer les délégations à la protection des données au montage des projets de recherche, et ce dès les premières

⁽²²⁾ BOIDIN R. (2019), *Protection et gouvernance des données dans la recherche en santé*, thèse soutenue pour l'obtention de son doctorat d'État en pharmacie.

⁽²³⁾ Loi n°78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés.

⁽²⁴⁾ Notamment le règlement (UE) 536/2014 du 16 avril 2014 relatif aux essais cliniques de médicaments à usage humain et abrogeant la directive 2001/20/CE.

⁽²⁵⁾ Le 21 avril 2021, la Commission européenne a publié sa proposition de règlement établissant des règles harmonisées en matière d'intelligence artificielle. Le 3 mai 2022, la Commission européenne a présenté une proposition de règlement dans le but d'instituer l'espace européen des données de santé (EHDS), https://ec.europa.eu/health/publications/proposal-regulation-european-health-data-space_fr

réunions de cadrage⁽²⁶⁾. Ce travail d'anticipation réalisé en équipe, dans l'esprit du RGPD et en conformité avec les lignes directrices du Comité européen de protection des données⁽²⁷⁾, est essentiel pour éviter les « erreurs d'aiguillage », qui obligent à des allers-retours et à des modifications ultérieures coûteuses en temps et en énergie. Cette écoute en amont des scientifiques permet de bien comprendre leurs besoins et de trouver avec eux les adaptations nécessaires des protocoles permettant de concilier ces besoins avec les exigences de la réglementation. Bien plus, cela permet aux chercheurs de faire des choix méthodologiques mieux éclairés, qui contribueront à l'élaboration plus rapide de dossiers solides permettant ultérieurement une valorisation agile des données personnelles dans le respect de la transparence à l'égard des personnes concernées. Cela facilite également le dialogue qu'entretiennent les DPO avec la CNIL, les partenaires, les participants aux études et les autres acteurs.

⁽²⁶⁾ LESAULNIER F. (2021), « La protection des données personnelles au cœur de la recherche Covid-19 à l'Inserm », *JDSAM*, n°29, p. 33.

⁽²⁷⁾ Lignes directrices 03/2020 sur le traitement de données concernant la santé à des fins de recherche scientifique dans le contexte de la pandémie de Covid-19. Adoptées le 21 avril 2020.

Mise en place d'un réseau de proximité constitué de référents spécialisés

Pour accompagner les scientifiques au niveau local, il est utile de mettre en place dans les organismes – au sein des centres qui traitent à grande échelle des données sensibles, des plateformes de services, des services support, des directions métier, etc. – des référents spécialisés dans la protection des données personnelles qui constituent autant de relais de proximité et qu'il appartient à la DPO de former et d'animer. Ces référents, dont les compétences diverses contribuent à la richesse du réseau, seront autant de points d'appui au niveau local pour diffuser les bonnes pratiques et faire remonter les questions qui permettront de construire des outils basés sur des retours d'expérience et qui seront adaptés aux réalités du terrain.

Face aux contraintes imposées par la réglementation et aux enjeux de la mise en place de politiques institutionnelles de protection des données, un besoin de renforcement des délégations à la protection des données, qui interviennent aux côtés des chercheurs, et de leur gouvernance se fait sentir, tout comme le besoin de la création d'instances locales d'appui à la recherche. Ce nécessaire renforcement s'impose aussi au niveau des services de la CNIL. Tous ensemble, nous pouvons fluidifier l'activité de recherche au bénéfice de la santé des populations, tout en assurant la promotion d'une recherche en santé humaine responsable.