

LES DONNÉES DE SANTÉ

DE QUOI PARLE-T-ON ?

"La donnée constitue le cœur de l'exercice de la médecine dans son rapport avec le patient. Il n'est donc pas surprenant, dans le cadre d'une numérisation globale de l'ensemble de la pratique médicale et de l'utilisation de capteurs de plus en plus nombreux et variés, d'assister à une augmentation exponentielle de la quantité et de la diversité des données disponibles".¹

En France, les données de santé proviennent de multiples sources : les bases de données médico-administratives comme par exemple le Sniiram² et ses 8,9 milliards de feuilles de soins, les images des 80 millions d'actes d'imagerie effectués chaque année, les cohortes et registres, les dossiers médicaux, les essais cliniques, les données patients collectées via les smartphones, les réseaux sociaux et les sites internet...

Elles sont donc disparates et d'une grande diversité de formats, car toutes ces données ont été recueillies pour un usage bien spécifique : diagnostiquer une maladie, détecter une mutation particulière dans le génome, rembourser des soins, mesurer son activité physique... L'intelligence artificielle (IA) vient bouleverser la donne en cherchant à utiliser toutes ces data dans l'objectif de faire progresser la recherche, les soins et l'innovation en santé.

Comment ? en annotant et en appariant les données pour obtenir des résultats plus fiables et de meilleure qualité, mais aussi pour faire surgir des hypothèses et des liens qui n'étaient pas envisagés. A une condition : recueillir un nombre suffisant de données exploitables pour faire tourner les algorithmes de l'IA.

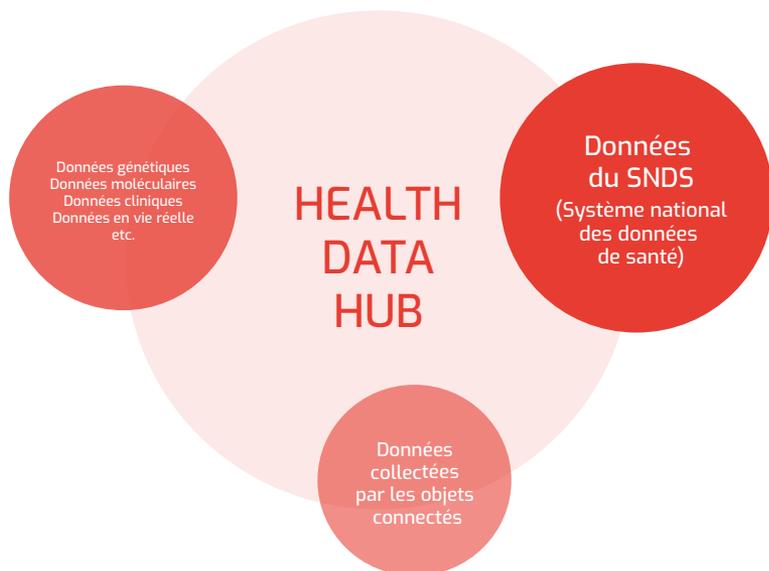
Quand on sait qu'il faut près de 100 000 images pour que les algorithmes de l'IA puissent apprendre à détecter un mélanome et poser un diagnostic sûr, on comprend qu'un hôpital seul ne peut collecter le nombre de données nécessaires ; les producteurs de données doivent donc s'associer pour collecter, échanger, partager leurs données. Il faut aussi s'assurer de pouvoir disposer de données "propres", rangées et bien étiquetées.

CE QUI SE PROFILE D'ICI 2030

- Les systèmes dotés d'intelligence artificielle, de techniques de machine learning et de puissances de calcul importantes permettront de synthétiser et de modéliser des données complexes pour affiner un diagnostic, identifier les mutations génétiques en cause, surveiller la croissance tumorale, aider à la prise de décision des médecins...
- Dans le champ de la cancérologie, l'utilisation de données "intelligentes" permettra d'avancer dans la connaissance du cancer, en intégrant les mutations du génome, les relations hôte/environnement, la capacité d'adaptation du vivant, décomposant le cancer en sous-ensembles de maladies et faisant presque du cancer un ensemble de maladies orphelines.
- On se dirige vers une redivision de la classification des maladies et l'émergence de nouvelles hypothèses de recherche.
- Des données bien exploitées vont de plus permettre d'aller toujours plus loin dans les soins donnés aux patients. Car, *"si les traitements sont bien codifiés au début de la maladie, plus on avance dans la maladie en utilisant les différentes lignes thérapeutiques disponibles, et moins on dispose de recommandations fondées sur l'evidence based-medicine et moins on dispose de cas identiques basés sur l'expérience".³*
- L'IA autorisera une nouvelle modélisation, une modélisation multi-échelle dans le domaine de la santé qui *"va consister à intégrer des données très hétérogènes ouvrant la voie à des approches révolutionnaires".⁴*

CE QUI EST EN COURS

- La création du Health Data Hub, début 2019, un dispositif de partage entre producteurs et utilisateurs de données de santé



1. Le Health Data Hub est un guichet d'accès unique à l'intégralité des données de santé soutenues par la solidarité nationale.

2. Le Health Data Hub collecte les données de santé. Il consolide le patrimoine des données de santé déjà disponibles.

3. Le Health Data Hub propose des outils pour créer des algorithmes d'analyse :

- des outils en open source (R, Milk, Python...)
- des outils en accès privé
- des outils de R&D

4. Le Health Data Hub permet d'apparier et d'analyser des données. Il met à disposition des jeux de données appariées et documentées.



LES DONNÉES DE SANTÉ

• La création d'une chaîne de valeur autour de la donnée et son interprétation

RECHERCHE



- Optimisation de la génération de leads (pistes de recherche)
- Plateformes collaboratives (innovation, financement)
- Jumeaux numériques humains



DÉVELOPPEMENT



- Digitalisation des processus de gestion des essais cliniques
- Essais cliniques adaptatifs et évolutifs



ACCÈS AU MARCHÉ



- Transformation du modèle de démonstration de la valeur médicale
- Démonstration de la valeur médicale en vie réelle



PRODUCTION



- Automatisation des lignes de production
- Maintenance prédictive
- Automatisation du contrôle qualité



DISTRIBUTION



- Gestion des stocks par réalité augmentée
- Nouveaux modèles de technologies du transport médical
- Système Datamatrix pour combattre la contrefaçon médicamenteuse



USAGES



- Transformation de l'offre
- Transformation des business models

INTÉGRATION VERTICALE

CE QU'IL FAUT DÉPASSER

- **Le bouleversement du modèle de l'imagerie médicale**

*"Les algorithmes déjà très présents dans la construction des images médicales sont aujourd'hui suffisamment puissants pour guider l'analyse des images médicales, aussi bien, voire mieux que des experts humains."*¹

La société Therapixel, par exemple, a réussi à entraîner des algorithmes à "apprendre" les 640 000 mammographies rassemblées pour un concours mondial, afin de lui faire distinguer les mammographies suspectes des normales.

Il est aussi possible de construire des modèles numériques (jumeau numérique) pour synthétiser des images médicales et pouvoir les transférer en vie réelle.

- **Le bouleversement à venir du modèle de recherche clinique**

L'IA va accélérer considérablement le développement clinique en permettant de tester de nouveaux médicaments sur des populations sélectionnées.

A terme, il est possible de penser que la phase III laissera la place à des études en vie réelle sur des patients sélectionnés en fonction de leur génétique, de leur phénotype, de leur profil génomique... Les coûts de la recherche clinique devraient donc baisser. Le modèle économique de la R&D va donc évoluer en retour.

- Résoudre les problèmes de partage de valeur. Si la donnée appartient juridiquement au malade, de fait, elle ne lui "appartient pas". Anonymisée, sécurisée, non divulguée, elle sert la communauté des patients et les avancées de la recherche. Ne pas pouvoir les utiliser serait synonyme d'une perte de chance pour les autres patients.

- Ne pas interpréter de façon trop rigide le Règlement général sur la protection des données (RGPD), qui multiplie les courriers, déclarations aux patients et ralentit le processus de collecte de données intelligentes.

- Relever les défis organisationnels afin de penser en amont la collecte et la capitalisation des données.

¹ Ibid.cité. p. 151.