Entre gouvernance des données et intelligence artificielle : quelle place pour la poursuite de l'intérêt général?

Actes du colloque de clôture des travaux de la Chaire Justice sociale et IA

Sous la direction scientifique d'Anne-Sophie Hulin

Fondation Abeona
Championing Responsible AI

ENS | PSL

obvia

Université Laval

Auteurs

Anita Burgun : MD, PhD. Professeure d'informatique médicale à l'Université de Paris. Chef du département d'informatique médicale des Hôpitaux Georges Pompidou et Necker (AP-HP).

Stéphanie Combes : Directrice du Health Data Hub.

Nathalie de Grove-Valdeyron : Professeure de droit public, Université Toulouse Capitole (IRDEIC), Chaire Jean Monnet (2023-2026).

Caroline Guillot : Directrice citoyenne du Health Data Hub.

Anne-Sophie Hulin : Titulaire de la Chaire Justice sociale et IA, professeure à la Faculté de droit de l'Université de Sherbrooke, coresponsable de l'axe Droit, cyberjustice et cybersécurité de l'Obvia.

Jacques Priol : Président du cabinet CIVITEO et de l'Observatoire Data Publica, Expert associé auprès du Partenariat mondial sur l'intelligence artificielle.

Jeanne Solofrizzo : Consultante du cabinet CIVITEO et étudiante du master de droit public du numérique de l'Université Paris I Panthéon-Sorbonne.

Grimaud Valat : Avocat associé DTMV Avocats et Senior advisor Human Technology Foundation.

Cette publication a été réalisée dans le cadre des travaux de la Chaire Justice sociale et intelligence artificielle, soutenus financièrement par la Fondation Abeona, l'Obvia, l'Université Laval et l'École normale supérieure de Paris (ENS-PSL). Ces travaux ont également bénéficié du soutien financier du Fonds de recherche du Québec.

Produit avec le soutien financier du Fonds de recherche du Québec

ISBN : 978-2-925138-53-2
DOI : 10.61737/UIWJ9558

Pour citer ce document : Hulin, AS (dir), (2024). *Entre gouvernance des données et intelligence artificielle: quelle place pour la poursuite de l'intérêt général?* Obvia.
https://doi.org/10.61737/UIWJ9558

Table des matières

Avant-propos 4

L'inaccessibilité de l'abondance : le supplice de Tantale 8

IA de confiance et action publique : choisir d'abord des IA utiles 12

Espace européen des données de santé : enjeux et défis pour l'utilisation secondaire des données de santé 18

Données et IA en santé : quels sont les nouveaux défis? 24

Citoyens et seconde vie des données de santé : les enjeux en matière d'information et de droits 30

Avant-propos

Anne-Sophie Hulin

Titulaire de la Chaire Justice sociale et IA, professeure à la Faculté de droit de l'Université de Sherbrooke, coresponsable de l'axe Droit, cyberjustice et cybersécurité de l'Obvia

Effervescence, stupeur et tremblements, passion et (dé) raison…

Voici quelques mots qui, à mon sens, capturent les sentiments tumultueux que génèrent les avancées récentes de l'intelligence artificielle et les enjeux de société qu'elles soulèvent. Aux côtés d'un vif enthousiasme pour le potentiel de l'IA à transformer nos vies, nos industries, sa réception fait autant preuve de sidération devant la rapidité des progrès et d'inquiétudes relatives au fait que toute innovation a un prix social. La croissance explosive de l'IA et de l'utilisation des données massives a en effet exacerbé et révélé des enjeux sociétaux complexes, incluant notamment des enjeux de vie privée, de sécurité des données, de transformation de l'industrie et des activités créatives, de respect de l'autonomie individuelle ainsi que des implications sur l'emploi et les inégalités sociales, etc. Ces questions soulèvent des dilemmes éthiques, juridiques et techniques sur la manière dont la technologie doit être développée, déployée et encadrée de sorte à garantir un niveau de compatibilité satisfaisant avec le bien-être collectif et la protection des droits et des libertés individuelles. La recherche d'un équilibre entre innovation technologique et respect des principes éthiques et juridiques constitue un défi majeur pour les sociétés modernes.

Une postmodernité de la gouvernance des données et de l'IA?

De ce point de vue, les dernières années constituent un moment charnière en matière de gouvernance des données et de l'IA face à la prise de conscience collective accrue des implications éthiques, sociales et économiques de ces technologies. La nécessité d'encadrer l'innovation technologique est ainsi devenue plus évidente que jamais, notamment avec le déploiement massif de l'IA générative auprès du grand public et la difficile maitrise des IA à usage général. Ce faisant, les réflexions et initiatives en termes d'encadrement ont foisonné partout dans le monde et à différentes échelles (nationales, régionales et/ou internationales). Parmi celles-ci, il faut noter l'offensivité juridique de l'Union européenne qui, au gré de l'édification d'un riche, novateur et strict arsenal juridique, s'est doté du premier cadre règlementaire relatif à la gouvernance des données et à l'intelligence artificielle. En effet, aux côtés du règlement européen sur la gouvernance des données[1] et sur les données[2], l'Union européenne est en voie d'adopter la première règlementation déterminant à quelles conditions les systèmes d'intelligence artificielle peuvent être mis sur le marché. Sous réserve des dernières étapes du processus législatif européen, l'entrée en vigueur est prévue pour 2026[3]. Le Canada a également opté pour l'approche législative, emboîtant le pas de l'Union européenne.

1 Règlement (UE) 2022/868 du Parlement européen et du Conseil du 30 mai 2022 portant sur la gouvernance européenne des données et modifiant le règlement (UE) 2018/1724.

2 Règlement (UE) 2023/2854 du Parlement européen et du Conseil du 13 décembre 2023 concernant des règles harmonisées portant sur l'équité de l'accès aux données et de l'utilisation des données et modifiant le règlement (UE) 2017/2394 et la directive (UE) 2020/1828 (règlement sur les données).

3 « Intelligence artificielle cx: les députés adoptent une législation historique », Parlement européen (13 mars 2024), en ligne : <https://www.europarl.europa.eu/news/fr/press-room/20240308IPR19015/intelligence-artificielle-les-deputes-adoptent-une-legislation-historique>.

Décriée sur le fond et sur la forme, la Loi sur l'intelligence artificielle et les données (LIAD) demeure encore en cours d'élaboration (PL c-27[4]). En cas d'adoption, la LIAD serait la première loi au Canada à poser des exigences communes pour la conception, le développement et le déploiement de systèmes d'intelligence. Comparativement au règlement européen, la LIAD jouit d'un champ d'application plus restreint et dont il n'est pas évident de prédire quel sera l'impact réel dans l'attente de la publication des règlements qui définiront l'essentiel de ses modalités d'application. D'autres juridictions ont également adopté des textes juridiques, mais non avec la même ampleur (ex. : Chine, États-Unis). D'autres au contraire s'y refusent pour préserver l'innovation sans pour autant ignorer les dérives potentielles de l'IA (ex. : Royaume-Uni[5]). L'agilité normative est ainsi mise de l'avant. Alors que les modèles de gouvernance de l'IA et des données se multiplient, une tendance, de plus en plus généralisée, s'installe vers la recherche d'un cadre de gouvernance des données et de l'IA qui, outre le degré de (ré) conciliation entre innovation et protection des droits et libertés individuelles, garantisse un déploiement et une utilisation de l'intelligence artificielle bénéfique pour la société dans son ensemble.

L'intérêt général comme levier de régulation de l'utilisation des données et du développement de l'IA

Cet ouvrage s'inscrit dans le contexte actuel où la définition des orientations et des objectifs relatifs au cadre de gouvernance des données et de l'intelligence artificielle revêt une importance capitale. Les contributions rassemblées illustrent une partie des échanges ayant eu lieu lors du colloque organisé sous l'égide de la Chaire justice sociale et intelligence artificielle, qui s'est tenu le 8 février 2024 à l'École Normale Supérieure de Paris. Réunissant des experts de divers domaines, ce colloque a pour objectif d'explorer quelques-uns des points d'intersection entre les données, l'IA et l'intérêt général. Ceci conduit aux questions suivantes : dans quelle mesure la poursuite de l'intérêt général peut-elle légitimer l'innovation numérique? Comment peut-elle influer sur son encadrement? Quelles perspectives sociales et/ou voies de tempéraments aux pratiques actuelles offre-t-elle du point de vue de l'utilisation des données et du développement de l'IA? Autrement dit : entre gouvernance des données et intelligence artificielle : quelle place pour la poursuite de l'intérêt général?

Au cours de cet après-midi d'échanges, les débats se sont articulés autour de trois thématiques principales, lesquelles matérialisent des préoccupations d'intérêt général. En effet, bien que la détermination de l'intérêt général soit empreinte de complexité, particulièrement dans un contexte de droit comparé, il demeure possible d'en identifier des manifestations concrètes.

Ce faisant, le premier panel s'est concentré sur les possibilités et les défis de l'altruisme des données. Ce concept se réfère à l'idée de partager volontairement des données personnelles ou d'entreprise à des fins bénéfiques pour la société, sans recherche de profit personnel. Ce partage peut être motivé par des raisons éthiques, humanitaires, ou de recherche scientifique, où les données peuvent être utilisées pour améliorer des services, développer de nouvelles technologies, ou encore aider dans des études de santé publique, de gestion de crises, ou de développement durable. Depuis son introduction en droit de l'Union européenne avec le règlement européen sur la gouvernance des données, l'altruisme des données interpelle. Non seulement, il s'érige en mode alternatif d'accès aux données des individus, mais il constitue un des leviers de la stratégie européenne des données établie en soutien l'innovation technologique. L'altruisme des données suggère une réorientation des pratiques et des politiques autour de la collecte, de l'analyse et du partage des données, en favorisant une approche qui valorise la contribution des données à l'intérêt général, sans pour autant négliger la protection des droits individuels.

4 Loi édictant la Loi sur la protection de la vie privée des consommateurs, la Loi sur le Tribunal de la protection des renseignements personnels et des données et la Loi sur l'intelligence artificielle et les données et apportant des modifications corrélatives et connexes à d'autres lois, projet de loi n°C-27 (deuxième lecture et renvoi en comité - 24 avril 2023), 1e sess., 44elégis. (Can.)

5 A pro-innovation approach to AI regulation, Policy paper, Presented to Parliament by the Secretary of State for Science, Innovation and Technology by Command of His Majesty on 29 March 2023, 2023, en ligne : <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>.

Dans ce contexte, l'altruisme des données devient un catalyseur pour repenser la gouvernance de l'information à l'ère du numérique et trouver des leviers pour garantir la transparence, la responsabilité et la participation citoyenne. En effet, l'altruisme des données contribue à l'encapacitation numérique des individus qui, à travers leurs dons de données, peuvent se doter d'un certain contrôle quant à l'utilisation secondaire de leurs données. Toutefois, ce mode alternatif de circulation des données pose également des questions importantes sur la vie privée, la sécurité des données. Il nécessite plus que jamais des cadres règlementaires clairs pour protéger les droits des individus, mais qui contribuent également à l'instauration d'un climat de confiance et d'inclusion. Sans ces éléments clés, les initiatives visant à promouvoir le partage altruiste des données ne bénéficient pas d'un contexte suffisamment incitatif pour les individus et présentent un fort risque d'échec. La faisabilité de ce partage fait aussi l'objet de doutes. Car, comment partager ses données quand on n'en a pas la possession ni le contrôle? L'altruisme des données met en exergue l'importance de doter le droit à la portabilité d'efficacité, faute de quoi il demeurera au rang d'ambition inachevée.

Le second panel a, pour sa part, mis de l'avant l'ambivalence du lien entre le développement de l'IA et la protection de l'environnement. L'IA, avec sa capacité à traiter de vastes quantités de données et à exécuter des tâches complexes rapidement, présente en effet des opportunités considérables pour l'innovation et l'amélioration des services dans de nombreux secteurs. Toutefois, son expansion rapide soulève également des questions importantes concernant la consommation énergétique et l'utilisation des ressources, ainsi que les implications éthiques et sociales. Alors que le rôle de l'IA dans la compréhension et la gestion des problématiques environnementales est mis de l'avant, l'empreinte environnementale de son développement pose la question de savoir dans quelle mesure l'IA doit et peut continuer à se développer au sein des sociétés. De manière commune, les échanges de ce panel ont mis de l'avant l'importance de conserver une approche critique, si ce n'est raisonnée, du développement et de l'utilisation de l'IA. Le recours à l'IA par les organisations est-il nécessaire en tout temps? Il semble impératif de considérer les contextes spécifiques et les besoins particuliers avant de décider de son adoption systématique. Car, aussi prometteuse que soit l'IA d'un point de vue social, économique et environnemental, il paraît crucial, qu'au regard de la grandissante gourmandise en données et en énergie des systèmes d'IA, de procéder à une évaluation holistique des solutions d'IA. Celle-ci ne se limiterait ni à leur efficacité technique ni à leur potentiel économique ou social, mais prêterait davantage attention à ses impacts sociaux, économiques et environnementaux conjointement. Une telle approche implique de considérer comment les technologies d'IA peuvent être conçues et déployées de manière à promouvoir le bien-être collectif présent et futur, tout en respectant les principes de justice sociale. Cet alignement de l'IA est d'autant plus pertinent que les technologies d'IA ne sont pas neutres : elles reflètent les valeurs des sociétés qui les développent et les utilisent. Outre des cadres de gouvernance qui intègrent ces considérations, l'élaboration de méthodologies pour accompagner les organisations désireuses d'opter pour l'intelligence artificielle paraît ainsi un exercice à ne pas sous-estimer en vue de garantir une utilisation responsable de l'IA.

Enfin, le troisième et dernier panel du colloque s'est concentré sur le domaine de la santé, lequel constitue un point de rencontre central entre la gouvernance des données, l'IA et l'intérêt général. La santé représente un terrain fertile pour l'intégration de l'IA, marquant une convergence cruciale avec la gouvernance des données et l'intérêt général. L'utilisation secondaire des données tout comme l'implantation de l'IA en santé offrent des opportunités uniques en termes d'amélioration des diagnostics et des traitements personnalisés, de gestion efficiente des ressources de santé, etc. Telle est la raison pour laquelle l'Union européenne a lancé comme chantier prioritaire la constitution de l'espace européen de données de santé, lequel entend d'ailleurs combiner initiative publique et altruisme des données.

Cependant, l'utilisation secondaire des données de santé soulève également des questions éthiques et légales importantes concernant la confidentialité, la sécurité des données, l'équité de l'accès aux technologies, la relation patient-praticien ainsi que des difficultés opérationnelles majeures (accès aux données pertinentes, sécurité, interopérabilité des données). Au gré de ces multiples enjeux s'érige alors la difficulté d'établir un équilibre socialement acceptable entre l'innovation et la protection des personnes dans le contexte où les liens entre les données, l'IA et la santé se resserrent continuellement. Cet équilibre est d'autant plus difficile à atteindre que les données de santé constituent certes des données sensibles, mais qu'elles sont également marquées d'une hybridité intrinsèque. Si elles sont le cœur de l'intimité de la personne et de sa santé, elles sont fondamentales du point de vue de la santé collective. Les attentes en termes de protection sont ainsi inversement proportionnelles aux besoins de leur utilisation secondaire pour des fins de recherche. Il est donc crucial de trouver un juste milieu entre confidentialité individuelle et bénéfice collectif, afin de permettre des progrès significatifs tout en respectant les droits des individus.

De manière commune, les échanges ont mis en lumière l'importance d'une gouvernance des données éthique, transparente, inclusive, durable, ainsi que le potentiel transformateur de l'IA pour répondre aux enjeux sociétaux majeurs que sont, entre autres, l'amélioration de la santé et la protection de l'environnement. Somme toute, l'ensemble des contributions soulève avec finesse les différentes tensions qui entourent le rapport que nous entretenons avec nos données, souligne l'importance de la responsabilité collective dans la gouvernance des données et le développement de l'IA tout en illustrant comment le numérique renouvèle l'implication citoyenne.

La Chaire justice sociale et intelligence artificielle

Pour conclure cet avant-propos, je tiens à exprimer ma profonde gratitude envers chaque intervenante et intervenant. à ce colloque, et particulièrement celles et ceux qui ont accepté d'en publier le contenu. Leur générosité à partager perspectives et recherches a profondément enrichi notre compréhension des défis éthiques, sociaux et environnementaux associés à l'intelligence artificielle sous l'angle de la justice sociale. Ces échanges ont non seulement posé les fondations de futures recherches, mais ont également alimenté un dialogue essentiel sur les modalités par lesquelles l'intelligence artificielle peut être orientée vers la promotion de la justice sociale. Il ne fait aucun doute que le contenu de cette journée constitue une ressource précieuse pour quiconque s'intéresse aux croisements entre gouvernance des données, intelligence artificielle et bien commun. Je me réjouis donc qu'en plus des enregistrements des conférences librement accessibles en ligne, une publication en libre accès, tant en français qu'en anglais, permette de partager largement les précieuses contributions de ce colloque.

Mes derniers mots sont destinés à souligner l'opportunité exceptionnelle que fut la Chaire Justice sociale et IA dans mon parcours de chercheuse. Chaque année, la Chaire Justice sociale et intelligence artificielle accueille une chercheuse ou un chercheur pour explorer les intersections entre l'IA et la justice sociale. Depuis sa fondation, la Chaire a eu l'honneur d'accueillir des chercheuses telles que Kate Crawford et Karine Gentelet, et j'ai eu l'extrême privilège d'en être la troisième titulaire pour l'année 2022-2023. Je suis extrêmement reconnaissante envers les partenaires de la Chaire — la Fondation Abeona, l'ENS PSL Paris, l'Obvia et l'Université Laval — pour leur soutien et confiance indéfectibles. Leur engagement dans la recherche pour une IA respectueuse des droits et intérêts tant des individus que de la collectivité a été une source d'inspiration continue et a contribué à enrichir ma réflexion et mes recherches. À cet égard, je souhaite exprimer ma gratitude à Anne Bouverot, Tanya Perelmuter, Thierry Poibeau et Lyse Langlois pour leurs conseils avisés vis-à-vis de la réalisation de mes travaux.

Finalement, je souhaite remercier toute l'équipe opérationnelle de l'Obvia, et tout particulièrement Guillaume Macaux pour m'avoir épaulé dans la production et la diffusion à grande échelle de mes travaux.

Je souhaite également remercier l'équipe de chercheuses de la Chaire (Alexia Argiolas, Émilie Bélanger, Émilie Guiraud et Samantha Dorinet) et salue leur réactivité en recherche compte tenu de la frénétique actualité juridique des derniers mois en l'espèce.

L'organisation du colloque du 8 février 2024 et la réalisation de cet ouvrage ont également bénéficié de la précieuse assistance de Samantha Dorinet. L'édition de l'ouvrage a également bénéficié de l'appui de Daniel Polillo pour la révision linguistique en anglais.

Consulter la page de l'événement pour plus d'information :

L'inaccessibilité de l'abondance : le supplice de Tantale

Grimaud Valat

Avocat Associé DTMV Avocats
Senior Advisor Human Technology Foundation

Où l'abondance des données récoltées et exploitées à des fins commerciales doit être mise au service de l'intérêt général par l'entremise d'une démarche volontaire de l'individu.

Quel supplice que d'être entouré, presque submergé, des ressources dont on a vitalement besoin sans pouvoir y accéder...

Tel fut le sort réservé à Tantale, fils de Zeus puni par les Dieux de l'Olympe à passer l'éternité dans le Tartare, debout au milieu d'un cours d'eau sans pouvoir s'y abreuver, et entouré d'arbres fruitiers dont les branches se dérobaient à lui dès qu'il en approchait la main.

Tantale, le supplice de l'inaccessibilité de l'abondance.

Ce supplice n'est-il pas celui subi aujourd'hui par tout porteur de projet essentiellement tourné vers l'intérêt général, et qui aurait besoin de données pour remplir les objectifs qu'il entend atteindre?

Quel paradoxe que de constater la presque infinie abondance des données collectées sans difficulté à des fins commerciales et l'inaccessibilité de ces mêmes données à des fins altruistes! Illustrons cela en prenant l'exemple, concret et simple, des données de circulation des individus au sein des villes. Celles-ci sont collectées, traitées, et utilisées par les éditeurs d'application de navigation pour nourrir leurs systèmes d'intelligence artificielle et tenter de proposer à leurs utilisateurs des itinéraires optimisés et moins encombrés. Ces données existent, sont d'une redoutable précision, mais ne sont pas accessibles aux institutions publiques en charge de la conception de plans de circulation adaptés aux nouvelles mobilités...

L'absence de partage de ces données conduit à promouvoir des services de contournement des embouteillages, là on l'on pourrait éviter leur formation. Le partage de ces données pourrait améliorer en même temps la qualité de vie des résidents et passants des villes, tout en contribuant à la réduction des émissions de CO2.

Si l'on entend dépasser cette situation paradoxale, il faut s'attacher à en comprendre l'origine : à savoir la défiance majeure des individus comme des entreprises vis-à-vis du partage de leurs données.

L'instauration de cette défiance était inévitable si l'on considère la façon dont les populations ont été initialement confrontées à l'usage massif de leurs données. Rappelons que notre premier contact avec l'usage massif de nos données personnelles fut pris avec les réseaux sociaux, fournisseurs privés de services de divertissement « gratuits ». Et les individus d'y déverser l'intégralité de ce qui constituait leur vie privée, dont des informations particulièrement sensibles comme leurs opinions politiques ou leur orientation sexuelle, afin de partager celle-ci avec leurs « amis » digitaux, protégés toutefois du regard de l'autre par un écran (de fumée?). La gratuité n'existant pas, le sous-jacent du modèle était évidemment une concession d'un droit au bénéfice des éditeurs de ces services à exploiter les données déversées dans ceux-ci pour créer et alimenter le système de publicité ciblée le plus perfectionné jamais créé.

Les données personnelles devenaient « l'or noir » du XXIe siècle, leur exploitation propulsant en un temps record (quelques années tout au plus) des start-ups au rang de sociétés disposant de chiffre d'affaires et de pouvoir d'influence supérieurs à des états.

Cette expression, « or noir » traduit au demeurant parfaitement la vision purement mercantile que l'on a des données disponibles et porte avec elle tout ce qu'il y a de négatif et d'inéquitable dans l'exploitation de ses gisements. Pourtant les données ne sont pas de l'or noir, leur exploitation ne peut pas être limitée par cette vision purement économique, là où leur collecte et leur partage peuvent être mis au service de la résolution de certains des plus grands défis auxquels notre civilisation se heurte.

Le partage de la donnée doit être mis au service de l'amélioration de la santé, des mobilités, et évidemment de la lutte contre le réchauffement climatique.

Mais cela n'est possible qu'en outrepassant la défiance. L'effort pédagogique est ici déterminant et doit servir la démonstration d'un concept qui est fort éloigné de ce qui constitue le socle de notre raisonnement habituel : en matière de donnée, la protection permet le partage. Contrairement aux biens dits rivaux, qui se détériorent voire disparaissent avec l'usage que l'on en fait, et dont la protection fait obstacle au partage, les données sont des biens non rivaux et peuvent être exploitées et partagées, sans disparaitre ni se détériorer.

La protection de la donnée ne passe ainsi pas par l'instauration d'une exclusivité d'usage, mais par la gouvernance de celui-ci.

Ceci valant tant pour les individus que pour les entreprises. Les premiers entendant préserver leur vie privée (dans le contexte tout aussi paradoxal d'utilisation massive des réseaux sociaux) et s'affranchir de toute « surveillance » d'état. Les secondes cherchant à préserver leur avantage concurrentiel.

Il faut donc apporter une solution permettant d'assurer la protection des données, et ce à travers des mesures techniques, certes, mais surtout une gouvernance préservant les intérêts de toutes les parties prenantes.

L'altruisme en matière de données (ou data altruisme) propose justement une solution de gouvernance permettant d'assurer une protection des données telle que leur partage au service de l'intérêt général devient finalement possible.

Le data altruisme c'est la proposition faite par la Commission européenne, dans le Data Governance Act, de créer une troisième voie de partage des données, entre la voie commerciale, subie par les individus, et la voie réglementaire, instaurant un partage obligatoire.

Le principe repose sur la création d'un tiers de confiance : l'organisation data altruiste. Cette organisation, pour bénéficier de ce titre, doit répondre à plusieurs critères, et notamment être à but non lucratif, être indépendante légalement et fonctionnellement de toute autre organisation ou activité et naturellement être créée pour servir une finalité d'intérêt général qu'elle doit déclarer. Le rôle de cette organisation est ensuite de collecter les données auprès de contributeurs, puis d'organiser la gouvernance de ces données pour qu'elles profitent à la cause d'intérêt général qu'elle promeut, tout en veillant au respect des droits des contributeurs.

Ces conditions viennent ainsi contrer la majorité des causes générant de la défiance : le collecteur de données ne recherche pas le profit, il est indépendant de toute société ou état, et il gouverne les données qui lui sont confiées pour les mettre au service de l'intérêt général, les protégeant pour mieux les partager.

Les conditions de la confiance sont ainsi créées, et soutenues par des obligations de transparence permettant aux contributeurs de savoir comment et par qui sont exploitées leurs données.

L'autre particularité du data altruisme est qu'il est basé sur une démarche volontaire des contributeurs, une démarche de partage active qui n'est ni subie, ni contrainte, mais mue par un choix et une envie de contribuer à une cause.

La Commission européenne crée ainsi, par la voie réglementaire, un cadre destiné à favoriser l'émergence d'initiatives de partage de données au service de l'intérêt général, dans un contexte favorisant l'émergence d'une confiance.

Si la confiance peut être reconquise, il convient de souligner qu'il existe, outre la défiance, un deuxième obstacle de taille au partage des données. Celui de la maitrise de ces données par leurs sujets.

Le constat est assez évident : les individus n'ont en réalité qu'un contrôle extrêmement limité sur leurs données, voire aucun contrôle du tout. Les données sont collectées par les fournisseurs de services numériques, mais ne sont pas accessibles directement par les individus.

Or le partage implique une maitrise.

La mise en place d'une identité numérique de confiance, centrée sur l'individu est probablement l'un des outils permettant le mieux de répondre à ce problème.

Outre les services d'identification ou d'authentification des individus permis par les systèmes d'identité numérique, ceux-ci offrent surtout la possibilité de réunir en un même endroit sécurisé, ce que l'on appelle un « portefeuille numérique » de données et attributs formant ce que l'on appelle l'identité numérique des individus, dans l'acceptation la plus large de ce terme.

La création d'un tel portefeuille numérique, tel que celui actuellement promu par l'Union européenne, permettrait de donner aux citoyens une voie d'accès et de contrôle unique à leurs données tout en bénéficiant d'un haut niveau de sécurité (substantiel à tout le moins) et d'une authentification des données qu'il contiendrait.

L'un des leviers de la confiance étant le caractère volontaire de l'utilisation des technologies, il importe que les systèmes d'identité numériques soient pensés pour les individus, et que ceux-ci puissent décider :

- d'abord de ce qu'ils entendent stocker dans leur portefeuille d'identité numérique
- et ensuite de ce qu'ils entendent partager, auprès de qui et pour quel usage.

Si l'identification officielle de l'individu est permise par les fonctions primaires du système d'identité numérique (compris comme une extension digitale des cartes d'identité et autres passeports), cela permet d'authentifier le lien entre les données contenues dans le portefeuille numérique et son détenteur. Partant de ce principe, certaines de ces données, et notamment certains attributs, peuvent être partagées sans que l'identité complète du divulguant ne doive être révélée, dans le strict respect des principes de minimisation des données et de protection de la vie privée. Cela permet d'augmenter considérablement les possibilités offertes par un tel outil, tout en assurant une protection maximale de la personne. À titre d'illustration, la preuve d'âge, qu'elle soit requise en ligne ou dans le monde réel pour accéder à certains lieux ou services, devient possible sans divulgation d'une autre information que le respect de la condition d'âge. La personne étant identifiée, ses données d'âge certifiées, il est possible de créer un certificat d'âge anonymisé.

Un tel système permet également aux individus qui contrôleraient les données contenues dans leur portefeuille numérique de les partager, par exemple avec des organisations altruistes en matière de données, ayant gagné leur confiance grâce aux règles de gouvernance mises en place et grâce aux rapports de transparence mis à disposition des contributeurs.

L'organisation altruiste sait, quant à elle, que les données transférées depuis un portefeuille d'identité numérique proviennent d'un individu dont l'identité est vérifiée, et que ces données sont certifiées. Elle peut donc les collecter et les mettre à la disposition d'un objectif d'intérêt général sans avoir à connaître l'identité du contributeur et en disposant d'un haut degré de confiance vis-à-vis de ces données.

Ceci démontre que lorsque les données sont protégées et gouvernées de bout en bout, elles peuvent être partagées.

Les données protégées peuvent être partagées, et les données partagées peuvent rester protégées.

C'est ainsi seulement, en appréhendant la question du partage de la donnée au service de l'intérêt général comme un système global impliquant la création d'un cadre de confiance de bout en bout, que l'on pourra mettre un terme à notre supplice Tantalien, et que l'on verra enfin l'accès à l'abondance libéré au service du bien commun.

Bibliographie indicative:

Rapports de la Human Technologie Foundation :

- Sur l'altruisme en matière de données, en ligne : https://www.human-technology-foundation.org/fr-news/rapport-data-altruisme.
- Sur l'Identité numérique, en ligne : https://www.human-technology-foundation.org/fr-news/rapport-sur-lidentite-numerique.

IA de confiance et action publique : choisir d'abord des IA utiles

Jacques Priol

Président du cabinet CIVITEO et de l'Observatoire Data Publica, expert associé auprès du Partenariat mondial sur l'intelligence artificielle

Jeanne Solofrizzo

Consultante du cabinet CIVITEO et étudiante du master de droit public du numérique de l'Université Paris I Panthéon-Sorbonne.

L'année 2023 restera dans l'histoire du progrès technologique comme celle de l'irruption des intelligences artificielles génératives au cœur des débats scientifiques, médiatiques, mais aussi politiques et sociétaux. En quelques mois, en France comme ailleurs, tout le monde ou presque a entendu parler de ChatGPT puis des autres larges modèles de langage. Dès le mois de mai, l'époque est encore à la sidération, 71 % des Français en ont entendu parler; mais 72 % s'estiment insuffisamment formés pour l'utiliser. À noter : 68 % de ceux qui l'utilisent dans l'entreprise ne le disent pas à leur supérieur hiérarchique[1].

Arrivent les premières polémiques. Elles portent d'abord en Europe sur la compatibilité de ces outils avec le Règlement général sur la protection des données (RGPD). Les données d'entraînement comportaient-elles des données personnelles collectées en dehors de tout cadre de consentement licite? Plus inquiétant : comment l'utilisation de ces IA génératives peut-elle garantir la licéité du traitement des données de ses propres utilisateurs? Ces questions, posées par les autorités de régulation de différents pays européens, provoqueront quelques ajustements et quelques retards dans le déploiement des LLM d'OpenAi ou de Google en Europe. Une grande ville française (Montpellier) interdit même temporairement l'utilisation de ChatGPT à tous ses agents publics.

D'autres débats surgissent, à l'échelle mondiale. Les bases d'apprentissages comportent quantité de données protégées, de sorte que le recours aux IA génératives expose leurs utilisateurs à des procès pour violation de droits d'auteur ou du secret des affaires. Le coût énergétique et l'impact carbone de chaque modèle, mais aussi de chaque requête, sont progressivement dévoilés. Plus important encore : après la sidération chacun prend conscience que ces IA ne sont pas infaillibles. Pire : elles souffrent d'hallucinations. Ce point est essentiel, car pour le comprendre il faut entrer dans la boîte noire. Il faut expliquer les principes de l'apprentissage machine et comprendre que ces IA génératives calculent des probabilités qu'un texte (une image, un son, une vidéo…) soit vraisemblable. Et grâce à l'accent mis sur ces outils violemment disruptifs, il devient possible d'expliquer d'autres systèmes d'intelligence artificielle, moins spectaculaires et qui avaient commencé à se déployer dans des pans entiers de la société à bas bruit.

Ainsi, le service public local en France n'avait pas attendu ChatGPT pour s'intéresser à l'IA. Mais personne, ou presque, n'en parlait. Quelques territoires pionniers, souvent de grandes métropoles et des régions, avaient testé des outils d'apprentissage algorithmique pour modéliser et prédire des situations et optimiser la gestion publique. Citons par exemple la Région Occitanie qui utilise de l'IA depuis 2020 pour analyser les besoins de formation des demandeurs d'emploi et proposer rapidement des stages et des parcours de formation adaptés aux besoins immédiats des entreprises. Citons aussi Nantes ou Toulouse qui travaillent sur la prédiction de la fréquentation des cantines scolaires pour lutter contre le gaspillage alimentaire.

1 Etude IPSOS – mai 2023

En 2022 et en 2023, l'Observatoire Data Publica[2] a réalisé une grande enquête auprès des collectivités locales françaises sur l'usage des données, l'open data, la cybersécurité et bien sûr le recours à l'intelligence artificielle. C'est une partie de ces résultats qui est présentée ici.

Un focus suit sur les conditions de déploiement des IA par les collectivités territoriales françaises. L'outil d'analyse qui est présenté a été développé par les consultants du cabinet CIVITEO lors d'ateliers de travail avec la Région Centre Val-de-Loire, l'agglomération de Paris-Saclay, les villes d'Angers, Montpellier, Montreuil et Clichy-sous-Bois.

I. État des lieux du recours à l'IA par les collectivités locales en France

L'enquête de l'Observatoire Data Publica a été administrée durant l'été 2023 auprès de deux cents collectivités locales françaises (régions, départements, communes ou établissements publics de coopération intercommunale) réparties sur la totalité du territoire métropolitain[3].

27 % des collectivités répondantes déclarent avoir déjà utilisé (ou testé) une IA, contre 15 % en 2022. Sans surprise, il s'agit principalement de grandes ou très grandes collectivités. Les métropoles[4] sont les plus nombreuses à avoir testé des IA (44 %). Tout comme 36 % des départements et 30 % des régions. Aucune commune de moins de 10 000 habitants ne déclare en revanche avoir testé l'IA. Au total, il y a, mi-2023, plus d'une cinquantaine de projets d'IA territoriale en cours de déploiement, contre 19 en 2022[5] et seulement 5 en 2021[6]. Les thématiques prioritaires sont les suivantes : l'analyse des flux de mobilité, des robots conversationnels à l'attention des usagers, des outils de modélisation pour optimiser la gestion des déchets, les cantines ou encore l'éclairage public. Plusieurs projets concernent la sécurité avec notamment le recours à la vision par ordinateur, objet de polémiques en raison de l'adoption en France d'une loi dérogatoire pour préparer l'accueil des Jeux olympiques de Paris.

Mais nous allons voir que cette photographie à l'été 2023 va rapidement être périmée. IA générative oblige!

Dans l'enquête de l'Observatoire, 30 % des collectivités territoriales déclarent vouloir tester des IA en 2024. Les trois thématiques prioritaires concernent le changement climatique : il s'agit d'utiliser l'IA pour piloter des politiques environnementales, pour la gestion de l'énergie et la gestion de l'eau. Courant 2023, le Gouvernement français a lancé un appel à projets permettant de financer des IA territoriales au service de la transition climatique. Chaque projet retenu est susceptible de bénéficier de plusieurs millions d'euros de subvention dès lors qu'il s'agira d'un démonstrateur reproductible. Les projets sont nombreux. Ils concernent la gestion des nappes phréatiques, la réduction des consommations énergétiques des services publics, l'impact des règles d'urbanisme sur l'artificialisation des sols ou la biodiversité, la réduction de la mobilité automobile, l'amélioration du tri et du recyclage des déchets… En parallèle, se poursuit à grande échelle la diffusion de systèmes d'IA permettant l'analyse d'images adossées aux dispositifs de vidéoprotection, à des fins de sécurité, mais aussi de fluidification intelligence du trafic (priorité aux vélos ou aux transports en commun par exemple) ou de détection des dépôts de déchets sauvages.

2 L'Observatoire Data Publica est une association à but non lucratif qui a été créée en 2020 par les cabinets CIVITEO, DATACTIVIST, INNOPUBLICA et PARME AVOCATS. L'Observatoire est l'un des partenaires en France de l'OBVIA.

3 L'enquête complète et l'ensemble des données sont disponibles en ligne : < https://enquete.data-publica.eu/rapport/Rapport.html >.

4 En France, les métropoles sont des collectivités territoriales avec un statut particulier pour la gestion des 21 aires urbaines les plus importantes.

5 Conseil d'État (2022). *Intelligence artificielle et action publique : construire la confiance, servir la performance*, disponible en ligne : < https://www.conseil-etat.fr/publications-colloques/etudes/intelligence-artificielle-et-action-publique-construire-la-confiance-servir-la-performance#:~:text=Intelligence%20artificielle%20et%20action%20publique%20%3A%20construire%20la%20confiance%2C%20servir%20la%20performance,-31%20août%202022&text=Si%20l%27intelligence%20artificielle%20suscite,impossibles%20à%20réaliser%20jusque%2Dlà >.

6 Commission européenne (2021). *Artificial intelligence in public services.*

Mais les projets les plus nombreux, les plus rapides et perçus comme les plus simples à déployer, prennent appui sur l'IA générative. Ils voient le jour à l'initiative des services publics. Mais souvent il s'agit d'une réponse à des sollicitations de grandes entreprises numériques comme de start-up locales. Les projets de LLM territoriaux se multiplient. Ainsi, DelibIA est un projet de génération de textes juridiques locaux, dont l'apprentissage repose sur l'ingurgitation de millions de délibérations municipales existantes. De nombreux projets de robots conversationnels sont évoqués, pour orienter les usagers dans le maquis des aides publiques. Le Gouvernement français expérimente lui-même ces outils, et développe son propre modèle de langage. Ce « ChatGPT administratif souverain » et open source est prénommé ALBERT. Il est mis à disposition de toutes les administrations.

Finalement, CIVITEO et l'Observatoire Data Publica ont identifié 7 grands domaines d'application possible de l'IA au sein des administrations publiques locales.

Shéma 1 : Les domaines possibles de recours à l'IA pour les missions de service public local.

Mais l'utilisation de ces outils ne va pas de soi. Les enjeux économiques, juridiques et éthiques, environnementaux, technologiques et même démocratiques et politiques sont nombreux. Ils existent pour le monde de l'entreprise, mais prennent une autre ampleur quand il s'agit d'utiliser l'IA au service de l'intérêt général. D'où les questions que se posent les décideurs publics.

II. Pour des IA utiles : la construction d'un outil d'aide à la décision pour les décideurs publics locaux

L'outil présenté ici est le fruit d'un travail concret mené avec des décideurs locaux de différentes collectivités par les équipes de CIVITEO. Nous avons ainsi pu travailler à l'élaboration de plusieurs projets d'utilisation de l'IA à une échelle territoriale, notamment le projet PrevizO porté par le Climate Data Hub de la Région Centre Val-de-Loire (outil de modélisation de la pénurie en eau sur un bassin versant de la Loire) ou le projet Urba(IA) de l'agglomération de Paris-Saclay (combinaison d'algorithmes de modélisation des impacts des règles d'urbanisme sur le climat, la biodiversité, l'artificialisation des sols…).

Nous avons aussi pu tester et coconstruire cet outil d'aide à la décision en mobilisant les équipes municipales de plusieurs communes. La Ville d'Angers, première métropole française à avoir voté une « stratégie métropolitaine de la donnée », a souhaité se doter d'une doctrine interne sur le recours à l'IA dans ses missions de service public. Les villes de Montreuil et Clichy-sous-Bois ont fait travailler leurs cadres au cours de séminaires sur l'IA. La Ville de Montpellier a constitué un panel test auprès d'agents volontaires représentant une grande variété de métiers dans le cadre d'une initiative plus globale de concertation sur l'intelligence artificielle au service de l'intérêt général.

A) La nécessité d'accompagner la réflexion

Chaque système d'intelligence artificielle susceptible d'être utilisé par le service public local est caractérisé par une multitude de critères susceptibles d'entrer en compte dans la décision de l'utiliser, ou non.

L'algorithme est-il ouvert ou non? A-t-il été développé par une entreprise française ou européenne? Respecte-t-il (par anticipation) les critères de la future réglementation européenne IA Act? Que sait-on des données d'entraînement? S'il s'agit d'une IA spécifique, a-t-elle été entraînée avec les données de villes françaises? Ou américaines, ou chinoises? Est-ce que l'utilisation de l'algorithme garantit le respect du RGPD ou au contraire fragilise la conformité du service rendu? Comment est contrôlée la performance de l'algorithme? Comment sont détectés les biais? Qui a la capacité de les corriger, de parfaire l'entraînement? Quel est l'impact environnemental de ce système d'IA? Sait-on le mesurer?

Sur tous ces points, il existe une littérature abondante, des chartes, des guides. En France, le Conseil d'État a publié un rapport[7] dans lequel sont listées les conditions d'une IA de confiance au service de l'action publique : primauté humaine, autonomie stratégique, soutenabilité environnementale, sécurité, transparence, équité et non-discrimination, performance.

Mais ce cadre théorique est complexe à décliner pour des administrations locales confrontées à des opportunités d'IA dont les promesses sont parfois très alléchantes, et la capacité à répondre ou même à identifier les enjeux face à des risques réels d'IA inappropriées.

C'est la raison pour laquelle la grille d'analyse construite avec quelques collectivités pionnières vise à mettre en balance des éléments appréhendables localement par des équipes formées a minima sur les concepts clefs de l'intelligence artificielle et sur le cadre juridique de leur déploiement (il existe par exemple en France une loi sur la transparence algorithmique, peu connue et peu appliquée, particulièrement stricte en cas de recours à l'IA). L'objectif n'est pas d'être « pour » ou « contre » le recours à l'IA pour un projet d'utilité sociale, mais de savoir dire que l'on peut le déployer « à condition que », ou qu'il ne faut pas le déployer « parce que ».

Et pour chaque projet, la première question est celle de l'utilité de l'IA.

7 Conseil d'État (2022), *supra*.

B) Un outil au service des collectivités

L'examen d'un projet de recours à l'IA débute par une estimation des bénéfices attendus. Ils sont de trois ordres. D'abord fonctionnels : l'IA peut améliorer des prévisions, réduire des délais de traitement, supprimer des tâches répétitives, permettre des économies substantielles. Les bénéfices peuvent aussi être en lien avec la stratégie du territoire, son attractivité, son écosystème. Ils peuvent bien sûr être politiques au sens où le recours à l'IA vient en appui d'une politique prioritaire portée par les élus.

Il y a aussi des risques. Et donc des conditions à réunir pour déployer le système en confiance. Dans la grille construite avec les collectivités, nous en avons retenu trois. Le premier risque porte sur la maîtrise et le contrôle des biais et erreurs (la supervision humaine étant l'une des conditions possibles), le second la maîtrise de l'algorithme et des données (permettant l'autonomie stratégique et condition de la transparence) et le troisième concerne la protection de la vie privée (avec la possibilité d'études d'impact approfondie pour garantir la conformité RGPD).

À ces six critères en opportunités et risques, s'ajoutent deux critères dont l'impact se mesure sous la forme d'une balance : l'impact environnemental et l'impact sur les métiers. Concernant l'impact environnemental, l'émergence d'outils performants mesurant l'impact carbone des algorithmes va faciliter le travail des décideurs. La notion d'IA frugale est fortement promue en France par le Ministère de la transition écologique[8] et elle repose sur une idée simple : un système d'IA qui crée plus de problèmes qu'il n'en résout doit être abandonné. La mesure globale est complexe, et en cas de doute, si aucun modèle frugal n'existe, la recommandation pourra être de renoncer à l'IA.

Concernant l'impact sur les métiers, la démarche est également complexe. Quelques collectivités françaises pilotes utilisent une version adaptée du projet MAIAT (« Mesure de l'Acceptabilité sociale de l'IA au Travail ») développé par l'Icam et Yann Ferguson dans le cadre du Partenariat mondial sur l'intelligence artificielle[9]. L'objectif est de sortir de la dialectique simpliste sur l'IA comme facteur de suppression immédiate et massive d'emploi, et de mesurer son impact sur plusieurs facteurs d'acceptabilité professionnelle et sociale. Notamment, pour le service public local, l'impact possible de l'IA sur la relation entre les usagers et les agents publics.

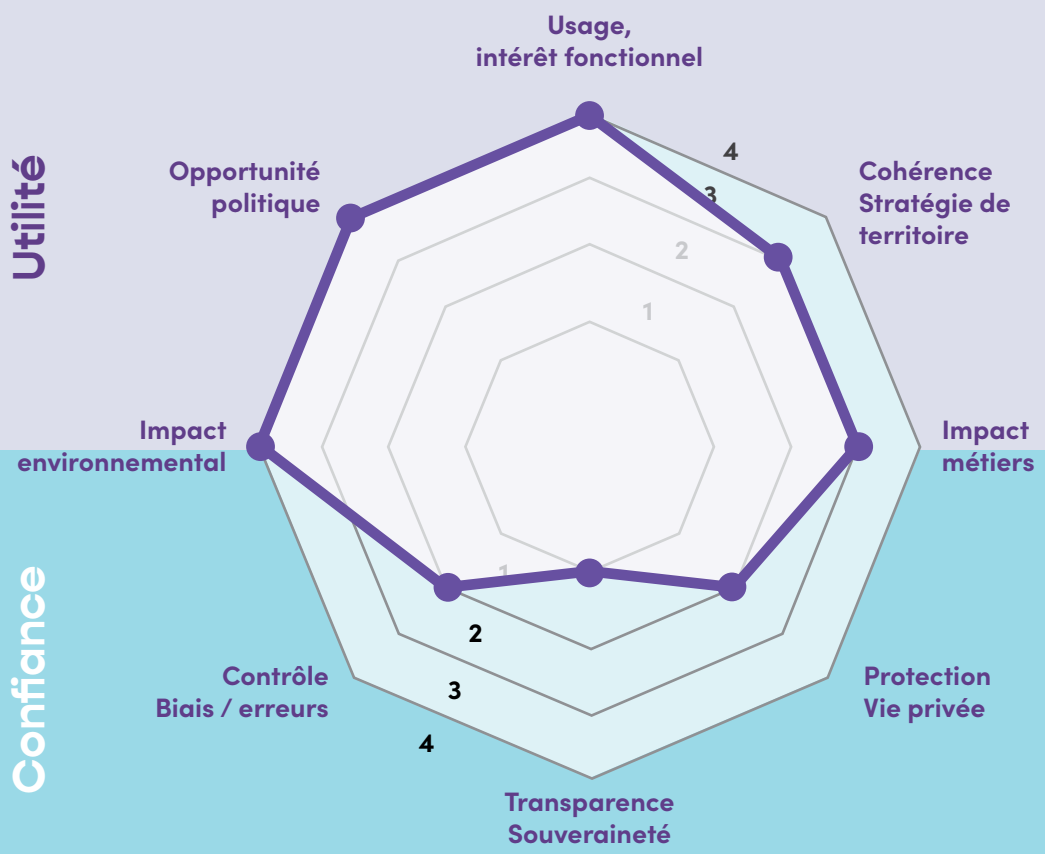
Pour les critères environnement et métier, on recherche donc à la fois un impact positif et la limitation des effets négatifs. D'où l'idée d'une balance.

De façon très concrète, l'outil imaginé par le cabinet CIVITEO avec les collectivités pionnières repose sur une grille d'analyse à quatre niveaux : utilité/confiance faible; utilité/confiance modérée; utilité/confiance moyenne; et utilité/confiance élevée. Elle se traduit par une représentation sous forme d'un radar qui permet in fine à la collectivité d'obtenir une appréciation de l'impact que pourrait avoir un système d'IA dans la mise en œuvre de telle ou telle mission de service public.

8 Ministère de la transition écologique et de la cohésion des territoires français (2023). *Appel à projets Démonstrateurs d'IA frugale pour les transitions écologique et énergétique dans les territoires.*

9 CRESPIN, P. et *Al.* (2022). *L'IA au travail : propositions pour outiller la confiance,* Conférence Nationale sur les Applications Pratiques de l'IA, Juin 2022, Saint-Etienne, France.

Shéma 2 : Le radar des systèmes d'IA territoriale.



Dans ce radar, la solution proposée peut sembler très séduisante en termes de bénéfices immédiats, d'affichage politique et de stratégie, mais le système d'IA offre très peu de garanties en termes de transparence, de maîtrise publique, de respect du RGPD de risques de discrimination, etc. Il s'agit peut-être d'un algorithme très en vogue et la tentation de succomber à l'effet de mode existe!

Bien sûr, aucun système n'est optimal. Et il y a toujours des vérifications et des garanties à apporter. Étude faite, deux scénarios sont possibles : « la collectivité décide d'expérimenter le recours à l'IA dans tel service, et voici comment » ou « la collectivité décide de renoncer à ce projet d'IA, pour l'instant, et voici pourquoi… ».

La vigilance est de mise. Certains systèmes d'IA n'apportent pas les garanties nécessaires à leur déploiement en confiance pour des missions d'intérêt général. Mais d'autres parfois, offrent ces garanties… et s'avèrent inutiles.

Espace européen des données de santé : enjeux et défis pour l'utilisation secondaire des données de santé

Nathalie De Grove-Valdeyron

Professeure de droit public, Université Toulouse Capitole (IRDEIC)
Chaire Jean Monnet (2023-2026)

Ursula Von der Leyen a fait du numérique une des principales priorités de son mandat en estimant que « les données et l'IA[1] sont les ingrédients de l'innovation qui peuvent aider à trouver des solutions aux enjeux sociétaux actuels [notamment] dans le domaine de la santé [..] ». Dès 2018, la Commission avait identifié, dans sa communication sur la transformation numérique des soins, les obstacles majeurs à l'adoption de solutions numériques dans le domaine des soins. Au titre des priorités figuraient déjà, d'une part, « l'accès sécurisé des citoyens à leurs données de santé et leur partage » et, d'autre part, « la qualité de ces données pour promouvoir la recherche, la prévention des maladies et la personnalisation des soins ». Ces deux priorités constituent les deux grands objectifs de la mise en place de l'espace européen des données de santé (European Health Data Space, ou EHDS selon l'acronyme anglais), qui a fait l'objet d'une proposition de Règlement présentée par la Commission européenne le 3 mai 2022[2] (ci-après, le règlement EHDS). Par ailleurs, dans sa communication du 19 février 2020 (COM [2020] 66), la Commission proposait une stratégie pour les données qui s'est concrétisée dans différents « Acts », en réalité des règlements, qui doivent être pris en compte à la fois pour appréhender le contexte dans lequel va s'insérer le règlement en cours d'adoption, mais aussi pour comprendre l'articulation, encore incertaine sur beaucoup d'aspects, entre le règlement EHDS et ces différents textes. Ainsi, sans pouvoir approfondir cette question dans le cadre de ce papier, on se contentera de mentionner les quelques textes clés, indissociables de la proposition de règlement. Il s'agit tout d'abord du « Data Act », règlement sur les données du 13 décembre 2023[3], qui met en œuvre certains objectifs du « Data Governance Act », règlement (UE) 2022/868 sur la gouvernance des données[4]. Il s'agit aussi de « l'Intelligence Artificial Act » du 21 avril 2021, qui a fait l'objet d'un accord politique le 8 décembre 2023, adopté à l'unanimité par les gouvernements des États membres en janvier 2024. Ces règlements ont une dimension transversale et ont vocation à s'appliquer au domaine de la santé pour certains de leurs aspects.

1 Voir la Communication de la Commission du 20 février 2020, COM (2020)65 final, *Livre blanc : Intelligence artificielle, Une approche européenne axée sur l'excellence et la confiance.* Ce livre blanc accompagne la stratégie européenne pour les données.

2 COM (2022),197 final du 3 mai 2022.

3 Règlement (UE)2023/2854 du 13 décembre 2023 du Parlement européen et du Conseil concernant des règles harmonisées portant sur l'équité de l'accès aux données et de l'utilisation des données, JOUE L 22 décembre 2023.

4 Exemple d'articulation : Selon le considérant 115 [...] Ce règlement (Data Act) s'entend « sans préjudice des règles plus spécifiques dans le cadre du développement d'espaces européens communs de données, ou sous réserve des exceptions prévues par le présent règlement, sans préjudice du droit de l'Union et du droit national prévoyant l'accès aux données et autorisant l'utilisation des données à des fins de recherche scientifique ». Voir par exemple la demande de mise à disposition de données à caractère personnel en vertu de l'article 14 pour besoin exceptionnel (situation d'urgence) et le traitement des données dans ce cadre (art.17, 1 point g et h) (pseudonymisation/anonymisation).

Il en est ainsi aussi, notamment, des règles relatives au partage volontaire des données telles que celles mises en place par des organisations altruistes en matière de données reconnues en vertu du règlement (UE) 2022/868 du 30 mai 2022 (strictement encadrées[5]) ou encore des règles relatives à la mise à disposition par certains détenteurs de données, de données à caractère personnel en cas de besoin exceptionnel[6]. En ce qui concerne l'IA, très tôt l'Union européenne a souligné l'importance des systèmes d'IA appliqués à de vastes bases de données notamment pour la recherche et le développement d'une médecine de plus en plus précise et personnalisée. Parmi les autres règlements qui devront s'articuler avec le règlement EHDS, il faut aussi mentionner, de manière plus spécifique, les règlements (UE) 2017/745 sur les dispositifs médicaux et (UE) 2017/746 sur les dispositifs de diagnostic in vitro de 2017 et, cela va de soi, le règlement (UE) 2016/679 sur la protection des données du 27 avril 2016 (RGPD)[7] ainsi que le règlement (UE) 2018/1725 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les institutions, organes et organismes de l'Union[8]. Une clarification des interactions entre ces différents textes de droit dérivé serait certainement bienvenue sous peine d'interprétations divergentes lors de la mise en application concrète du règlement par les États membres. La directive 2011/24/UE relative à l'application des droits des patients en matière de soins de santé transfrontaliers retient aussi l'attention en ce qu'elle est à l'origine de la création du réseau e-Health qui rassemble, sur la base du volontariat, les représentants des États membres (experts nationaux de la santé en ligne). Ce réseau a été le principal organe stratégique et de gouvernance notamment sur les questions d'interopérabilité des données de santé électroniques, mais son rôle, limité à une action d'appui et à l'adoption d'actes non contraignants, n'a pas permis d'atteindre ses objectifs, ce qui a été un des éléments qui a conduit à la proposition de règlement EHDS.

Il est certes trop tôt pour apprécier l'impact concret qu'aura la création de ce nouvel espace commun de données dans le domaine de la santé, particulièrement pour l'usage secondaire des données de santé, au stade actuel d'élaboration du règlement, au moment où sont écrites ces lignes[9]. Néanmoins, en nous affranchissant volontairement des canons académiques français, il nous paraît intéressant de nous pencher sur les enjeux de ce nouvel espace. Qu'apportera l'EHDS, notamment dans le cadre de l'utilisation secondaire des données de santé (II) et comment la mise à disposition de ces données pourra-t-elle se réaliser compte tenu des compétences respectives de l'Union européenne et des États? Répondre à ces deux questions implique préalablement de rappeler pourquoi la création d'un tel espace s'imposait (I).

I. Un espace européen des données de santé pour quoi faire?

L'objectif du règlement est ambitieux : il s'agit de faire en sorte que « les personnes physiques de l'Union aient davantage de contrôle sur leurs données de santé dans la pratique », mais il s'agit aussi de « garantir un cadre juridique consistant en des mécanismes de gouvernance fiables aux échelons de l'Union et des États membres et en un environnement de traitement sécurisé ». La construction de cet espace devrait en outre contribuer à la mise en place d'un marché unique des produits et services de santé numérique en facilitant les flux transfrontières de données de santé électroniques. C'est donc naturellement l'article 114 du TFUE qui a été retenu comme base juridique du règlement puisqu'il s'agit d'améliorer la libre circulation des biens et des services tout en tenant compte d'un niveau élevé de protection de la santé pour tirer parti du marché intérieur et mettre ainsi fin aux obstacles à l'accès et l'utilisation des données de santé électroniques. Cette base juridique est accompagnée de l'article 16 du TFUE (droit à la protection des données personnelles que sont les données de santé).

5 Hulin A-S. et Castets Renard C. (2022). *Le règlement sur la gouvernance des données ou la promesse d'un marché des données à l'européenne*, Dalloz IP/IT, 575.

6 Voir par exemple la demande de mise à disposition de données à caractère personnel en vertu de l'article 14 pour besoin exceptionnel (situation d'urgence) et le traitement des données dans ce cadre (art.17, 1 point g et h) (pseudonymisation/anonymisation).

7 JOUE L 119, du 4 mai 2016, p. 1.

8 JOUE L 295, du 21 novembre 2018, p. 39.

9 Au moment de mettre sous presse, la proposition a été adoptée par le Parlement européen en première lecture le 24 avril 2024. Cf. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0331_FR.html (*document P9-TA(2024)0331*). Le texte est en attente de la position du Conseil en première et unique lecture.

La proposition de règlement distingue deux types d'utilisation des données de santé, l'utilisation dite « primaire » et l'utilisation « secondaire » des données de santé. L'utilisation qualifiée de « primaire » vise à assurer la continuité des soins et à améliorer la prise en charge des patients dans un contexte transfrontalier. En effet, il apparaît, à la lumière des études qui ont été commandées par la Commission, que cette continuité est loin d'être acquise, même au niveau des États membres. Est notamment pointée du doigt la diversité des normes et des spécifications techniques utilisées (logiciels incompatibles entre médecine de ville et hôpital, systèmes d'exploitation inutilisables, etc.). L'interopérabilité et la portabilité des données, qui sont un préalable, n'étant pas assurées de façon satisfaisante à l'échelon national, voire régional et local, il est aisé d'imaginer l'amplification des problèmes dans un contexte transfrontalier. Ce constat permet de mesurer le caractère ambitieux de la proposition de la Commission, ne serait-ce que sous l'angle du défi technique qu'elle représente. La crise sanitaire européenne liée à la COVID 19 a montré la nécessité de renforcer la coordination européenne afin, notamment, de mettre en place une politique concertée en réaction aux urgences sanitaires et elle a permis de mettre en œuvre, grâce à cette coordination, des solutions techniques garantissant la transférabilité des données et leur interopérabilité, dans le respect de la confidentialité. Les défis que représentent les aspects de nature technique pourront donc, au dire des spécialistes, être relevés. Le certificat Covid numérique de l'Union est sans doute l'illustration la plus tangible[10] des fruits de cette coopération européenne et des avantages qu'il a représentés en termes de facilitation de la liberté de circulation dans l'Union. Aujourd'hui la plateforme numérique sécurisée, Masanté@UE (*MyHealth@EU*)[11], initialement dénommée « e-HDSI », a été conçue pour permettre le transfert transfrontalier des données de santé, dans le cadre de l'utilisation primaire des données de santé. Elle n'attend plus qu'à être utilisée par l'ensemble des États membres, pour « fluidifier » le parcours des soins au sein de l'Union, au rythme qui sera celui des États membres puisque le recours à cette plateforme se fait, à ce jour, sur la base du volontariat. Cette plateforme connecte en effet déjà aujourd'hui les États membres via leurs points de contact nationaux (NCPe-H) (*National contact point for eHealth*),[12] mais attend l'adoption du règlement pour pouvoir déployer tout son potentiel. Elle devrait, à l'horizon 2025, permettre grâce au recours à un format européen d'échange des données de santé, le transfert des prescriptions électroniques, des dossiers médicaux électroniques, des rapports de sortie d'hôpital, des comptes rendus de laboratoire et des rapports d'imagerie médicale. L'utilisation dite secondaire des données de santé apparaît comme une question encore plus sensible.

II. Qu'apporte le règlement s'agissant de l'utilisation dite « secondaire des données de santé »?

L'utilisation secondaire des données, deuxième objectif de l'EHDS, est présentée dans la proposition de la Commission comme (l'utilisation) « [...] » à d'autres fins dans l'intérêt de la société telles que la recherche, l'innovation, l'élaboration de politiques, la sécurité des patients, la médecine personnalisée, les statistiques officielles ou les activités règlementaires ». Elle trouve sa base juridique de traitement dans l'article 9 § 2 points g, h, i et j du RGPD.

L'idée mise en avant est de permettre ainsi aux chercheurs, aux innovateurs, aux décideurs et aux organismes de règlementation d'avoir accès à des données de qualité pour leurs travaux de manière sécurisée, avec une gouvernance fiable et à des coûts inférieurs à ceux d'un système fondé sur le consentement.

10 Règlement (UE) n° 2021/953 du 14 juin 2021 relatif à un cadre pour la délivrance, la vérification et l'acceptation de certificats COVID-19 interopérables de vaccination, de test et de rétablissement (Certificat Covid numérique de l'UE) afin de faciliter la libre circulation pendant la pandémie de Covid-19, *JOUE* n° L 211, 15 juin 2021, p. 1-22.

11 Cette plateforme était appelée E-HDSI (eHealth Digital Service Infrastructure) voir, De Grove-Valdeyron N. (2022). « Santé Publique », Annuaire de droit de l'Union européenne, dir. C. Blumann et F. Picod, Panthéon Assas, 847 à 876, en ligne : https://www.cairn.info/annuaire-de-droit-de-l-union-europeenne--9782376510567-page-847.htm.

12 En France, l'Agence du numérique en santé.

La notion même d'utilisation secondaire a donné lieu à des débats et a des recommandations par l'action conjointe TEHDaS[13]. Elle est définie dans le règlement comme une utilisation en vue de la recherche scientifique pour des finalités strictement mentionnées dans le règlement (voir *infra),* mais elle sous-entend la possibilité de réutiliser des données pour des raisons autres que celles pour lesquelles elles ont été initialement collectées ce qui pose la question essentielle du consentement ou, à tout le moins, de l'information du patient à l'utilisation de ses données en vue de la recherche. En France, la question du consentement a fait l'objet d'un avis le 9 mai 2023[14] du CCNE (comité consultatif national d'éthique) et du comité national pilote d'éthique du numérique (CNPEN). Cet avis, qui comporte de nombreuses recommandations, fait état des différents types de consentements qui pourraient être requis dans le cadre de plateformes de données de santé : consentement général (absence de connaissance des spécificités des études au moment de la demande), consentement dynamique (via une plateforme de communication en ligne spécialisée qui facilite le consentement entre chercheurs et participants), consentement dynamique spécifique (possibilité de préciser les accès autorisés), et méta-consentement (options de consentements selon les modèles et formes de projets) qui est un modèle utilisé au Canada par la plateforme du GRIIS[15]. Aujourd'hui, la façon dont le consentement est recueilli au niveau des États membres ne fait pas l'objet de dispositions communes; ainsi s'agissant de la création du Dossier médical partagé qui a connu un départ difficile en France, la technique de « l'*Opt out*[16] » a été retenue, ce qui a permis la généralisation de « mon espace santé », 98 % des assurés bénéficiant en février 2023 de cet espace en ligne (sauf opposition à sa création).

Par ailleurs, il convient de rappeler que la façon dont le consentement est recueilli ne fait pas aujourd'hui l'objet de dispositions européennes, car il s'agit d'un domaine relevant de l'organisation des systèmes de santé, et donc de la compétence des États, conformément à l'article 168 § 7 du TFUE. Cette question du consentement s'avère un des points particulièrement sensibles de la proposition et le législateur européen semble s'orienter aujourd'hui vers la possibilité d'une marge de manœuvre laissée aux États dans ce domaine, compte tenu de leurs particularités nationales, en imposant toutefois la mise en place de mécanismes permettant aux personnes d'exercer leurs droits et de revenir sur un refus de consentement[17]. La Commission réfléchit par ailleurs de son côté, en s'appuyant sur le réseau e-health, à la mise en place d'une « harmonisation[18] » dans le traitement du consentement qui s'apparenterait à un formulaire unique permettant un traitement comparable du consentement dans l'ensemble des États membres. Le règlement « DGA » invite aussi pour sa part, en son article 22, la Commission à adopter des actes délégués formant un « recueil de règles » fixant, certes dans le cadre spécifique de l'altruisme en matière de données, « des exigences appropriées en matière d'information pour veiller à ce que les personnes concernées et les détenteurs de données reçoivent, avant qu'un consentement ou une autorisation ne soit donnée en matière d'altruisme de données, des informations suffisamment détaillées, claires et transparentes concernant l'utilisation des données [et] les outils permettant de donner et de retirer le consentement [...] (art.22§1, a et pour la finalité voir art.21§1 a)). Il pourrait s'agir de données de santé et ces actes délégués pourraient, selon nous, servir d'inspiration au recueil de consentement dans le cadre de l'utilisation secondaire des données de santé, si une forme de consentement est exigée par un État membre conformément à son droit national.

13 TEHDaS (Towards European Health Data Space), (2023). Qualitative Study to Assess Citizens' Perception of Sharing Health Data for Secondary Use and Recommendations on How to Engage Citizens in the EHDS, deliverable 8.1, en ligne: < https://tehdas.eu/tehdas1/app/uploads/2023/03/tehdas-study-to-assess-citizens-perception-of-sharing-health-data-for-secondary-use.pdf>.

14 CCNE et CNPEN (2023). *Plateformes de données de santé : enjeux d'éthique*. Avis 143 du CCNE, Avis 5 du CNPEN, en ligne : < https://www.ccne-ethique.fr/sites/default/files/2023-05/CCNE-CNPEN_GT-PDS_avis_final27032023.pdf>.

15 Voir Barton A., Cumyn A. et Ethier J.-F, *Le métaconsentement en lumière*, en ligne : https://ssaquebec.ca/wp-content/uploads/2018/05/Cumyn_A.pdf. Voir aussi Barton A., Cumyn A. et Ethier J.-F., *Atelier B3 : Le métaconsentement : point de citoyen sur une approche alternative au consentement*, en ligne : https://www.msss.gouv.qc.ca/professionnels/documents/comites-d-ethique-de-la-recherche/journees-detude/2019-9e-journees/13h30-24oct_B3-Cumyn_JECER2019.pdf.

16 Voir : https://www.ameli.fr/assure/protection-donnees-personnelles.

17 Selon la dernière version disponible « *les personnes physiques auront le droit de s'opposer à tout moment et sans motivation au traitement des données de santé électroniques à caractère personnel les concernant à des fins d'utilisation secondaire [...]* ». Les États membres devront prévoir un mécanisme d'opposition accessible et facilement compréhensible permettant d'exercer ce droit (art.48 bis-1), mais les États conservent une possibilité de mise à disposition, malgré opposition, pour certaines données (art. 34§1 a) à c)) (art.48 bis-3).

18 Le terme « d'harmonisation » ne doit pas être compris ici sous un angle juridique au sens de l'article 114 du TFUE mais au sens de rapprochement.

III. L'utilisation secondaire des données de santé : quelles données pour quelles finalités?

La réalisation de l'EHDS implique une « gouvernance » et donc l'adoption de règles permettant la mise à disposition des données dans des conditions fixées par le règlement. Sans pouvoir ici nous livrer à une analyse de cette gouvernance, nous nous concentrerons sur deux points délicats dans le cadre de celle-ci : les données visées et la finalité du traitement. Tout d'abord, seules certaines catégories de données (de santé) électroniques sont visées à des fins d'utilisation secondaire (voir art.33). Il s'agit, sans toutes les énumérer ici, de données provenant des dossiers médicaux électroniques (DME), des données administratives relatives aux demandes et remboursements; des données génétiques, génomiques, protéomiques, des données générées automatiquement par des dispositifs médicaux, etc. Quant aux seules finalités autorisées, elles sont visées à l'article 34[19], il s'agit ainsi : a) des raisons d'intérêt public dans le domaine de la santé publique et de la santé au travail (activités destinées à la protection contre les menaces transfrontières graves pour la santé et à la surveillance de la santé publique, ou des activités destinées à la garantie d'un niveau élevé de qualité et de sécurité des soins de santé y compris de sécurité des patients et des médicaments ou dispositifs médicaux; b) de l'élaboration des politiques et les activités règlementaires afin d'aider les organismes du secteur public ou les institutions, organes et organismes de l'Union, dont les autorités règlementaires, dans le secteur de la santé ou des soins, à accomplir les tâches inscrites dans leur mandat; il s'agit aussi c) de produire des statistiques officielles en rapport avec le secteur de la santé et des soins; ou encore, de e) mener la recherche scientifique ayant trait aux secteurs de la santé ou des soins, contribuant à la santé publique ou à l'évaluation des technologies de la santé (ou garantissant un niveau élevé de qualité et de sécurité des soins de santé, des médicaments ou des dispositifs médicaux) y compris les activités de développement et d'innovation pour les produits ou services; la formation, les tests et l'évaluation des algorithmes, les systèmes d'IA et applications de santé numérique…

Le règlement précise ensuite les obligations pesant sur *les organismes responsables de l'accès aux données*, désignés par les États membres, qui doivent mettre à disposition des utilisateurs de données, dans un environnement de traitement sécurisé, les données "pertinentes". Ils tiennent un catalogue national des ensembles de données (source, nature, conditions de mise à disposition) et sont soumis à certaines obligations détaillées dans le règlement.

Les détenteurs de données doivent, sur demande, mettre à disposition des demandeurs de données les données sollicitées aux fins précédemment indiquées dans un délai raisonnable limité à 3 mois à compter de la réception de la demande (ce délai est jugé irréalisable pour l'instant, le délai de mise à disposition avoisinant plutôt les 18 mois en France).

Le demandeur doit quant à lui fournir une explication détaillée de l'utilisation prévue des données de santé, et protéger des risques de réidentification, etc. Une demande unique peut être sollicitée quand elle concerne plus d'un État membre (via la plateforme sécurise dédiée DonnéesDeSanté@UE). Les États membres et la Commission s'efforcent de garantir l'interopérabilité des données de santé avec les autres espaces communs de données à venir.

Si les principales règles, très rapidement évoquées ci-dessus, ont bien été posées dans le règlement, on comprend aisément que la mise en œuvre concrète de celui-ci et son efficacité dépendront des nombreux actes délégués et d'exécution confiés à la Commission européenne. Tant que ceux-ci n'ont pas été adoptés, la mise en application effective du règlement risque d'être retardée de plusieurs mois voire de plusieurs années, car à ce stade un certain nombre d'incertitudes subsistent. Pour l'heure une nouvelle action TEDHaS 2 a été mise en place et elle devrait aider les États en anticipant les problèmes susceptibles de se poser en amont.

19 En tenant compte des amendements proposés par le Parlement européen.

Seul un cadre solide de gouvernance des données de santé permettra de maintenir la confiance du public dans ce nouvel espace européen des données de santé qui devrait permettre de tirer le meilleur parti du potentiel de la santé numérique. La confiance dans cet EHDS passe aussi par l'information des patients/citoyens et leur adhésion pour partager des données de santé indispensables à la recherche (dans le cadre d'une utilisation dite "secondaire" des données de santé). La juste mesure est à trouver entre l'intérêt public qui résulte de l'exploitation de vastes bases de données de santé[20] et les perspectives vertigineuses qui peuvent découler de cette exploitation au niveau de l'Union européenne (notamment en y appliquant des systèmes d'intelligence artificielle) d'une part, et la préservation des intérêts privés, d'autre part. Ce délicat équilibre devrait intégrer dans la balance le respect du principe de proportionnalité[21] en gardant à l'esprit le recours possible par l'Union ou les États membres à l'article 89§2 du RGPD qui prévoit, la possibilité de dérogations, dans le cas de la recherche scientifique, à certaines conditions.

La participation des citoyens et de l'ensemble des parties prenantes notamment à travers le Comité de l'EHDS et le forum des parties prenantes devrait renforcer la confiance dans ce nouvel espace européen de données propre au domaine de la santé. Les infrastructures de données de santé existantes, mises en place dans un cadre règlementaire national, encouragent déjà les activités de recherche et d'innovation dans le domaine de la santé, mais elles doivent être pensées aujourd'hui dans un contexte transfrontalier avec des mécanismes de gouvernance adaptés pour exploiter au mieux les bases de données au service des finalités prévues par le règlement. C'est à ce prix que des progrès majeurs pourront être réalisés dans le domaine de la recherche en santé en Europe.

20 Supiot E. (2022/1). « Les données génétiques au service de la recherche scientifique », Droit, santé et société, n° 1, 53-58.

21 Il s'agit aussi d'éviter des coûts excessifs ou des formalités trop lourdes liées au recueil du consentement, qui pourraient être mises en place par certains États membres, ce qui pourrait nuire à la cohérence du règlement et réintroduire une nouvelle fragmentation juridique de l'accès aux données que le règlement vise justement à combattre.

Données et IA en santé : quels sont les nouveaux défis?

Anita Burgun

MD, PhD. Professeur d'informatique médicale à l'Université de Paris. Chef du département d'informatique médicale des Hôpitaux Georges Pompidou et Necker (AP-HP).

Introduction

La majorité des informations qu'on qualifie de données de santé est, en dehors de l'imagerie médicale et des résultats de laboratoire, saisie par les professionnels de santé. Ceux-ci vivent souvent ce temps d'alimentation du dossier patient comme une charge de travail qui diminue leur disponibilité pour des tâches plus médicales. Pour les urgentistes du *Whittington Hospital* de Londres par exemple, le temps de saisie par le médecin pour un patient vu aux urgences avait été plus que multiplié par 3 lors du passage du dossier papier au dossier informatisé[1]. Récemment, une étude menée aux USA, en pratique ambulatoire, et médecine d'adulte, portant sur 100 millions de consultations pratiquées par 155 000 médecins, a montré qu'un médecin passe en moyenne 16 minutes et 14 secondes par consultation sur son ordinateur à interagir avec le système d'information clinique. Ce temps est consacré essentiellement à l'analyse du dossier patient (33 %), à la documentation du dossier (24 %) et à la prescription informatisée (17 %). Au total elle/il y passe entre 3,5 et 6 heures par jour, la médiane étant de 4h1/2[2].

L'arrivée de l'Intelligence Artificielle (IA) en santé a encore fait croitre la « pression de la donnée », elle a laissé penser qu'il suffit de rendre les données disponibles pour créer de beaux modèles très utiles pour prendre en charge les patients, et certains tenants du « solutionnisme technologique » ont imaginé naïvement qu'on pourrait remplacer les cliniciens par des modèles d'intelligence artificielle pour certaines tâches, en ignorant les problématiques de production des données et de décision médicale. Contrastant avec ces prises de position simplistes, les sociétés savantes médicales américaines alertent actuellement sur la nécessité de redéfinir les priorités de l'IA en santé. En novembre 2023, *l'American Medical Informatics Association (AMIA)* s'est associée avec *l'Association of Medical Directors of Information Systems et l'Alliance for Nursing Informatics*, pour évaluer les capacités de l'IA à réduire la charge de travail des cliniciens et limiter le burn-out généré par l'informatisation. L'objectif de l'IA devient alors d'alléger les tâches de recherche et de production des données dans le dossier patient. En 2023 également, *l'American Medical Association* (AMA) a demandé que les cliniciens aient une voix prépondérante pour définir quel rôle l'IA doit tenir en médecine. L'AMA a aussi insisté sur la nécessité de disposer de systèmes d'information clinique matures comme préalable à l'implémentation de toute IA et d'évaluer les bénéfices pour les cliniciens de l'IA en vie réelle.

1 Armstrong S. (2017). « Data, data everywhere: the challenges of personalised medicine », BMJ, 359, en ligne <https://www.bmj.com/content/359/bmj.j4546>.

2 Overhage J-M. et McCallie D. Jr (2020). « Physician Time Spent Using the Electronic Health Record During Outpatient Encounters: A Descriptive Study », Ann. Intern. Med., vol. 172-3, 169174, en ligne : < https://pubmed.ncbi.nlm.nih.gov/31931523/>.

Notre objectif est de construire le cadre permettant d'implémenter les Systèmes de Santé Apprenant (*learning health care systems*), qui sortent de la vision schématique de la réutilisation des données de soins pour la recherche en proposant un objectif intrinsèquement translationnel. Nous abordons en particulier trois points : la prise en compte des caractéristiques multimodales des données patients, les approches fédérées de partage, l'aide à la décision clinique dans les systèmes apprenants.

I. Des approches multimodales

La crise COVID a généralisé l'idée de réutiliser pour la recherche les données patients générées dans le cadre du soin plutôt que de resaisir des données comme on le fait dans les essais cliniques. Le dossier patient est source de données structurées sous forme de codes provenant d'un thesaurus, de données produites par des moniteurs (pouls, tension artérielle, etc.) ou de résultats de laboratoire, et de données non structurées, qu'il s'agisse d'images ou de documents textuels. Les données individuelles présentes dans les systèmes d'information cliniques décrivent à la fois l'état du patient au moment de l'examen, son histoire clinique, et ses antécédents familiaux. Elles sont par nature multimodales, incluant les différentes modalités d'imagerie médicale, les comptes-rendus textuels, les résultats des analyses biologiques et « omiques », les mesures continues, les prescriptions thérapeutiques (médicaments, actes) et les autres données structurées.

Leur caractère riche et multimodal donne à ces données brutes une place de choix pour entraîner les modèles d'Intelligence Artificielle[3] et ouvrir de nouvelles opportunités d'applications[4]. Alors que la grande majorité des modèles d'IA ont jusqu'à maintenant été entraînés sur une source de données, l'imagerie en particulier (et on utilisait le terme de multimodalité si on combinait par exemple scanner et IRM) et pour une tâche précise, par exemple classifier des lésions en malignes ou bénignes, il s'agit maintenant de transcender les modèles monotâches et unimodaux pour inclure le maximum de sources de données. Par exemple, il est utile d'intégrer des données génomiques, de l'imagerie médicale et d'autres données cliniques du dossier patient pour faire des choix thérapeutiques en oncologie de précision; des données biologiques, cliniques, d'imagerie, et d'objets connectés pour créer des jumeaux numériques.

Les maladies génétiques rares constituent un domaine médical qui requiert tout particulièrement d'intégrer un maximum de données pour établir les associations entre phénotypes et génotypes, et réutiliser la connaissance extraite de ces données dans le cadre de la prise en charge de nouveaux patients. La recherche menée à l'Institut Imagine porte sur les maladies génétiques et l'Institut travaille de façon très étroite avec l'hôpital Necker Enfants Malades de l'Assistance Publique-Hôpitaux de Paris. On compte au total environ 8000 maladies génétiques rares et chaque maladie est caractérisée par une incidence très faible. Chaque cas doit donc être décrit avec un maximum de précision pour obtenir une représentation la plus complète possible. La totalité des sources d'information doit être de ce fait intégrée, ce qui inclut toutes les données cliniques du dossier patient et l'ensemble des données génétiques disponibles.

3 Acosta J. N. et Al. (2022). « Multimodal biomedical AI », Nat Med, vol. 28-9, 17731784.

4 Topol E. J (2023). « As artificial intelligence goes multimodal, medical applications multiply », Science, vol. 381, 6663, en ligne : <https://pubmed.ncbi.nlm.nih.gov/37708283/>.

II. Une approche fédérée

Une représentation précise du patient ne peut pas être considérée comme exempt de risque de réidentification. Un exemple est celui des déficits intellectuels syndromiques avec dysmorphie faciale où il est nécessaire de disposer des photographies de visage pour établir le diagnostic. L'infrastructure de stockage des données et le modèle de gouvernance des données doivent donc être adaptés pour répondre de manière optimale à la combinaison de deux desiderata suivants :

1. Les données de patients porteurs de maladies génétiques rares sont extrêmement sensibles et doivent être ultra-protégées ;

2. Pour une maladie rare donnée, le nombre de cas par centre voire par pays peut être extrêmement faible et seule une mise en commun permettra d'obtenir des avancées significatives en recherche.

Une approche fondée sur des bases de données distribuées et un apprentissage fédéré, dans laquelle les données restent locales, mais les modèles sont partagés entre institutions, est la solution qui permet de répondre à ces injonctions contradictoires[5]. L'objectif est que tous les centres participants harmonisent leurs données en adoptant un modèle de données commun[6], pour que les mêmes requêtes puissent être utilisées par tous les centres et que l'algorithme puisse être entraîné dans chaque centre puis partagé.

Dans un système fédéré, les données restent hébergées et contrôlées localement et ne sont exportées ni vers d'autres sites ni vers un hub de données central. La fédération de bases de données (distribuées) donne de meilleures garanties contre de possibles brèches de confidentialité et permet de faciliter le contrôle par les producteurs de données, qui peuvent décider de participer ou non à un projet. Par conséquent, les collaborations entre centres sont renforcées, le passage à un niveau international facilité et la confiance des populations renforcée ou préservée.

L'approche fédérée permet de disposer de données multimodales produites au niveau de chaque site sans avoir à les stocker dans une mégabase de données collectant les données individuelles de plusieurs sites et plusieurs modalités. Il est possible ainsi de recruter plus facilement de nouvelles sources de données, qu'il s'agisse de modalités ou de sites.

Enfin, l'analyse fédérée des données repose sur une meilleure maîtrise du pipeline de production des données et de la qualité qui en résulte. Cette maîtrise de la chaîne de production et d'analyse des données est de nature à produire de meilleurs modèles d'IA. La validation locale est nécessaire pour vérifier la transposabilité de modèles externes[7]; elle est facilitée par cette approche distribuée dans laquelle chaque site est responsable.

L'approche fédérée a été adoptée par des réseaux internationaux, par exemple lors de la crise COVID, par le *Consortium for Clinical Characterization of COVID-19 by EHR* (4CE). Elle a sans retard permis de conduire dès le début de la crise, une recherche internationale sur la COVID, basée sur la réutilisation des données produites en routine par les hôpitaux, donc en conformité avec la préconisation de l'OMS d'exploiter les données de soin[8]. Cette approche a permis non seulement de produire des connaissances nouvelles sur la maladie, mais aussi de générer des modèles de prédiction et surtout d'en évaluer la transposabilité/généralisabilité[9].

5 Sheller M. J. et Al. (2020). Scientific Reports, vol. 10, 12 598, en ligne: https://www.nature.com/articles/s41598-020-69250-1 ».

6 Voir Kaliyaperumal R. et Al.. (2022). « Semantic modelling of common data elements for rare disease registries, and a prototype workflow for their deployment over registry data », Journal of Biomedical Semantics, vol 13, art. 9, en ligne: <https://jbiomedsem.biomedcentral.com/articles/10.1186/s13326-022-00264-6>.

7 Youssef A. et Al. (2023). « External validation of AI models in health should be replaced with recurring local validation », Nature Medecine, vol. 29-11, 26862687, en ligne : <https://www.nature.com/articles/s41591-023-02540-z>.

8 Brat G. A. et Al. (2020). « International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium », NPJ Digit. Med., vol 3-109, en ligne : <https://pubmed.ncbi.nlm.nih.gov/32864472/>.

9 Weber G.M. et Al. (2022). « International comparisons of laboratory values from the 4CE collaborative to predict COVID-19 mortality », NPJ Digit. Med., vol 5-74, 18, en ligne: <https://www.nature.com/articles/s41746-022-00601-0>.

Plus généralement l'approche distribuée/fédérée est préconisée au niveau national par certains pays comme l'Allemagne et le Canada. Le réseau Miracum allemand[10] rassemble des partenaires académiques, hospitaliers et industriels pour créer des centres de données locaux, utilisant des principes d'interopérabilité et d'harmonisation des données (création d'un modèle de données commun en particulier) qui permettent de partager aisément les outils de requêtage des bases de données et des modèles d'IA[11].

III. Tendre vers un système de santé apprenant

Les chercheurs, les médecins, les instances de régulation et les pouvoirs publics ont pris l'habitude de séparer le système de soins de la recherche : le système de santé produit des données de santé dans le cadre de la prise en charge des patients et ces données peuvent être utilisées dans un second temps pour faire de la recherche. Dans cette vision dichotomique, la question est d'ouvrir les données au chercheur en garantissant leur sécurité et la confidentialité et bien entendu les bonnes pratiques. C'est ce qui a donné naissance à des missions ministérielles qui visent à poser les bases d'une feuille de route en matière de réutilisation des données de santé, et ont conduit à des rapports comme le récent « Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire des données de santé » en France. Ce rapport préconise d'alléger la règlementation et d'accélérer la mise à disposition des données pour la recherche[12].

Il est cependant souhaitable de proposer une vision alternative et plus intégrative basée sur le modèle du système de santé apprenant pour que le « dernier kilomètre » qui vise à faire bénéficier le clinicien des produits générés par la mise à disposition des données de soins soit pris en compte.

Prenons l'exemple de l'hôpital Necker Enfants Malades, de l'Assistance Publique Hôpitaux de Paris (AP-HP). L'hôpital Necker est associé à un Institut de recherche sur les maladies génétiques, l'Institut Imagine. Celui-ci développe une recherche sur les maladies génétiques et ses chercheurs produisent des modèles biologiques « mécanistiques » de ces maladies à partir des expériences de « wet lab ». Le phénotypage des patients est réalisé à partir de l'entrepôt de données cliniques de l'hôpital Necker. Des méthodes de similarité sur les profils phénotypiques des patients servent à identifier des sous-types de patients dans les maladies pléiotropiques[13]. Les mêmes méthodes sont utiles en clinique devant un patient non diagnostiqué pour retrouver dans l'entrepôt de données des patients similaires pour lesquels un diagnostic a été établi[14]. La réutilisation des données de l'entrepôt de santé sert dans ce cas à réduire le délai avant le diagnostic et à mieux prendre en charge de nouveaux patients admis à l'hôpital[15]. La *figure 1* illustre le concept de système de santé apprenant tel qu'il a été développé à l'Institut Imagine et à l'hôpital Necker. Son intérêt pour les cliniciens a été démontré (un tiers des dossiers patients parmi le million que compte l'entrepôt en 2024 ont été « touchés » par une requête) ainsi que sa capacité à accélérer le diagnostic[16].

10 Prokosch H-U. et Al. (2018). « MIRACUM: Medical Informatics in Research and Care in University Medicine », Methods Inf. Med., vol. 57-S 01, en ligne: <https://pubmed.ncbi.nlm.nih.gov/30016814/>.

11 Kapsner L. A et Al. (2021). « Linking a Consortium-Wide Data Quality Assessment Tool with the MIRACUM Metadata Repository », Appl. Clin. Inform., vol. 12-4, 826835, en ligne: <https://pubmed.ncbi.nlm.nih.gov/34433217/>.

12 Marchand-Arvier J. et Al. (2023). Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire des données de santé, Inspection Générale des Affaires Sociales, 2023, en ligne : <https://sante.gouv.fr/IMG/pdf/rapport_donnees_de_sante.pdf>.

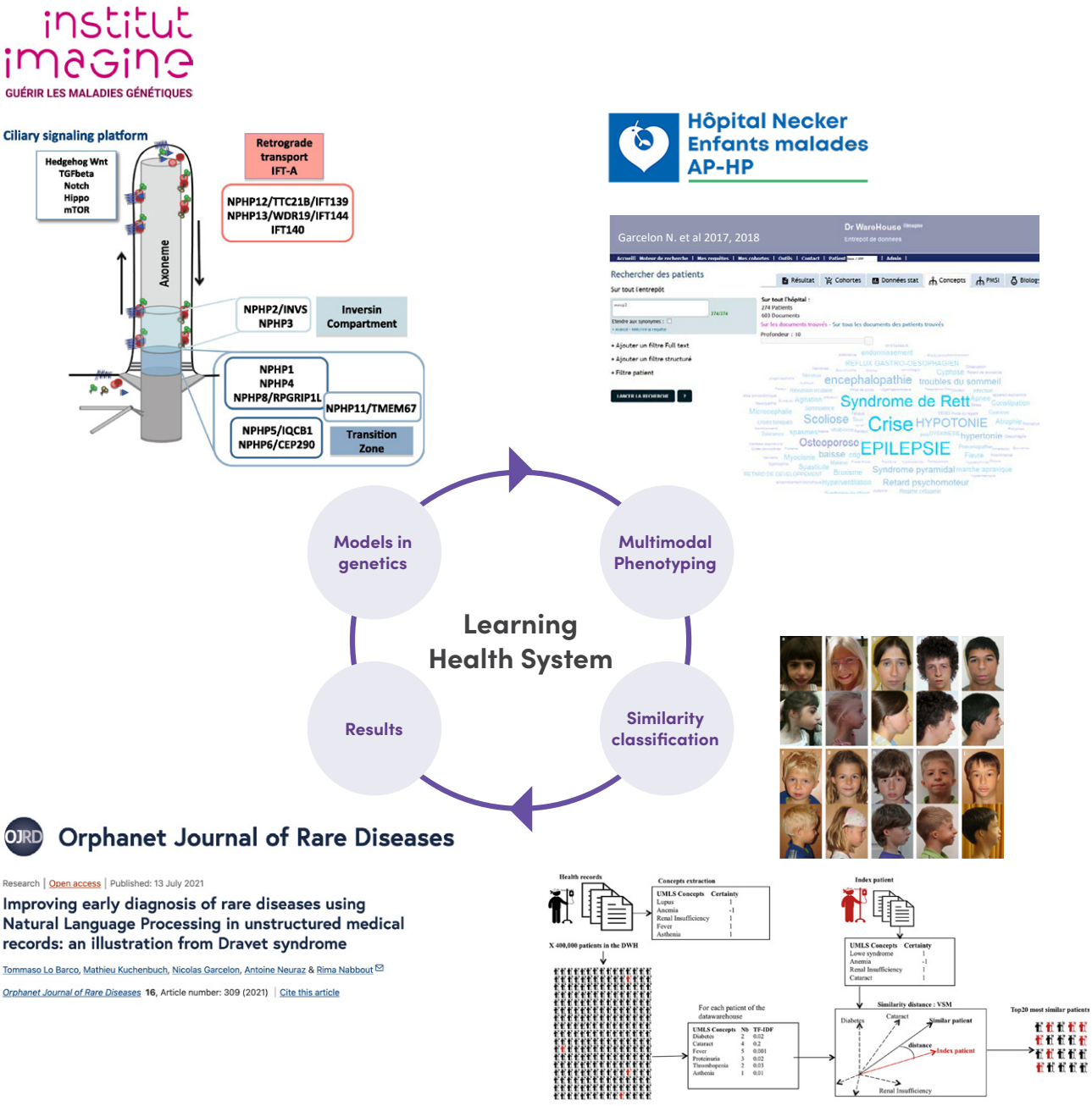
13 Chen X. et Al. (2019). « Phenotypic similarity for rare disease: Ciliopathy diagnoses and subtyping », J. Biomed. Inform., vol. 100, en ligne: https://pubmed.ncbi.nlm.nih.gov/31622800/>.

14 Lo Barco T. et Al. (2024). « Natural history of rare diseases using natural language processing of narrative unstructured electronic health records: The example of Dravet syndrome », Epilepsia, vol. 65-2, 350361, en ligne: <https://pubmed.ncbi.nlm.nih.gov/38065926/>.

15 Faviez C. et Al. (2024). « Performance and clinical utility of a new supervised machine-learning pipeline in detecting rare ciliopathy patients based on deep phenotyping from electronic health records and semantic similarity », Orphanet. J. Rare Dis., vol. 19-1, 55, en ligne: <Performance and clinical utility of a new supervised machine-learning pipeline in detecting rare ciliopathy patients based on deep phenotyping from electronic health records and semantic similarity>.

16 Lo Barco T. et Al. (2021). « Improving early diagnosis of rare diseases using Natural Language Processing in unstructured medical records: an illustration from Dravet syndrome », Orphanet. J. Rare Dis., vol. 16-1. 309, en ligne: <https://pubmed.ncbi.nlm.nih.gov/34256808/.

Figure 1 : le système apprenant de l'Institut Imagine et de l'hôpital Necker pour les maladies génétiques.



Sources :

https://www.sciencedirect.com/science/article/pii/S153204641930228X#f0020.

https://ojrd.biomedcentral.com/articles/10.1186/s13023-021-01936-9.

https://www.has-sante.fr/upload/docs/application/pdf/2019-11/colloque-has-a-burgun.pdf.

https://www.researchgate.net/figure/Facial-features-of-individuals-with-Meier-Gorlin-syndrome-Frontal-and-lateral-view-of-10_fig1_221831959.

https://www.sciencedirect.com/science/article/pii/S1532046417301764.

Conclusion

Comme l'ont souligné des prises de position récentes de cliniciens sur l'implémentation en clinique des modèles d'IA prédictifs[17], il apparait nécessaire d'adapter les procédures concernant des algorithmes d'intelligence artificielle alors que le risque de non transposabilité du modèle existe et de garantir leur interopérabilité avec les systèmes d'informatique clinique. Le système apprenant pour les maladies rares basé sur la recherche de patients similaires (*figure 1*) répond à ces demandes ainsi qu'à l'exigence d'explicabilité formulée par les cliniciens. Ce système apprenant tient compte de la totalité des informations disponibles pour chaque patient, qu'elles soient produites dans le cadre du soin ou lors de la recherche et il en intègre toutes les modalités. Si les algorithmes doivent être partagés, la complexité des données requiert une gouvernance locale et une approche fédérée. Une approche centralisatrice risquerait de diminuer la confiance des patients et des cliniciens et, surtout, de nous priver de la richesse des données brutes ce qui est indispensable pour produire des algorithmes à haute valeur ajoutée clinique.

17 Nong P. et Al. (2024). « How Academic Medical Centers Govern AI Prediction Tools in the Context of Uncertainty and Evolving Regulation », NEJM AI, vol. 1-3, en ligne : <https://ai.nejm.org/doi/full/10.1056/AIp2300048>.

Citoyens et seconde vie des données de santé : les enjeux en matière d'information et de droits

Caroline Guillot
Directrice citoyenne
du Health Data Hub

Stéphanie Combes
Directrice du Health Data Hub

Chaque fois qu'un patient interagit avec le système de santé, des données personnelles sont produites. Ces données combinent des informations précises sur une personne et celles relatives à son état de santé. Au départ, cette donnée sert à prendre en charge un patient de manière optimale[1]. Néanmoins, cette donnée peut aussi servir à la recherche; c'est l'exploitation secondaire des données de santé, sa « seconde vie »[2]! Les analyser ensemble, c'est se donner les moyens de tirer des enseignements qui permettront de mieux soigner l'ensemble des patients; des enseignements utiles à tous.

La France a été précurseur de l'exploitation secondaire des données de santé à travers de nombreuses initiatives depuis plus de 20 ans : la création du Sniiram en 1999, élargi en Système National des Données de Santé (ci-après, SNDS) en 2016, et auquel l'Assurance Maladie a ouvert davantage l'accès et la mise à disposition. Puis, la loi a élargi le périmètre des données de santé partageables, notamment les données de certaines cohortes, registres et entrepôts; tout en encourageant les projets en intelligence artificielle (ci-après, IA) qui peuvent notamment recourir à l'offre de service du Health Data Hub (ci-après, HDH)[3], créé à cet effet. L'Assurance Maladie et le HDH ont ainsi largement œuvré, en lien avec les autres acteurs de l'écosystème et ses partenaires, à poser les fondations pour permettre une exploitation plus importante du patrimoine de données.

Les enjeux en lien avec l'usage secondaire des données de santé sont nombreux : réduire les délais d'accès aux données de santé, formuler la vision à moyen/long terme du patrimoine national des données de santé et les règles du jeu du partage, accompagner les initiatives locales (les entrepôts de données de santé hospitaliers ou de données de ville), ouverture des données, etc. Des défis existent également au niveau européen, avec l'espace européen des données de santé (ci-après, EHDS). Un autre de ces défis est d'embarquer la société civile, notamment en lui permettant d'être informée des projets menés avec les données de santé, et de faire valoir ses droits. Cet article propose d'expliciter ce défi, et particulièrement à travers l'angle de l'information et des droits. Il exposera ce qui est déjà fait, et ce qu'il reste à faire pour relever ce défi sur ce champ spécifiquement.

1 Pour les informations utiles au suivi médical – on parle alors d'usage primaire des données de santé. Pour les données de santé - d'origines multiples réunies, croisées ou transformées en prévision d'une analyse ultérieure, faisant appel ou non aux techniques les plus innovantes telles que l'intelligence artificielle, on parle alors d'usage secondaire des données de santé.

2 Menager K. et *Al.* (2023). *Qualitative study to assess citizens' perception of sharing health data for secondary use and recommendations on how to engage citizens in the EHDS, Deliverable 8.1*, Tehdas, en ligne: < https://tehdas.eu/tehdas1/app/uploads/2023/03/tehdas-study-to-assess-citizens-perception-of-sharing-health-data-for-secondary-use.pdf>.

3 Créé par la *loi du 24 juillet 2019 relative à l'organisation et la transformation du système de santé*, le HDH est une structure publique permettant aux acteurs d'accéder facilement à des données de santé non nominatives, hébergées sur une plateforme sécurisée. Les porteurs de projets peuvent y croiser les données et les analyser pour améliorer la qualité des soins et l'accompagnement des patients.

L'enjeu de mieux documenter les bénéfices des projets d'utilité publique

Dans le domaine de la réutilisation secondaire des données de santé, la France dispose d'atouts considérables : qu'il s'agisse de la grande base médico-administrative de l'Assurance Maladie, des vastes cohortes de recherche ou encore de la dynamique actuellement observée en matière de déploiement d'entrepôts de données de santé hospitaliers, tous constitutifs du SNDS défini dans la loi. L'exploitation de ce patrimoine de données est extrêmement prometteuse.

En effet, l'utilisation secondaire des données est source de nombreux usages qui ont aussi été mis en lumière par la crise sanitaire. Un récent article[4] les résume ainsi : « les bénéfices attendus de l'accessibilité de ces données :

1. Elles facilitent la surveillance épidémiologique des grandes pathologies, base de beaucoup de décisions en santé publique;
2. elles permettent de mesurer l'efficacité des soins et de détecter des situations locales ou régionales s'écartant significativement de la moyenne;
3. de même pour le suivi de l'évolution des pratiques dans le temps et leur impact;
4. elles permettent de documenter les effets positifs et négatifs des innovations qui sont introduites par l'évolution des pratiques, qu'elles soient médicamenteuses, techniques ou organisationnelles;
5. elles informent sur l'accès des patients aux interventions de santé : on peut calculer des taux et délais, et les croiser avec des variables géographiques, des données sociales et économiques, et ainsi identifier des inégalités, ce qui fonctionne ou non, et sur quels territoires doivent porter les efforts d'amélioration en priorité;
6. elles contribuent directement à l'amélioration des connaissances, étape indispensable pour améliorer les pratiques et concevoir des politiques publiques efficaces et protectrices;
7. elles sont indispensables pour développer les usages de l'intelligence artificielle grâce à de nouveaux algorithmes de bonne qualité, car entrainés et testés sur des populations représentatives, ce qui n'est pas toujours le cas actuellement;
8. elles permettent d'évaluer l'efficacité des campagnes de prévention ou de dépistage, et l'observance des recommandations de bonnes pratiques ».

Pour illustrer les projets qui utilisent les données de santé, prenons l'exemple de l'un de ceux que le HDH accompagne[5]. Le projet Jumeau numérique « État de santé d'un territoire » porté par le Groupe VYV et le Datalab Normandie, vise à mieux prévenir les allergies aux pollens grâce aux données de santé et environnementales. En France, 4 millions de personnes sont diagnostiquées asthmatiques et 12 millions souffrent de rhinites allergiques. Les allergies respiratoires se classent en tête des affections chroniques chez les enfants. En augmentation, ces dernières impactent le quotidien des patients : épisodes récurrents de sifflements, essoufflements, oppressions thoraciques et toux. Il s'écoule généralement sept ans entre un premier symptôme allergique et la pose d'un diagnostic : une errance diagnostique et thérapeutique qui aggrave la maladie et le risque de développer des pathologies connexes.

Ce projet combine données de santé et environnementales dans l'optique de mettre en relation les prescriptions de médicaments, les épisodes polliniques issus du Réseau National de Surveillance Aérobiologique (RNSA) (pollens par types, volumes, etc.), et les caractéristiques du territoire : plantations, arbres, cultures, météo. La base constituée est anonymisée de sorte à pouvoir rendre plus accessibles des vues dans un outil de restitution. Plus précisément, des cartes interactives en 2D et 3D sont fournies par commune, ainsi que des projections en lien avec les prévisions météorologiques.

---

4 Aymé S. et *Al.* (2023). *Bénéfices et risques de l'utilisation des données de santé à des fins de recherche : Rapport du Conseil scientifique Consultatif du Health Data Hub*, en ligne : <https://hal.science/hal-04345572v1>.

5 Le projet Jumeau Numérique a été lauréat de l'appel à projets du Datalab Normandie lancé en 2021, et soutenu par le Health Data Hub. Dans ce cadre, l'apport du Health Data Hub a consisté à fournir des données de santé anonymisées : dénombrement des patients ayant eu au moins une délivrance de médicaments relatifs aux allergènes entre 2010 et 2019 par département, sexe et âge; classement des communes de la région dans lesquelles au moins une délivrance a été dispensée.

Ainsi, le projet dévoile des données nouvelles sur la localisation et la diffusion d'allergènes, et propose des outils de visualisation et de prévention pour les problèmes de santé liés notamment à la qualité de l'air. Le démonstrateur créé a été présenté dans les villes de Caen et Rouen. Les analyses réalisées sur huit ans et les simulations prévisionnelles sont prometteuses pour améliorer la génération d'alertes polliniques pour le grand public et pour mieux intégrer ce type de facteurs dans l'aménagement du territoire.

Plusieurs centaines de projets par an sont ainsi menés avec les données du patrimoine collectif. Ces données sont accessibles à tous, dans un cadre règlementaire strict. Etablissements de soins, start-ups et industriels, associations d'usagers du système de santé, tous les acteurs peuvent et ont la légitimité d'utiliser les données à des fins d'utilité publique. Des projets qui sont le fait d'équipes de personnes aux compétences variées. Certains d'entre eux bénéficient d'un soutien du HDH, notamment en termes d'accompagnement juridique, d'expertise en matière d'utilisation de données, de mise à disposition d'une plateforme technologique à l'état de l'art pour les analyses, d'accès à des indicateurs chiffrés basiques, etc.

Pour mener et concrétiser ces projets les équipes franchissent progressivement des jalons tels que l'appariement des données, les tests d'usage des applications conçues, etc. Les usagers sont informés de ces projets surtout du fait de l'obligation de transparence qui incombe aux responsables de traitement. La transparence résulte notamment de l'application du règlement général sur la protection des données (ci-après, RGPD) qui en fait une obligation et en pose les lignes directrices dans ses articles 12, 13 et 14. Elle consiste en particulier à mettre à disposition sous une forme claire et aisément accessible des informations sur les projets de recherche et sur les droits des personnes à l'égard des données. S'agissant du SNDS, le référencement est réalisé dans un répertoire qui est accessible au public. À ce jour, les citoyens ont ainsi accès à différentes informations sur les projets qui utilisent leurs données de façon dé-identifiées, dont son cadre règlementaire de traitement de la donnée, ses objectifs et intérêts pour la santé publique, etc. Ils accèdent aux résultats, souvent sous la forme d'une publication scientifique.

Toutefois, la réalisation concrète de ces bénéfices reste assez peu étudiée et relatée. Seuls certains faits ont bénéficié d'une visibilité particulière comme, par exemple, l'utilisation des données de l'Assurance Maladie dans la crise du Médiator[6], ou plus récemment les études d'Epiphare pendant la crise sanitaire. Pourtant ces exemples où les données de santé ont joué un rôle déterminant pour le système de santé sont essentiels, entre autres parce que cela permettra aux citoyens d'être sensibilisés au rôle des données, mais aussi pour disposer d'une évaluation des bénéfices liés à l'usage de ces données permettant la prise de décision éclairée par ceux qui formulent les politiques publiques relatives aux données de santé. Des travaux seraient à mener pour identifier une série d'indicateurs permettant de décrire un bénéfice concret de l'utilisation des données de santé pour les usagers du système de santé. Il se peut que cela dépende de catégories d'usage qu'il faudra dans ce cas proposer. Typiquement, dans le cas de l'affaire Médiator, les données de santé ont permis de mettre en avant un effet secondaire et donc le bénéfice peut être approché par le nombre de personnes impactées par cette découverte. Il pourrait aussi être celui d'avoir sensibilisé les personnes à l'intérêt de l'exploitation secondaire des données de santé. L'enjeu est de mettre en lumière les bénéfices réellement observés pour le système de santé de l'utilisation secondaire des données, qui pourrait devenir un moyen d'évaluer si les bases de données sont optimalement utilisées.

6 France Info et AFP, (2023). « Affaire du Mediator : le groupe Servier condamné en appel pour escroquerie à rembourser plus de 415 millions d'euros à l'assurance-maladie et aux mutuelles », en ligne : < https://www.francetvinfo.fr/sante/affaires/affaire-mediator/affaire-du-mediator-le-groupe-servier-condamne-en-appel-pour-escroquerie-a-rembourser-415-millions-d-euros-a-l-assurance-maladie-et-aux-mutuelles-_6256377.html >.

L'enjeu de permettre la place au dialogue avec la société civile

L'effort d'information peut consister, comme précédemment mentionné, à mettre en avant des bénéfices des projets, en expliquant clairement en quoi ils répondent à un intérêt de santé publique, en publiant des résultats obtenus de ces projets (qui d'ailleurs, s'agissant du SNDS, est une condition obligatoire), de le faire selon des outils différents comme des vidéos de sensibilisation[7] et de construire des indicateurs. Mais plus globalement, il semble nécessaire de construire une culture de la donnée de santé et de la diffuser, pour permettre aux personnes de prendre pleinement conscience des bénéfices de l'utilisation des données pour la recherche et à terme, de mieux comprendre leurs droits relatifs aux données de santé et les exercer de façon éclairée. Un enjeu d'autant plus important qu'en 2022, d'après le baromètre de connaissances sur les données de santé du HDH et de France Assos Santé (ci-après, FAS), la moitié seulement des Français déclarait être au fait de leurs droits concernant leurs données personnelles, et seulement un sur cinq les connaissait en détail. De plus, 77 % des personnes interrogées estimaient être insuffisamment informées sur les projets de recherche utilisant les données de santé, et une majorité d'entre elles exprimait le désir d'en apprendre davantage notamment sur l'utilisation des données pour la recherche.

À ce propos, un nouvel outil est en cours de développement par le HDH, pour faciliter l'exercice des droits des citoyens sur leurs données de santé. Il associe un formulaire numérique pour exercer ses droits à un portail pédagogique d'information sur les données de santé et la recherche. Cet outil innovant vient répondre aux besoins d'information des acteurs de l'écosystème et de la société civile sur les données de santé. Ainsi, en 2023, grâce notamment aux retours d'un groupe de travail, une première version du portail patient et du formulaire national de gestion des droits, permettant un exercice des droits facilités pour les personnes concernées, ont été mis en production. Deux comités ont été constitués. Le comité de pilotage composé de représentants de FAS, de la Direction de la recherche, des études, de l'évaluation des statistiques (DREES), de la Commission Nationale de l'informatique et des libertés (CNIL), et du HDH, et le comité citoyen composé de représentants d'usagers du système de santé et d'associations d'étudiants se réunissent depuis janvier 2023.

Les comités se sont entendus sur l'importance de construire une information pédagogique en mettant particulièrement l'accent sur le fait que les données de santé peuvent servir à la recherche. C'est en sensibilisant le public de cette façon que les droits seront compris et que les citoyens pourront les faire valoir s'ils le souhaitent. Parmi les autres questions considérées comme prioritaires par les comités qui se sont réunis récemment se trouvent : quelles règles doivent être respectées pour l'utilisation de ces données? Quelles données sont concernées par cet outil et quels droits y sont associés? Quels projets sont actuellement en cours? Toutes ces questions seront abordées de façon accessible et pédagogique dans un portail d'information citoyen. Son objectif est de fournir toutes les réponses nécessaires pour prendre des décisions éclairées lors de l'exercice des droits ou simplement lors de la recherche d'informations. Il vise à éviter que les choix relatifs à l'exercice des droits ne se fassent par défaut ou par manque d'information.

L'utilisateur de ce portail d'information pourra être dirigé vers un espace dédié pour faire valoir ses droits. Il s'agira d'un formulaire permettant de faire une demande d'exercice des droits relatifs aux données personnelles de la part d'un citoyen. Ce formulaire sera envoyé au délégué à la protection des données (DPO) de la base concernée. Une procédure technique sécurisée permettra de faire valoir les droits tel que demandé dans le formulaire. Un email sera envoyé au demandeur pour le prévenir que cela a été réalisé.

7 Ford E. et *Al.* (2019). 'Our data, our society, our health: a vision for inclusive and transparent health data science in the United Kingdom and beyond', *Learning Health Systems,* en ligne : <https://doi.org/10.1002/lrh2.10191>.

Outre l'intérêt de permettre aux personnes de lire une information pour laquelle elles se sentiraient concernées du fait de cette éducation à la donnée de santé, la construction d'une culture de la donnée de santé, et l'introduction de plus de pédagogie, participerait très certainement à déconstruire l'image d'une appréhension, voire d'une réticence citoyenne[8] au partage qui reste ancrée chez d'autres parties prenantes[9]. Il s'agit donc d'encourager un mouvement sociétal d'acquisition d'une « culture de la donnée de santé » en accélérant la constitution d'un cadre de pensée qui donnerait du sens à l'idée même d'une exploitation secondaire des données de santé, notamment pour qu'une action éducative des données puisse avoir des effets concrets[10]. Aussi la question de l'inscription des acteurs dans le système de gouvernance des données de santé, et les types d'acteurs collectifs capables de créer une autre représentation se pose. Les usagers et leurs associations de patients sont d'excellents vecteurs pour faire connaître une solution numérique à la communauté. Grâce au développement de réseaux de communications spécifiques, ils représentent un gage supplémentaire pour donner de la visibilité à l'outil numérique. Ils encouragent d'autres usagers à l'utiliser en partageant leurs expériences positives.

Enfin, créer un dialogue autour de ce sujet, et plus globalement de l'IA, avec des millions de Français constitue une autre réponse. Elle est d'ailleurs formulée dans une des propositions du rapport de la Commission de l'intelligence artificielle[11]. Il y est indiqué que nous ne pourrons pas avancer collectivement et construire un système solidaire si nous n'avons pas un projet commun, une narration commune — un projet de progrès collectif; et qu'il est donc nécessaire de débattre. Cette ouverture et ce dialogue peuvent prendre des formes diverses dont des études[12] (sur les perceptions, les besoins, les attentes, les connaissances des citoyens vis-à-vis des données de santé), des ateliers participatifs et de codesign (selon des publics cibles ou plus larges pour sensibiliser aux données de santé, pour former, et porter les engagements citoyens vis-à-vis des données), des consultations et des concertations. Peuvent aussi être mis en place des ateliers de vulgarisation scientifique, et des initiatives de transparence, et d'ouverture des données, des cafés citoyens…

Conclusion

La question de l'information des citoyens de l'usage de leurs données, notamment à des fins de recherche, est multidisciplinaire, car juristes, communicants, responsables de structures de soins et d'entrepôts, associations, professionnels recueillant les données s'en emparent. L'un des enjeux en ce qui concerne l'exploitation secondaire des données de santé est de permettre à la société civile de prendre une part active dans la compréhension des données. Et le défi est bel et bien celui d'une information pédagogique et participative; un défi qui concerne toutes les structures amenées à utiliser ou mettre à disposition ces données. Deux pistes semblent y répondre : la piste qui consiste à trouver la juste articulation entre l'information légale et pédagogique; et la piste participative qui consiste à renforcer le dialogue.

8 État d'esprit Statis et Health Data Hub (2023). *Conférence de consensus : Comment sensibiliser la société au partage des données de santé, Rapport d'étude*, 2023, en ligne : < https://www.health-data-hub.fr/actualites/conference-de-consensus-les-propositions-citoyennes-pour-parler-du-benefice-des-donnees>.

9 Neves A. L. et al. (2019). 'Health Care Professionals' Perspectives on the Secondary Use of Health Records to Improve Quality and Safety of Care in England: Qualitative Study', *Journal of Medical Internet Research* 21(9), e14135, en ligne: <https://www.jmir.org/2019/9/e14135>.

10 Combes S. et Guillot C. (2021). « Utilisation secondaire des données de santé : quelles modalités de participation citoyenne », Third, n°7, en ligne : < https://third.digital/numero-sept/utilisation-secondaire-des-donnees-de-sante-quelles-modalites-de-participation-citoyenne/>.

11 Commission de l'intelligence artificielle (2024). *IA, notre ambition pour la France*, en ligne : <https://www.info.gouv.fr/upload/media/content/0001/09/4d3cc456dd2f5b9d79ee75feea63b47f10d75158.pdf>.

12 Guillot C., SCHWEIGHOFER A. et LUCAS J. (2024) « Comment parler des données de santé aux citoyens? Le défi de l'information créative et participative des structures de santé » dans Bubien Y. et Bataille-Ambert A., dir, *Innovations et communication en santé*, Éditions LEH

obvia

obvia.ca