



# Dossier de presse

## Clinique des données

*Octobre 2025*

## LA CLINIQUE DES DONNÉES DU CHU DE NANTES

### Le CHU de Nantes, premier CHU de France à avoir mis en place un véritable entrepôt de données en santé dès 2018. Retour sur un bouleversement historique de l'usage des données en santé.

#### L'édito de Pierre-Antoine Gourraud,

Fondateur de la Clinique de données, préfigureur du Comité éthique et Scientifique des entrepôts de données de santé du CHU de Nantes.

Et si les données de santé devenaient nos meilleures alliées pour mieux soigner demain ? Depuis quelques années, les hôpitaux et les chercheurs disposent d'une ressource souvent méconnue du grand public : les données issues des soins réels, collectées chaque jour dans les établissements de santé. Ces données (anonymisées et protégées,) constituent une source précieuse d'informations pour faire avancer la recherche et, in fine, améliorer la prise en charge des patients. Ces données issues du soin sont comme des partitions de musique que l'on pourrait rejouer à l'infini.



Dans notre système de santé, chaque traitement, chaque imagerie, chaque diagnostic produit une trace. Lorsque nous les rassemblons et les retraitions pour les rendre parlantes, ces informations racontent une histoire bien plus grande que celle d'un seul individu : elles permettent de mieux comprendre l'évolution des maladies, la manière de les prendre en charge, d'identifier plus rapidement les effets secondaires rares d'un médicament, ou encore de repérer des profils de patients qui répondent mieux à tel ou tel traitement. C'est précisément ce que permet le partage massif et encadré des données entre établissements de santé.

Ce partage est aujourd'hui essentiel pour construire une médecine réellement personnalisée. Car pour adapter un traitement à chaque patient, encore faut-il comprendre les différences qui existent entre les individus — dans leur parcours de soins, leur génétique, leurs réactions aux traitements. Plus les bases de données sont vastes, plus les modèles que nous développons sont précis.

Bien entendu, cette révolution ne peut se faire sans garanties. Les données utilisées dans nos recherches sont sécurisées et strictement encadrées.

Aujourd'hui, en tant que chercheur, je suis convaincu d'une chose : plus nous partageons intelligemment les données, plus nous donnons une chance à chaque patient de bénéficier d'un soin plus juste, plus rapide et plus adapté à sa situation. Les données ne remplacent pas le médecin — elles l'éclairent. Et grâce à elles, c'est toute la chaîne du soin qui peut progresser, pour que demain, chacun soit mieux soigné, c'est ce que j'appelle la « solidaritarité ».

**Le contexte de la création de la clinique des données** : depuis le déploiement du Dossier Patient Informatisé (DPI) en mai 2015 dans les unités de soins, l'informatisation des données de santé s'est accélérée. À l'ère du tout numérique, l'accès à ces données ouvre un champ de recherche particulièrement prometteur. Dans un contexte marqué par le Big Data, les objets connectés, la médecine personnalisée, les premiers projets d'intelligence artificielle et l'ouverture nationale des bases de données de la CPAM, le CHU de Nantes a saisi l'opportunité, via le GIRCI (Groupements interrégionaux pour la recherche clinique et l'innovation), de mettre en place un entrepôt de données à vocation de recherche, grâce au logiciel « eHOP » développé par les professionnels de Rennes.

Dans ce paysage, un besoin s'est progressivement affirmé : celui de développer les compétences en statistiques et en mathématiques cliniques afin de **mieux accompagner les porteurs de projets de recherche**, tant en amont (par exemple pour le calcul du nombre de sujets nécessaires) qu'en aval (réalisation des analyses statistiques).

Jusqu'au milieu des années 2010, la plupart des professionnels étaient formés dans un environnement où les données de santé étaient rares, coûteuses à collecter et difficilement accessibles. Cette réalité limitait naturellement les ambitions en matière d'analyse quantitative. C'est dans ce contexte que les premiers profils biostatisticiens ou épidémiologistes ont été recrutés, sans toujours bénéficier d'un écosystème favorable à l'exploitation de données à grande échelle.

**Un changement de paradigme s'est opéré à partir du moment où les entrepôts de données en santé (EDS) ont commencé à se développer. En l'espace de quelques années, l'arrivée de données massives a profondément bouleversé les pratiques.** Les volumes disponibles sont passés de quelques centaines à plusieurs millions de patients, modifiant radicalement l'échelle des analyses possibles.

Ce renversement a entraîné un repositionnement stratégique majeur : le CHU de Nantes, jusqu'alors producteur ponctuel de données, est devenu **une véritable source de données de santé à fort potentiel de valorisation. L'objectif n'est plus seulement de produire de nouvelles données pour chaque projet, mais aussi — et surtout — de tirer parti de celles déjà collectées, en développant des usages secondaires.**



Équipe - Clinique des données

C'est dans cette dynamique que s'inscrit la Clinique des Données, conçue comme un espace de rencontre entre les données disponibles, les besoins méthodologiques, et les ambitions scientifiques et cliniques.

## Repères sur la clinique des données

### A quoi ça sert ?

La clinique des données, par l'intermédiaire de l'entrepôt de données en santé qui stocke les données, permet de **mieux repérer des besoins et des caractéristiques particulières pour l'ensemble des patients**. Ces données permettent, in fine, de déployer des projets de recherche ou encore de développer la médecine personnalisée.

### Comment ?

En mettant en place une équipe d'experts et de nouveaux métiers comme le traitement du langage (« *data analyst* » par exemple) au service de tous ceux qui se posent des questions auxquelles le CHU de Nantes pourrait répondre (partenaires publics ou privés).

### La clinique des données en quelques dates clés

2015 : arrivée du Pr Gourraud au CHU de Nantes – Lancement du projet

2016 : choix de donner accès aux documents textuels de l'hôpital

2018 : autorisation du déploiement par la CNIL

2019 : premier recrutement d'une compétence en traitement automatique du langage

2021 : 2M de données analysées – [Publication de l'algorithme INSEE-HOP actant la valeur de la transformation des données issues du soin repris au sein du Ouest data Hub](#)

2022 : projet malades\* – grand modèle souverain en santé – [Publication du modèle de langue Bio-Mistral](#)

2023 : les données de 3M de patients sont analysables – Publication d'une méthode de substitution des données réelles par des données simulées complètement non réidentifiantes.

2024 – 2025 : réorganisation avec recrutement d'un directeur opérationnel – déploiement des collaborations avec des partenaires privés\*\*

\* Le projet Malades vise à intégrer et adapter les grands modèles de langue pour des applications biomédicales en développant des outils de traitement automatique du langage adaptés au domaine médical français. Ce projet coordonné par Nantes Université bénéficie d'une aide financière de l'État gérée par l'ANR dans le cadre du Programme Intelligence Artificielle – AAP TSIA – Giga-modèles Edition 2023

\*\* Exemple d'un partenariat mis en place en 2025 entre le CHU de Nantes et HEVA – La Poste Santé & Autonomie pour simplifier et accélérer les projets de recherche sur données. Ce partenariat s'inscrit dans le cadre de l'offre innovation du CHU de Nantes : La Fabrique de l'innovation en santé®. Concrètement :

- Le CHU de Nantes facilite à HEVA la mise à disposition des données de son catalogue, tout en s'assurant du respect de la réglementation en vigueur. Ces données sont issues du soin, via l'entrepôt de données de santé du CHU de Nantes, ou de la recherche.

- Les équipes du CHU de Nantes apportent leurs expertises pour évaluer, dans des délais optimisés, la faisabilité technique et réglementaire des demandes d'études sur données. HEVA permet de recourir à son expertise reconnue notamment en analyse et appariement de données.

## Quatre cas d'usage de l'entrepôt de données de santé du CHU de Nantes

### 1. Recenser le nombre de codes rouges en maternité grâce aux données de santé

Une césarienne est réalisée en « code rouge » si le pronostic vital du fœtus et/ou de la mère est immédiatement engagé. Afin de suivre le nombre de césariennes « codes rouges » réalisées au CHU de Nantes au cours de l'année 2018, les rapports cliniques stockés dans l'entrepôt de données de santé se sont révélés être une ressource indispensable. En analysant ces rapports grâce au traitement automatique du langage naturel\*, 934 enregistrements correspondant au nombre de césariennes non programmées ont ainsi pu être identifiés. Parmi tous les rapports cliniques de césariennes non planifiées recensés, 59 codes rouges ont été dénombrés. Ainsi, l'entrepôt de données de santé du CHU de Nantes et les outils de traitement automatique du langage naturel ont été utiles pour générer automatiquement un indicateur de performance de l'unité obstétrique.

\* Domaine multidisciplinaire impliquant la linguistique, l'informatique et l'intelligence artificielle, qui vise à créer des outils de traitement de textes et de la parole pour diverses applications.

### 2. Anévrisme intracrânien, les données de santé une ressource essentielle pour identifier des patients « cas témoins »

L'anévrisme intracrânien désigne une dilatation localisée d'une artère au niveau du cerveau, qui prend la forme d'une poche de sang dans laquelle le sang circule sous pression. Celle-ci peut se rompre dans environ 1% des cas, conduisant dans un tiers cas au décès du patient ou pour un autre tiers provoquer d'importantes séquelles. La prédiction du risque d'anévrisme intracrânien est donc considérée comme un enjeu de santé publique majeur. Dans le cadre du projet de recherche ICAN\*, coordonné par le CHU de Nantes, les données issues de l'entrepôt de données de santé du CHU ont été utilisées pour identifier des patients non porteurs d'anévrisme intracrânien afin de constituer le groupe des patients « cas témoins ». L'utilisation des données issues de l'entrepôt a ainsi permis d'obtenir rapidement une liste de patients répondant aux critères généraux d'inclusion dans l'étude.

\*Publié dans la revue scientifique [Neurosurgery](#)

### **3. Facteurs de récurrence de l'hémangiome infantile : les data, un train d'union entre les établissements de santé pour répondre ensemble aux questions scientifiques**

Grâce au Ouest Data Hub, plateforme interrégionale des données de santé d'HUGO (groupement des Hôpitaux Universitaires du Grand Ouest), près de 130 millions de documents et 1,2 milliards de données structurées sont disponibles afin de répondre à des questions scientifiques.

Cette ressource interrégionale a notamment permis d'identifier les facteurs de récurrence de l'hémangiome infantile (tumeur vasculaire bénigne). Les données de près de 225 enfants, atteints d'hémangiome et pris en charge entre juin 2014 et décembre 2021 au sein des établissements du groupement de coopération sanitaire HUGO, ont pu être ainsi analysées. Les résultats de l'étude menée sur les données disponibles mettent notamment en évidence qu'un hémangiome infantile profond est un facteur de risque de la récurrence précoce.

Les résultats de cette étude ont été publiés dans la revue Annales de Dermatologie et de Vénérologie.

### **4. Intelligence artificielle générative**

Le CHU de Nantes et le Laboratoire des sciences du numérique de Nantes (LS2N) ont lancé en octobre 2024 un appel à idées innovantes destiné à identifier auprès des hospitaliers des cas d'usages de l'utilisation de modèles d'IA générative à l'hôpital. Ces modèles d'IA se baseront sur les données disponibles au sein de l'entrepôt de données de santé du CHU. Après étude des 50 projets proposés par les professionnels du CHU, 4 cas d'usage ont été retenus. Ils concernent notamment l'amélioration de la réponse aux patients, l'aide à la rédaction et au diagnostic, ou encore l'optimisation du codage des actes médicaux. Ces projets sont actuellement en cours d'accompagnement.

## **Quelles perspectives pour la donnée en santé dans les établissements hospitaliers ?**

### **Une place croissante dans le pilotage de l'établissement, au bénéfice des patients**

**Les données de santé occupent aujourd'hui une place de plus en plus centrale dans la stratégie de pilotage des établissements hospitaliers.** Elles permettent d'objectiver les pratiques, d'identifier des marges de progrès, et d'évaluer l'impact réel des actions engagées. Grâce à une meilleure connaissance des parcours de soins et des besoins spécifiques des patients, les décisions prises au niveau institutionnel deviennent plus éclairées, plus réactives et mieux adaptées aux besoins des patients. Loin d'être un simple outil administratif, **la donnée devient ainsi un levier direct d'amélioration de la qualité des soins.**

### **Poursuivre et intensifier le développement des usages**

L'enjeu actuel ne réside plus seulement dans la collecte des données, mais bien dans leur valorisation. Cela implique de diversifier les usages : **soutien à la recherche clinique, aide à la décision médicale, optimisation des flux logistiques, anticipation des ruptures de parcours ou encore détection précoce des risques.**

### **Favoriser le développement de synergies inter-établissements**

Les enjeux liés à la donnée ne s'arrêtent pas aux frontières d'un établissement. **La mutualisation des ressources, des compétences et des jeux de données entre hôpitaux,** centres de santé et structures médico-sociales permet d'enrichir les cohortes et le potentiel des recherches qui peuvent être menées. En favorisant les collaborations inter-établissements, il devient possible de mener des projets de recherche à grande échelle, de mieux coordonner les parcours patients sur un territoire, et d'accélérer l'innovation dans les pratiques de soins de façon plus simple.

### **Animer la circulation des données à l'échelle d'un territoire, au service de tous**

Le véritable défi à venir est d'organiser, dans un cadre **éthique et sécurisé,** une **circulation fluide des données de santé à l'échelle des territoires.** Cela suppose non seulement des infrastructures techniques adaptées, mais aussi une gouvernance partagée et une culture commune autour de la donnée. L'objectif est clair : **permettre à chaque acteur du système de santé – du médecin de ville à l'hôpital, du chercheur au décideur – d'accéder à des informations pertinentes, à jour, et utiles pour mieux répondre aux besoins des populations locales.**

Ces perspectives s'illustrent par exemple dans le projet Synopse coordonné par Nantes Métropole. La crise du Covid 19 a fait émerger le besoin de créer un outil pour faciliter la mise en réseau et le partage de données de santé environnementale à l'échelle du territoire. Débuté en juillet 2024, le projet Synopse vise ainsi à mieux partager la connaissance du territoire (plateforme partenariale), organiser et outiller la gestion de crise, ainsi que proposer des services pour aider les citoyens à être acteurs de leur santé. Le CHU de Nantes, partenaire de ce projet, est particulièrement impliqué dans le développement d'un outil de veille et de gestion de crise afin de suivre l'évolution des indicateurs et prévoir l'évolution des prises en charges à l'hôpital. Référence à la publication en français : [en savoir plus](#)



## **Contact presse**

Zakaria Gambert  
zakaria.gambert@chu-nantes.fr  
07 77 25 95 47