

**État d'avancement  
et premières réalisations  
prometteuses des  
projets accompagnés  
par le Health Data Hub**

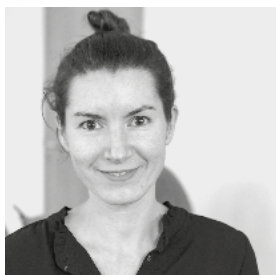


JANVIER 2024

# Édito



**Gérard Raymond**  
Vice-Président  
du Health Data Hub

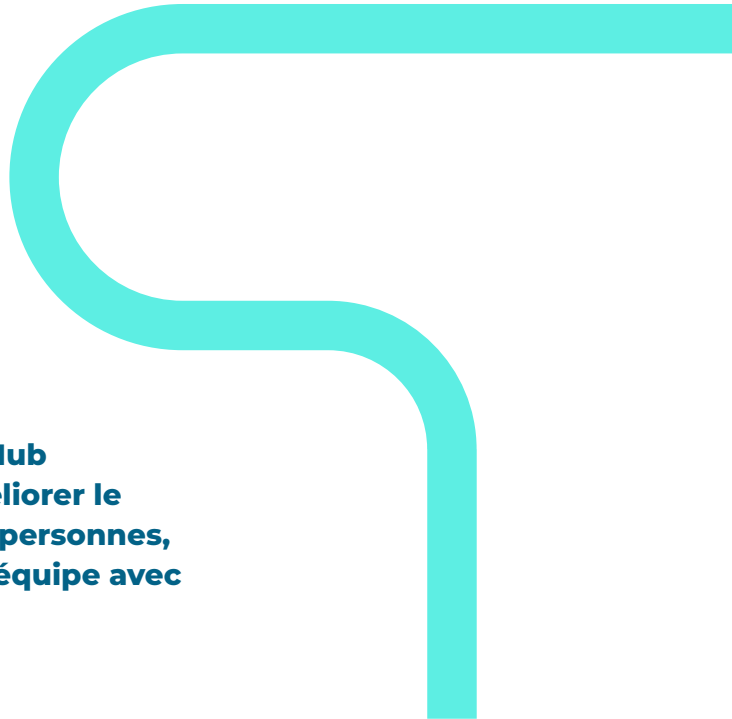


**Stéphanie Combes**  
Directrice  
du Health Data Hub

Le numérique – et en particulier la génération, circulation et l'exploitation de la donnée – transforme aujourd'hui très profondément la santé publique et environnementale. Avec des innovations très diverses, dont la mise en place de plateformes numériques publiques permettant de stocker, selon des modalités variées, il est possible de partager toutes sortes de données de santé. Il peut s'agir des informations utiles au suivi médical – on parle alors d'usage primaire des données de santé –, ou de 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 –. Le Health Data Hub a été créé pour cela, comme une plateforme de données de santé permettant ces usages de manière facilitée.

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 Système national des données de santé (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. Les données sont un facteur majeur d'accélération de la recherche médicale, à l'aide d'algorithmes d'intelligence artificielle qui peuvent aider à croiser les informations des patients atteints de maladies rares, par exemple. Le croisement de données sert également à créer de nouveaux outils pour les professionnels de santé pour déterminer la meilleure prise en charge, que ce soit lors du dépistage, de la pose du diagnostic, de l'interprétation de résultats d'imagerie médicale ou pour les prescriptions médicamenteuses. Elles aident, enfin, à l'amélioration continue du système de santé, par le développement de l'utilisation des données en vie réelle dans l'évaluation des produits de santé mis sur le marché (médicaments et dispositifs médicaux), dans la surveillance d'une crise sanitaire ou encore dans le suivi et l'évaluation de l'efficacité des politiques publiques sanitaires, notamment en terme de prévention. Par ailleurs, l'utilisation secondaire des données de santé constitue une opportunité économique importante.



**« Les projets que le Health Data Hub accompagne contribuent à améliorer le système de soin et la santé des personnes, et leur succès tient au travail d'équipe avec les porteurs de projets. »**

Nous souhaitons donner à voir, à travers l'exposition de quelques projets, le champ des possibles des bénéfices de l'utilisation de certaines bases de données de santé, souvent chaînées à celles de l'Assurance Maladie. C'est une occasion, aussi, de sensibiliser le lecteur sur tout ce qui se joue autour de la réutilisation des données de santé en termes d'ouverture aux acteurs, de créativité et d'innovations dont sont capables les équipes pour utiliser les données et les chaîner, de jalons progressifs à franchir pour concrétiser ces projets, etc. En effet, ces données sont accessibles à tous, dans un cadre réglementaire strict, comme les projets exposés dans ce rapport en témoignent. Établissements de soins, start-ups et industriels, associations d'utilisateurs 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.

Ils peuvent, c'est le cas de ceux narrés ici, bénéficier d'un soutien du Health Data Hub, 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. Enfin, ce document permet de sensibiliser le lecteur au temps de la recherche et aux jalons que les équipes qui les portent doivent franchir... Donner à voir que les bénéfices des projets ne se limitent pas aux résultats finaux, car il faut aussi intégrer le temps de la recherche, le temps de la mise au point de ces projets comprenant des phases telles que l'appariement des données, les tests d'usage des applications conçues, etc. Ces succès dans les avancées des projets valent la peine d'être exposés ; d'où l'objet de ce rapport.



# Sommaire

ÉDITO	2
ORDEI	4
BACTHUB	5
DEEPSARC	7
JUMEAU NUMÉRIQUE “ÉTAT DE SANTÉ D’UN TERRITOIRE”	9
HYDRO	11
TARPON	13
INNERVE	15
NHANCE	17
DEEP.PISTE	19
REXETRIS	20
E-CORE	22
EHDEN	24
PROJET MICI - AFA CROHN RCH FRANCE	26
PROJET MALADIES NEUROMUSCULAIRES - AFM TÉLÉTHON	28
PROJET VASCULARITES - FRANCE VASCULARITES	29
LE DATA CHALLENGE TISSUENET	30
LE DATA CHALLENGE VISIOMEL	32





## ORDEI

**CLIQUEZ ICI !** En savoir plus sur les effets indésirables des médicaments et les ruptures de stock : [data.ansm](https://data.ansm.fr)

### PARTENAIRE

**Agence nationale de sécurité du médicament et des produits de santé (ANSM)** est l'acteur public qui permet, au nom de l'État, l'accès aux produits de santé en France et sécurise leur usage tout au long de leur cycle de vie. Elle s'assure que les produits de santé disponibles en France sont fiables, efficaces, accessibles et bien utilisés.

Les effets indésirables, ou effets secondaires, sont des réactions nocives et non voulues, susceptibles de se produire à la suite de la prise d'un médicament. Ces effets peuvent être des nausées, des maux de tête, des démangeaisons, des éruptions cutanées, des diarrhées, une vision floue, de la fatigue, etc. Les professionnels de santé, les patients ou associations de patients et les laboratoires pharmaceutiques peuvent signaler aux autorités sanitaires tout événement sur le site [www.signalement-sante.gouv.fr](https://www.signalement-sante.gouv.fr).

Afin de rendre ces informations accessibles à tous, l'ANSM a entrepris une démarche d'ouverture de ces données.

Le projet ORDEI (Outil d'Information Des Effets Indésirables), porté par l'ANSM, vise à développer un outil capable de restituer et visualiser, de manière compréhensible, des indicateurs portant sur les effets indésirables des médicaments déclarés auprès des autorités sanitaires, tout en respectant l'anonymat des patients.

Trois bases de données ont été utilisées pour cet outil : la Base Nationale de Pharmacovigilance de l'ANSM et des produits de santé (base BNPV) ; la base Open Medic contenant les informations sur le remboursement des médicaments (données anonymes issues de la base principale du SNDS) ; ainsi que la base de l'ANSM contenant les informations sur les auto-risations de mise sur le marché des médicaments (base CODEX).

Le projet ORDEI a été lauréat du premier appel à projets du Health Data Hub, lancé en 2019. Dans ce cadre, l'apport du Health Data Hub a consisté à la mise à disposition d'expertises pendant plus d'un an, notamment pour contribuer à la préparation et au traitement des données, ainsi qu'au développement de l'outil de restitution des indicateurs. À

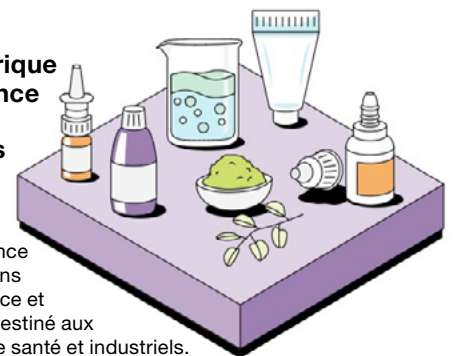
partir de la base CODEX, les informations par médicament dans les deux autres bases de données ont, par exemple, été standardisées et recoupées. Cette étape a permis de consolider les dénombrements agrégés du nombre de consommateurs et des effets indésirables par substance active, classe d'âge et sexe, ainsi que par type d'effets indésirables recensés. Un programme d'extraction de la base principale du SNDS permettant de calculer le nombre de consommateurs par substance active de manière plus précise a également été conçu. Il sera prochainement intégré dans l'outil.

La plateforme [data.ansm](https://data.ansm.fr) est disponible en accès libre depuis avril 2023. Elle permet de consulter des informations et des données chiffrées sur l'historique des déclarations d'effets indésirables de médicaments, des erreurs médicamenteuses et des ruptures de stocks de médicaments depuis 2014.

[data.ansm](https://data.ansm.fr)

### Accès public à l'historique des données de l'Agence nationale de sécurité du médicament et des produits de santé

[data.ansm](https://data.ansm.fr) est un site d'information et de transparence sur l'historique des déclarations relatives à la pharmacovigilance et aux ruptures de stock. Il est destiné aux particuliers, professionnels de santé et industriels.



Ce site n'est pas un site de consultation et d'information médicales, demandez conseil à votre médecin ou pharmacien.

# 600 000

effets indésirables déclarés sur la période 2014-2022



## BACTHUB

Mieux comprendre la résistance aux antibiotiques

### PARTENAIRES

**L'Assistance Publique-Hôpitaux de Paris (AP-HP)** est un établissement de santé public et un centre hospitalier universitaire (CHU) d'Île-de-France. Elle regroupe 38 hôpitaux, situés à Paris, en petite couronne et en province.

**L'Institut national de la santé et de la recherche médicale (Inserm)** est un organisme public de recherche français entièrement dédié à la santé humaine. Son objectif est d'améliorer la santé de tous par le progrès des connaissances sur le vivant et sur les maladies, l'innovation dans les traitements et la recherche en santé publique.

La résistance des bactéries aux antibiotiques, aussi appelée antibiorésistance, représente un enjeu majeur de santé publique. En effet, une étude menée en 2019 estime que 5 millions de décès dans le monde sont associés à l'antibiorésistance. L'usage inapproprié des antibiotiques contribue au développement et à la dissémination de bactéries résistantes aux traitements. Cette résistance se manifeste lorsqu'une bactérie se transforme et développe des mécanismes de défense, diminuant ou annulant l'action des antibiotiques qui la combattent.

Les Français sont parmi les plus grands consommateurs d'antibiotiques en Europe, selon [l'European Centre for Disease Prevention and Control](#). Pour contrôler l'augmentation des bactéries résistantes, le Gouvernement français a mis en place des stratégies de santé publique, parmi lesquelles la lutte contre le mésusage et la surconsommation d'antibiotiques. En lien avec ces politiques, le projet BACTHUB vise à clarifier le rôle des antibiotiques dans l'acquisition d'une infection à bactérie résistante aux antibiotiques.

Le projet BACTHUB, porté par l'Inserm et l'AP-HP, vise à étudier le lien entre la prise d'antibiotiques et le développement d'une bactériémie (infection dans le sang) à bactérie résistante aux antibiotiques. Les résultats obtenus pourront guider les stratégies de lutte contre l'antibiorésistance, mais aussi aider les soignants à bien choisir les antibiotiques qu'ils prescrivent aux malades.

Pour mener ce projet, deux grandes sources de données sont mobilisées : les données de l'entrepôt de données de santé de l'AP-HP et celles de la base principale du SNDS.

La première étape du projet, c'est-à-dire la constitution de la base de données hospitalière propre à ce projet, a été franchie. Il s'agit de la première base en France, et l'une des rares à l'échelle internationale, comportant des données aussi détaillées pour un grand nombre de patients hospitalisés avec une bactériémie. En effet, elle inclut à la fois des données sur les patients, sur les séjours hospitaliers, sur le(s) bactérie(s) responsable(s) de l'infection et sur leur statut vis-à-vis de la résistance aux antibiotiques.

Le projet BACTHUB est un projet institutionnel lancé en 2019. Dans ce cadre, l'apport du Health Data Hub a consisté à mettre à disposition deux data engineers pendant plus d'un an. Elles ont été notamment chargées du nettoyage, de l'évaluation et de l'amélioration de la qualité des données, de la structuration de la base, et de sa documentation. Cette base, qui comprend des données cliniques, pharmacologiques, biologiques, microbiologiques et socio-démographiques, est appariée aux données de la base principale du SNDS, grâce à l'appui de l'Assurance Maladie.

---

# 670 000

**infections à bactéries résistantes  
aux antibiotiques en Europe en 2015**

---



— L'analyse de la base de données hospitalière créée en première étape a permis de spécifier les principales caractéristiques de près de 30 000 patients, de leur séjour hospitalier, de leur(s) bactérie(s) et de leur lieu d'acquisition (en ville ou à l'hôpital). À titre d'exemple, on y apprend que parmi ces patients, 41 % avaient acquis leur bactériémie en ville et 59 % à l'hôpital. Les patients avec une bactériémie présentaient de forts taux de mortalité, atteignant 20 % à trois mois de l'infection dans le cas d'une bactérie acquise en ville et 26 % à trois mois de l'infection dans le cas d'une bactérie acquise à l'hôpital. Le taux de mortalité était 1,5 à 2 fois plus faible pour les patients sans passage en soins intensifs par rapport à ceux avec passage en soins intensifs, selon le cadre d'acquisition de l'infection (ville ou hôpital). Ces premiers résultats ont fait l'objet d'une publication scientifique : [Abbara, S., Guillemot, D., El Oualydy, S., Kos, M., Poret, C., Breant, S., ... & Watier, L. \(2022\). Antimicrobial resistance and mortality in hospitalized patients with bacteremia in the greater Paris area from 2016 to 2019. \*Clinical Epidemiology\*, 1547-1560.](#)

Grâce au chaînage de cette base de données à la base principale du SNDS, le parcours de soins de ces patients devient quasi-exhaustif. En effet, l'ensemble des hospitalisations, qu'elles soient à l'AP-HP, ou non, peut être connu. Par ailleurs, l'ensemble des soins remboursés en ville est renseigné, en particulier les antibiotiques délivrés aux patients en ville avant leur hospitalisation avec bactériémie. Cette information permet d'étudier avec précision le lien entre la prise d'antibiotiques et le risque de faire une bactériémie à bactérie résistante aux antibiotiques. Ainsi, dans la deuxième phase de ce projet, il pourra être estimé le nombre de bactériémies à bactéries résistantes aux antibiotiques évitées si l'on diminuait la prescription d'antibiotiques.

---

# 125

infections à bactéries multirésistantes, et plus de 5 500 décès associés, en France en 2015

---





## DEEPSARC

Améliorer le parcours de soins des patients atteints de sarcomes : des tumeurs rares et multiformes

### PARTENAIRES

**Le Centre Léon Bérard** est un centre de lutte contre le cancer (CLCC) qui est situé en Auvergne-Rhône-Alpes et a une triple mission de soins, de recherche et d'enseignement. Il est membre d'Unicancer.

**Le réseau NETSARC+ (Network Sarcoma)** est le réseau de référence clinique coordonné par le **Centre Léon Bérard**. Ce réseau est constitué des équipes régionales de cliniciens référents dans la prise en charge des patients atteints de sarcomes ou tumeurs conjonctives.

**L'Institut Bergonié** est le centre régional de lutte contre le cancer (CLCC) de la Nouvelle-Aquitaine. Il est chargé d'une triple mission comme le Centre Léon Bérard et est également membre d'Unicancer.

**L'équipe REPERES** (Recherche en Pharmaco-épidémiologie et recours aux soins, UPRES EA-7449) est une unité intégrée de recherche, d'expertise et de formation issue de la convergence de deux équipes, l'une basée au CHU de Rennes et l'autre issue de l'Ecole des hautes études en santé publique (EHESP, campus de Rennes, Département Méthodes quantitatives en santé publique, METIS).

Les sarcomes sont des tumeurs rares et multiformes touchant les tissus mous, les os et les viscères. En France, près de 5 000 personnes chaque année sont diagnostiquées pour ce type de cancer. Avec une survie moyenne à 5 ans de 50 à 60 %, le pronostic des patients demeure sévère. Ce type de cancer étant rare, les études cliniques menées concernent des cohortes de petites tailles (inférieures à 300 patients). La prise en charge des sarcomes repose donc sur un niveau de preuve encore faible, malgré une recherche clinique ancienne et active. Environ la moitié des patients ne répondent pas aux traitements standards.

Le projet DEEPSARC, coordonné par le Centre Léon Bérard et soutenu par l'Institut Bergonié, le réseau NETSARC+ et l'équipe REPERES, vise à identifier les traitements les plus adaptés à chaque profil de patient, grâce aux "données de vie réelle" (par opposition aux données recueillies dans les essais cliniques), favorisant ainsi ses chances de survie.

À terme, les résultats de cette étude permettront d'améliorer le pronostic des patients porteurs de sarcome ne répondant pas aux traitements standards.

**50 à 60 %**

en moyenne de taux de survie à 5 ans

Pour définir le meilleur parcours de soins thérapeutique au service des patients atteints de sarcome, le projet consiste à utiliser une approche complémentaire à celle des essais cliniques. Pour ce faire, le projet s'appuie sur les données existantes de l'ensemble des patients atteints de sarcome traités en France entre 2010 et 2017 (contenues dans la base issue du réseau NETSARC+) croisées avec les données de la base principale du SNDS.

La base du réseau NETSARC+ contient des données anatomocliniques qui ont été recueillies lors du recensement quasi exhaustif des cas de sarcome par les 25 équipes régionales de cliniciens du réseau référents dans la prise en charge de ces patients. La base principale contient quant à elle des données de remboursements et de facturation hospitalière, ainsi que les causes médicales de décès.

Le projet DEEPSARC a été lauréat du premier appel à projets du Health Data Hub lancé en 2019. Dans ce cadre, l'apport du Health Data Hub a notamment consisté à fournir plusieurs espaces d'analyse au sein de sa plateforme technologique. Ces espaces étaient notamment nécessaires pour réaliser la première phase de croisement entre les deux sources de données. Cette phase essentielle et particulièrement technique a été réalisée grâce à l'expertise de l'équipe REPERES et à l'outil de chaînage CUBR-LINK de la société CUBR qui a pu être mis à disposition au sein de la plateforme technologique dans le respect de hautes conditions de sécurité. Ce logiciel





s'avérait nécessaire afin d'être en capacité d'explorer le très grand nombre de possibilités d'appariement pour un patient de NETSARC+ avec son correspondant au sein de la base principale. L'équipe du projet DEEPSARC a aussi pu bénéficier de l'appui de l'Assurance Maladie pour la mise à disposition des données issues de la base principale.

— L'étape relative à l'appariement des données de 19 000 patients est désormais finalisée, tandis que l'analyse statistique supervisée par l'Institut Bergonié à Bordeaux est en cours. Les premières courbes de survie ont été produites.

---

**5 000**

---

nouveaux cas de sarcomes en France par an

---

**50 à 60 %**

---

en moyenne de taux de survie en 5 ans



# JUMENTAU NUMÉRIQUE "ÉTAT DE SANTÉ D'UN TERRITOIRE"

Mieux prévenir les allergies aux pollens grâce aux données de santé et environnementales

## PARTENAIRES

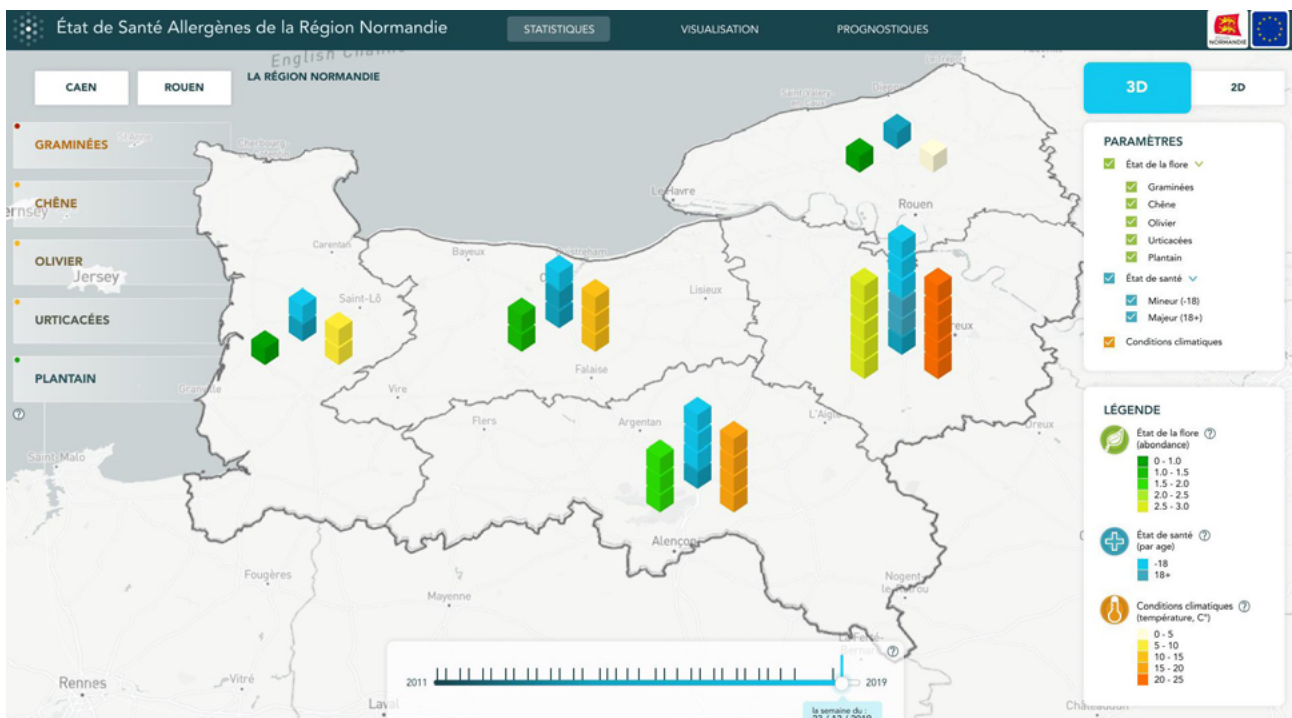
**Le Groupe VYV** est un acteur mutualiste en santé et protection sociale qui opère dans trois secteurs : assurance, soins et accompagnement, logement. Il propose des offres personnalisées pour accompagner et protéger ses adhérents.

**Le DataLab Normandie** réunit, au sein d'un consortium, des acteurs normands et un écosystème d'acteurs spécialistes de la donnée et de l'intelligence artificielle. Son objectif est de faciliter l'usage des données pour favoriser l'émergence de services numériques innovants.

En France, 4 millions de personnes sont diagnostiquées asthmatiques et 12 millions souffrent de rhinites allergiques. Les allergies respiratoires sont au premier rang des maladies chroniques de l'enfant. 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. Les coûts d'hospitalisation liés aux allergies respiratoires sont estimés à un milliard d'euros.

Les allergies respiratoires sont souvent liées à l'environnement, notamment aux pollens. Le Réseau national de surveillance aérobiologique (RNSA) observe les concentrations de pollens sur 80 sites en France depuis plus de 30 ans. Grâce à cette vigilance, le RNSA évalue les risques d'allergies en fonction de l'exposition aux pollens par zone géographique.

C'est dans ce contexte que le projet Jumeau numérique "État de santé d'un territoire" a été lancé sous l'égide d'un consortium réunissant : l'université de Caen et son laboratoire de recherche en informatique GREYC, le RNSA pour son expertise dans les pollens et son réseau d'allergologues, la société





OREKA Ingénierie basée à Cherbourg, experte en réalisations virtuelles et la société SAAGIE pour sa solution de HUB DATA. Le Groupe VYV, chef de file du consortium, apporte son expertise en datascience et en données de santé. Mené à l'initiative de la Région Normandie, le projet est soutenu par la communauté européenne.

■ Ce projet combine données de santé et données environnementales dans l'optique de mettre en relation les prescriptions de médicaments, les épisodes polliniques issus du RNSA (pollens par types, volumes, etc.) et les caractéristiques du territoire : plantations, arbres, cultures, météo. La base ainsi 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.

L'étude des corrélations est permise via la superposition des cartes géographiques, elle-même obtenue via un travail algorithmique de modélisation, dans une logique de création d'un "jumeau numérique". Un jumeau numérique est une représentation numérique d'un objet physique ou d'un système. Les jumeaux numériques collectent des données en temps réel et permettent de montrer l'évolution d'un système dans le temps. Des

---

En 2017,

**1 personne sur 3**

atteinte de rhinite allergique

---

approches issues de l'intelligence artificielle, de l'apprentissage automatique et plus généralement de l'analyse des données sont utilisées pour créer des modèles de simulation numérique qui se mettent à jour et changent à mesure que leurs contreparties physiques évoluent.

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 neuves 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 prendre en compte ce type de facteurs dans l'aménagement du territoire.

---

En 2022,

**4 millions**

de personnes diagnostiquées asthmatiques en France

---



## HYDRO

Mieux prédire les crises d'insuffisance cardiaque

### PARTENAIRE

**Implicity**, start-up MedTech dans le secteur de la cardiologie, est un acteur de la télésurveillance des dispositifs électroniques cardiaques implantables (DECI) grâce à sa plateforme de télésuivi disponible pour l'ensemble des fabricants de DECI en France et à l'international.

L'insuffisance cardiaque est un problème majeur de santé publique en France, qui concerne près de 2 millions de personnes, et qui est à l'origine de 70 000 décès chaque année. Avec plus de 160 000 hospitalisations annuelles, elle représente la première cause d'hospitalisation des plus de 65 ans. Le coût annuel des hospitalisations dues à cette maladie chronique est estimé à 1,6 milliard d'euros.

L'insuffisance cardiaque est une pathologie liée le plus souvent à une perte d'efficacité de la fonction "pompe" du cœur qui affecte le quotidien du patient (essoufflement, fatigue disproportionnée, etc.). Malgré de grands progrès, la prise en charge thérapeutique des patients reste complexe et les hospitalisations difficiles à anticiper.

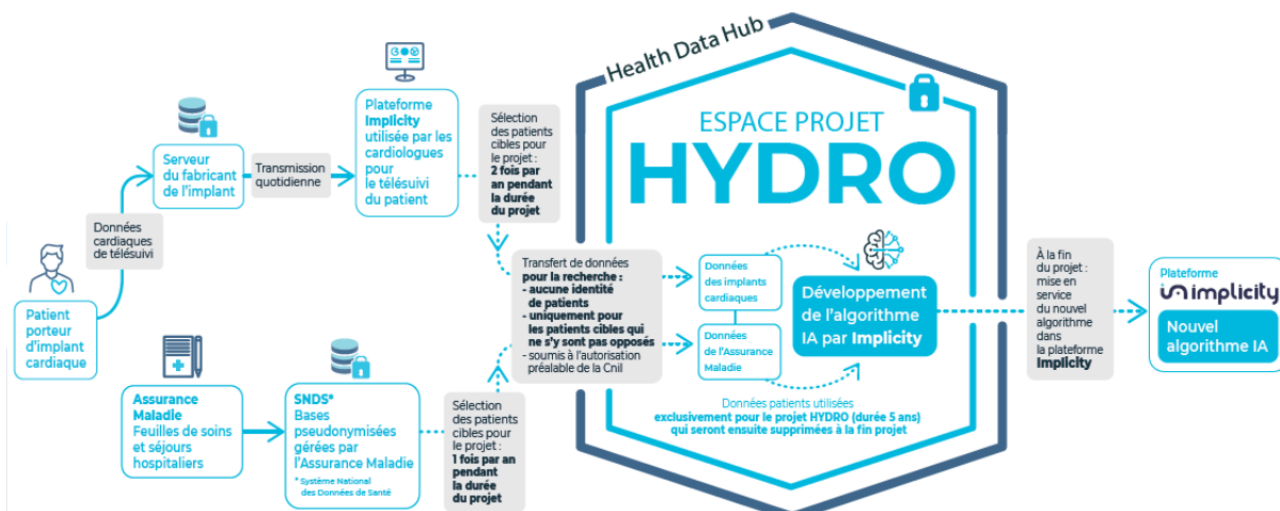
L'objectif du projet HYDRO, porté par Implicity, est de prédire les crises d'insuffisance cardiaque menant à une hospitalisation pour les patients porteurs d'un pacemaker grâce à un score de risque de décompensation cardiaque, calculé et mis à jour très régulièrement pour chaque patient. Une

compréhension plus fine des facteurs prévalents à une insuffisance cardiaque permettra une prise en charge thérapeutique plus personnalisée et préventive qui contribuera à réduire la mortalité et les hospitalisations.

Pour prévenir les crises d'insuffisance cardiaque, des solutions existent mais reposent généralement sur un suivi quotidien du poids et des signes cliniques du patient via un questionnaire.

Le projet HYDRO utilise les données issues des outils de télésuivi (capteurs cardiaques connectés) apportés par la plateforme d'Implicity qui permet aux cardiologues de suivre leurs patients de manière centralisée, et celles de la base principale du SNDS qui portent sur les remboursements et la facturation hospitalière. 35 000 patients ont ainsi été informés de l'utilisation de leurs données dé-identifiées à des fins de recherche. À partir de ces deux sources de données, pourra être développé un algorithme de prédiction des hospitalisations pour insuffisance cardiaque. Les données de l'Assurance

### PARCOURS DES DONNÉES PATIENT DANS LE CADRE DU PROJET





Maladie permettront d'évaluer la performance de l'algorithme. Dans le cas où ces résultats de performance seraient satisfaisants, un tel algorithme pourrait être intégré à la plateforme de télésuivi à des fins de prévention.

Le projet HYDRO a été lauréat du premier appel à projets du Health Data Hub lancé en 2019. Dans ce cadre, l'apport des équipes du Health Data Hub s'est concrétisé par un accompagnement dans les différentes phases de la mise en oeuvre du projet et notamment, aux côtés des experts d'Implicitity et de l'Assurance Maladie, pour obtenir un haut niveau de qualité dans le cadre de la réalisation du croisement entre les deux sources de données. En effet, s'agissant d'un appariement indirect, c'est-à-dire qui ne pouvait pas se baser sur un identifiant commun aux deux bases de données, cette étape a requis un travail approfondi de plusieurs mois pour réaliser la

jointure entre les bases. Dans la phase de création de l'algorithme, le Health Data Hub a aussi fourni son appui et expertise en datascience à l'équipe d'Implicitity pour traiter les grandes volumétries de données mobilisées.

Les performances des algorithmes obtenus étant très prometteuses, Implicitity a soumis un dossier d'autorisation pour leur dispositif médical au BSI et à la Food and Drug Administration (FDA). BSI est un organisme notifié évaluant le dossier réglementaire dans le but d'obtenir un marquage CE nécessaire à la mise sur le marché du dispositif médical en Europe. La FDA est l'administration américaine des denrées alimentaires et des médicaments. Cet organisme a notamment pour mandat d'autoriser la commercialisation des médicaments sur le territoire des États-Unis.

---

En 2021, en France, l'insuffisance cardiaque touche près de  
**2 millions** de personnes

---

---

**200 000**  
hospitalisations et 70 000 décès  
chaque année en France sont liés  
à un problème d'insuffisance cardiaque

---



## TARPON

Réduire les risques traumatiques grâce à l'intelligence artificielle

### PARTENAIRES

**L'Institut national de la santé et de la recherche médicale (Inserm)** est le seul organisme public de recherche français entièrement dédié à la santé humaine. Son objectif est d'améliorer la santé de tous par le progrès des connaissances sur le vivant et sur les maladies, l'innovation dans les traitements et la recherche en santé publique.

**Le CHU de Bordeaux** regroupe trois sites hospitaliers qui assurent la prise en charge des soins de la population bordelaise et d'Aquitaine. Étroitement lié à l'université de Bordeaux, il participe de manière constante à la formation médicale, chirurgicale, pharmaceutique et odontologique de la région ainsi qu'à la recherche fondamentale.

**Le Centre de recherche sur la santé des populations de Bordeaux**, aussi appelé **Bordeaux Population Health (BPH)**, est une unité de recherche co-hébergée par l'université de Bordeaux et l'Inserm. Il se concentre sur les principales priorités de santé publique dans un large éventail de disciplines.

Sur plus de 20 millions de passages aux urgences annuels en France, 35 à 40% sont dus à des traumatismes. Les traumatismes non intentionnels entraînent chaque année environ 40 000 décès, dont 21 000 à la suite d'un accident de la vie courante. Cependant, peu d'informations sont disponibles pour analyser l'origine de ces accidents traumatiques et ainsi mieux les prévenir. En effet, les informations sur la cause du traumatisme existent bien dans le dossier médical du patient, mais sous forme de texte libre noté par les personnels soignants à l'occasion du passage aux urgences. Il n'existe pas de traitement automatisé de ces informations pourtant cruciales.

Le projet TARPON, porté par l'Inserm, le CHU de Bordeaux et le Bordeaux Population Health, vise à mesurer les risques liés à la consommation de médicaments pouvant expliquer certains traumatismes. À titre d'exemple, il s'agira d'analyser si la prise de certains médicaments peut entraîner une baisse de la vigilance en situation de conduite, pouvant causer un accident de la route. Ces travaux permettraient d'envisager à terme la mise en place d'un système national de surveillance des traumatismes.

Lorsqu'une personne arrive aux urgences, le professionnel de santé qui l'accueille lui ouvre un dossier patient pour y noter les premières informations sur son état. Dans le cadre du projet TARPON, cette information sera analysée automatiquement

afin de comptabiliser le nombre exact de personnes arrivées aux urgences suite à un traumatisme et de caractériser ce dernier.

À l'aide de ces annotations et après en avoir évalué la qualité, le porteur du projet et ses partenaires vont pouvoir notamment les croiser avec les données de la base principale du SNDS afin d'apprécier les corrélations entre les types de traumatismes et la consommation de médicaments les précédant. L'analyse automatique des dossiers médicaux fait appel à une technologie de traitement automatique du langage (Generative Pre-trained Transformer).

Ce projet a été lauréat du second appel à projets du Health Data Hub, lancé en 2020 en partenariat avec le Grand Défi "amélioration des diagnostics médicaux par l'intelligence artificielle" et Bpifrance. Dans ce cadre, le projet a bénéficié d'un financement et d'un accompagnement dans ses démarches. La phase d'analyse post croisement pourra être conduite sur la plateforme technologique du Health Data Hub.

Dans le cadre de la première phase du projet, 69 110 notes cliniques liées à des visites aux urgences du CHU de Bordeaux ont fait l'objet d'une annotation manuelle par une équipe d'infirmiers et d'infirmières des urgences pendant cinq mois. Parmi ces notes cliniques, 22 481 concernaient des traumatismes. Cette base de données annotées sert de référence pour évaluer l'algorithme de prédiction qui identifie à partir de la note qu'il s'agit d'un traumatisme



et son type. L'algorithme développé a permis de classer correctement 97 % des comptes rendus, ce qui est nettement supérieur à ce qui était obtenu avec des méthodes moins innovantes. Ces résultats ont fait l'objet de plusieurs articles scientifiques :

› [Bourdois, L., Avalos, M., Chenais, G., Conrand, B., Gil-Jardiné, C., Guennec-Jacques, A., ... & Lagarde, E. \(2021, Juin\). Traitement automatique des résumés de passages aux urgences: focus sur la désidentification. In \*PFIA 2021-Santé et IA\*.](#)

› [Chenais, G., Touchais, H., Avalos, M., Bourdois, L., Revel, P., Gil-Jardiné, C., & Lagarde, E. \(2021, Juin\). Performance of BERT models for French in the classification of textual data from emergency room visits. In \*PFIA 2021-Journée Santé et IA\*.](#)

› [Chenais, G., Gil-Jardiné, C., Touchais, H., Fernandez, M. A., Conrand, B., Tellier, E., ... & Lagarde, E. \(2023\). Deep Learning Transformer Models for Building a Comprehensive and Real-time Trauma Observatory: Development and Validation Study. \*JMIR AI\*, 2\(1\), e40843.](#)

L'appariement avec les données de la base principale du SNDS a par ailleurs démarré grâce à l'appui des équipes de l'Assurance Maladie.

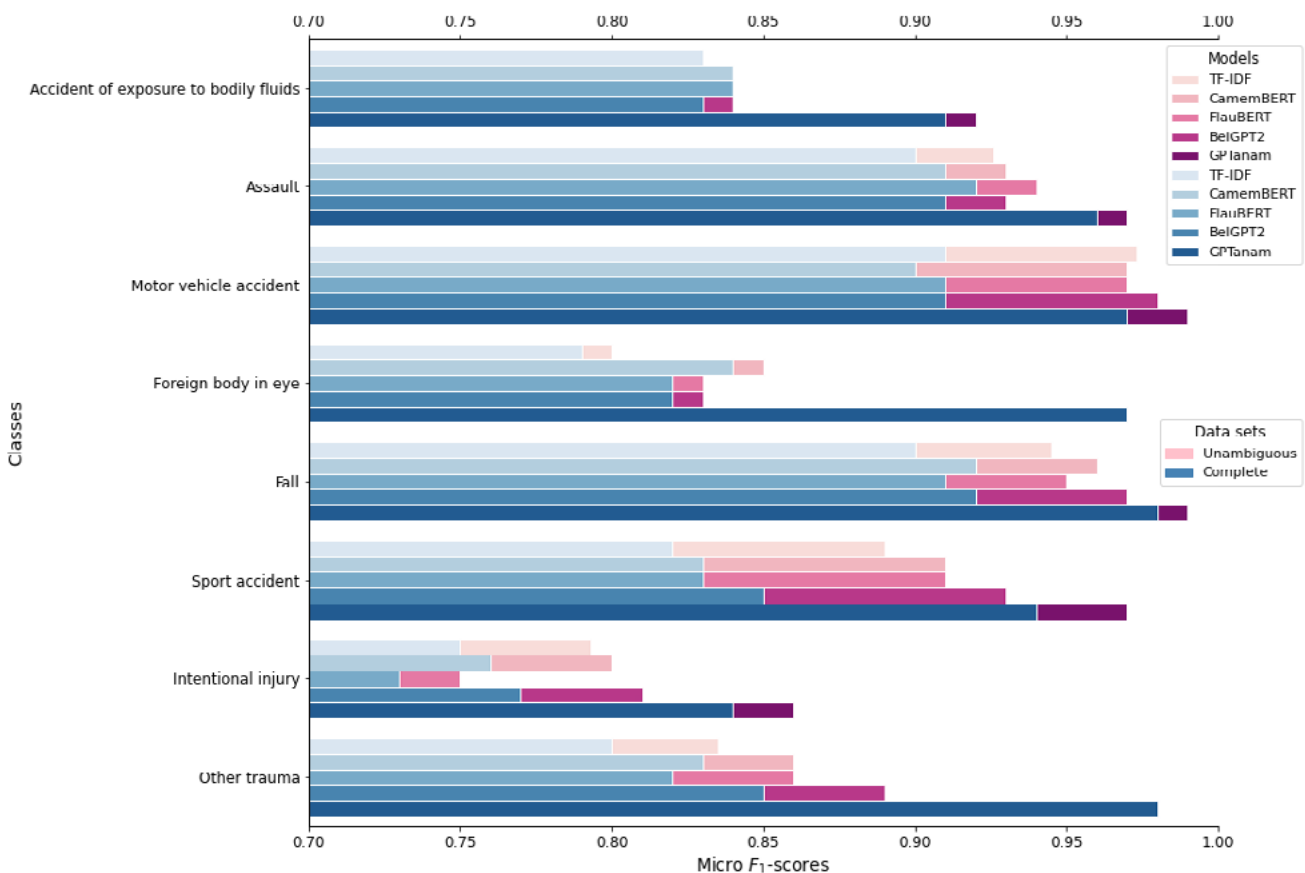
Plus de

**20 millions**

de passages annuels aux urgences en France en 2022

**35 à 40 %**

sont dus à des traumatismes







## INNERVE

Améliorer la prise en charge des neuropathies grâce à l'intelligence artificielle

### PARTENAIRES

**Quantmetry** est un cabinet de conseil en IA qui accompagne ses clients, de la stratégie à l'industrialisation, depuis plus de 10 ans.

**Assistance Publique-Hôpitaux de Paris (AP-HP)** est un établissement de santé public et un centre hospitalier universitaire (CHU) d'Île-de-France. Elle regroupe 38 hôpitaux, situés à Paris, en petite couronne et en province.

Les neuropathies et leurs douleurs associées font partie des pathologies les plus fréquentes vues par les neurologues. La plupart des neuropathies affectent les fibres de petites tailles qui permettent de ressentir la douleur et la température.

Pour diagnostiquer cette affection, les professionnels de santé prescrivent une biopsie cutanée, qui est ensuite examinée au scanner pour compter le nombre de petites fibres atteignant la surface de la peau. Il s'agit d'un examen de référence pour ce type de pathologies. Néanmoins, l'analyse des résultats est un travail fastidieux et chronophage prenant en moyenne au professionnel de santé 1h30 par patient. De plus, l'errance thérapeutique reste importante : les patients peuvent en souffrir et ne parviennent que tardivement à obtenir un diagnostic.

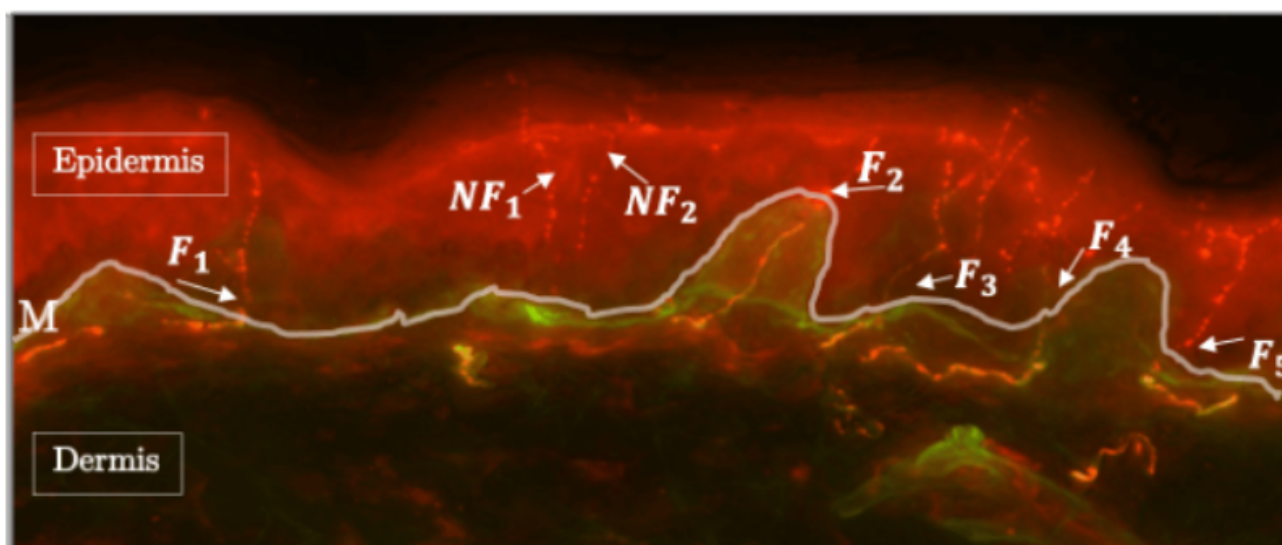
Le projet INNERVE, porté par Quantmetry et l'AP-HP, vise à développer un logiciel s'intégrant directement à l'environnement de travail du médecin,

lui permettant d'analyser plus facilement les images de biopsies produites par le scanner. Il comptera automatiquement et rapidement ces petites fibres, et extraiera les biomarqueurs pertinents pour affiner le diagnostic (type de neuropathie et sévérité de l'affection). Le logiciel sera également équipé d'une interface permettant aux professionnels de santé de corriger les erreurs du logiciel.

Un tel logiciel sera donc de nature à améliorer la pose du diagnostic en le rendant plus rapide mais également plus précis, réduisant par conséquent l'errance thérapeutique.

Le logiciel développé constituera la première solution industrielle dédiée au décompte des fibres nerveuses de petites tailles.

Pour y parvenir, un algorithme d'intelligence artificielle est développé sur la base d'une base de données anonymisées issues de 80 biopsies et de 5 900 annotations posées par les neurologues.





Le projet INNERVE a été lauréat du second appel à projets du Health Data Hub lancé en 2020 en partenariat avec le Grand Défi “amélioration des diagnostics médicaux par l’intelligence artificielle” et Bpifrance. Le Health Data Hub a principalement apporté son concours pour le développement d’une méthodologie permettant d’évaluer le caractère anonyme des données utilisées. Un guide a été par la suite publié pour que ce travail puisse bénéficier à l’ensemble de la communauté.

Le développement d’une première version de l’outil est désormais achevé et son déploiement a commencé à l’AP-HP, au Kremlin-Bicêtre, partenaire de Quantmetry depuis le début du projet. Les équipes vont débiter la phase de certification du dispositif médical. Il est utilisé pour l’instant dans un contexte de recherche uniquement.

Le projet est également dans une phase d’étude rétrospective visant à documenter de façon précise et quantifiée le fonctionnement de l’outil et sa plus-value. Elle se terminera à la fin de l’année 2023. Ces travaux pourront alimenter un futur article scientifique.

Par ailleurs, une partie des données anonymes pourront être mises en open data dans les prochains mois de sorte à favoriser l’innovation dans la communauté scientifique.

---

En 2019, on estime

**35 000 cas**

de neuropathies de petites fibres  
en France

---



## NHANCE

Faciliter la pose d'un diagnostic grâce aux échographies

### PARTENAIRES

**Assistance Publique-Hôpitaux de Paris (AP-HP)** est un établissement de santé public et un centre hospitalier universitaire (CHU) d'Île-de-France. Elle regroupe 38 hôpitaux, situés à Paris, en petite couronne et en province.

**Inria** est l'Institut national de recherche en sciences et technologies du numérique. Au sein de 215 équipes projets, pour la plupart partagées avec de grandes universités de recherche, plus de 3 900 chercheurs et ingénieurs explorent de nouvelles voies, souvent de manière interdisciplinaire et en collaboration avec des partenaires industriels, pour relever des défis ambitieux.

**Nhance** est une association dont l'objectif est de rendre les soins de santé accessibles à tous grâce à l'IA. Elle est composée d'experts en échographie pour annoter les images et d'experts en deep learning.

— L'échographie est une technique de recueil d'images qui n'irradie pas le patient. Elle peut servir tout autant à un obstétricien pour voir un fœtus dans le ventre, qu'aux autres spécialités médicales et chirurgicales. Toutefois, interpréter une image reste complexe, avec des erreurs possibles.

— Le projet NHANCE, porté par l'AP-HP, Inria et l'association Nhance, vise à aider les professionnels de santé à faire la meilleure interprétation possible des échographies abdominales en leur mettant à disposition un outil d'assistance au diagnostic. Plus précisément, il permettra de guider la localisation des organes abdominaux lors de l'examen. Pour deux organes importants, en l'occurrence le foie et les reins, l'outil servira à identifier automatiquement des maladies que l'on peut habituellement détecter à partir d'une échographie afin d'aider les médecins et rendre plus rapide et précis leur diagnostic.

— Concrètement, le projet s'appuie sur une base d'images échographiques issues des hôpitaux de l'AP-HP pour améliorer la fiabilité de premiers modèles d'intelligence artificielle qui reconnaissent déjà certains organes et certaines maladies.

Afin de constituer la base de données indispensable au projet, le développement d'un outil permettant d'extraire et de pré-traiter semi-automatiquement les données a été nécessaire. Ainsi, c'est 120 000 images échographiques de l'hôpital Saint-Louis qui ont pu être extraites en seulement quelques semaines au lieu des deux ans nécessaires si cette extraction avait dû être faite manuellement.

270 000 images échographiques ont également pu être collectées à partir de l'entrepôt de données de santé de l'AP-HP grâce au ciblage des données.

Le projet NHANCE est un projet institutionnel lancé en 2019. Dans ce contexte, le Health Data Hub a apporté son soutien sous la forme d'expertise pour les deux chantiers techniques permettant la constitution de la base de données et mettra à disposition sa plateforme technologique pour l'entraînement final des algorithmes.

Environ

**28 millions**

d'échographies réalisées en France en 2014, constituant le poste de dépense le plus important en matière de radiologie médicale



Alors que la préparation de la base de données est en cours de finalisation, cinq articles scientifiques soulignant la plus-value de la mobilisation de techniques d'apprentissage profond dans la caractérisation des échographies du foie ont été publiés :

› [Dadoun, H., Rousseau, A. L., de Kerviler, E., Correas, J. M., Tissier, A. M., Joujou, F., ... & Ayache, N. \(2022\). Deep learning for the detection, localization, and characterization of focal liver lesions on abdominal US images. \*Radiology: Artificial Intelligence\*, 4\(3\), e210110.](#)

› [Dadoun, H., Delingette, H., Rousseau, A. L., de Kerviler, E., & Ayache, N. \(2023\). Deep clustering for abdominal organ classification in ultrasound imaging. \*Journal of Medical Imaging\*, 10\(3\), 034502-034502.](#)

› [Dadoun, H., Delingette, H., Rousseau, A. L., de Kerviler, E., & Ayache, N. \(2021, Avril\). Combining Bayesian And Deep Learning Methods For The Delineation Of The Fan In Ultrasound Images. In 2021 IEEE 18th International Symposium on Biomedical Imaging \(ISBI\) \(pp. 743-747\). IEEE.](#)

› [Dadoun, H., Delingette, H., Rousseau, A. L., de Kerviler, E., & Ayache, N. \(2023, Avril\). Joint representation learning from french radiological reports and ultrasound images. In \*IEEE ISBI 2023-International Symposium on Biomedical Imaging\*.](#)

› [Dadoun, H. \(2022\). Analyse d'images d'échographie abdominale basée sur l'IA pour aider au diagnostic médical \(Doctoral dissertation, Université Côte d'Azur\).](#)

Ils attestent que certains outils mobilisant des technologies d'apprentissage peuvent égaler l'expertise de praticiens pour la détection et la caractérisation de lésions du foie.

Près de

**6 millions**

d'échographies concernaient l'appareil digestif et les grossesses en 2014

RSNA

## Deep Learning for the Detection, Localization, and Characterization of Focal Liver Lesions on Abdominal US Images

### Key Result

Pretrained deep neural networks had high performance for detection, localization, and characterization of focal liver lesions (FLLs) on abdominal US images.

### Patients:

- Adults with lesions in liver parenchyma:
- Training and development set: 1026
  - Test set: 48

### Methods:

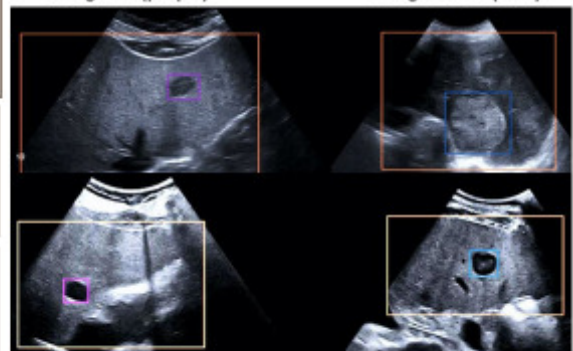
- Faster R-CNN and DETR neural networks were fine-tuned, and model performance for detection, localization, and characterizations of FLLs was compared to that of expert and nonexpert caregivers.

### Results:

- DETR showed higher performance for all tasks compared to Faster RCNN.
- For detection of liver parenchyma with or without lesions, DETR met or exceeded the performances of two experts, with 90% specificity and 97% sensitivity.

Benign FLL (purple)

Malignant FLL (blue)



- DETR achieved the highest performance among all raters with 81% specificity 82% sensitivity for characterization of FLLs.

Dadoun, H and Rousseau AL et al. Published Online: March 2, 2022  
<https://doi.org/10.1148/ryai.210110>

Radiology: Artificial Intelligence



## DEEP.PISTE

Adapter le parcours de dépistage organisé des cancers du sein grâce à l'IA

### PARTENAIRES

**Le Centre Régional de Coordination des Dépistages des Cancers (CRCDC)** en Occitanie est chargé de l'organisation des dépistages du cancer du sein, du col de l'utérus et colorectal de la région. L'objectif du dépistage est de diagnostiquer un cancer à un stade précoce, avant l'apparition même des symptômes, afin d'augmenter les chances de guérison à l'aide d'un traitement moins lourd.

**Epiconcept** est une société qui contribue à l'amélioration des programmes de santé publique en renforçant la connaissance des maladies et en mettant à la disposition des différents acteurs, des experts métiers et des solutions technologiques.

Le cancer du sein est le plus mortel chez la femme en France : près de 12 000 femmes en décèdent chaque année. Le dépistage organisé a été généralisé à l'ensemble du territoire en 2004. Il vise à effectuer une mammographie de dépistage tous les deux ans chez les femmes de 50 à 74 ans, ce qui devrait permettre de faire baisser la mortalité de près de 21 %. Néanmoins, encore aujourd'hui, par exemple, les faux positifs (les personnes diagnostiquées alors qu'elles n'ont pas de cancer) suite à un examen par mammographie avoisinent les 80 % ; et environ un cancer sur six n'est pas détecté lors de cet examen (c'est un faux négatif). Il y a donc encore matière à progresser, et il est nécessaire de mesurer l'efficacité du programme de dépistage.

Le projet DEEP.PISTE porté par le CRCDC Occitanie et Epiconcept, vise à améliorer le programme de dépistage du cancer du sein en développant une analyse automatique des mammographies et en affinant la compréhension des facteurs de risque.

Concrètement, le projet DEEP.PISTE croise les données issues des mammographies de la banque de patients du CRCDC Occitanie avec les données de l'Assurance Maladie. Une analyse approfondie des données combinées permettra d'identifier les patientes nécessitant un suivi médical renforcé et celles pouvant bénéficier d'un allègement du parcours.

**5 millions**

de femmes bénéficient d'un dépistage gratuit annuel en France

Après avoir franchi toutes les étapes réglementaires, les données du CRCDC ont été transférées vers la plateforme du Health Data Hub, où elles seront combinées aux données de l'Assurance Maladie. Chaque mammographie positive ou associée à un cancer sera ensuite annotée par un radiologue afin de déterminer la localisation du cancer. Une fois ces deux étapes réalisées, il est possible de reconstruire le parcours de chaque femme de manière détaillée pour pouvoir tester plusieurs améliorations possibles.

Le projet DEEP.PISTE a été lauréat du premier appel à projets du Health Data Hub lancé en 2019. L'équipe du projet DEEP.PISTE a aussi pu bénéficier de l'appui de l'Assurance Maladie pour la mise à disposition des données issues de la base principale du SNDS.

Alors que l'étude est en cours sur la plateforme du Health Data Hub, **le logiciel développé pour dé-identifier les mammographies**, a pu être partagé en open source. Il pourra être réutilisé par tout chercheur utilisant des outils DICOM (outils de gestion informatique des données issues de l'imagerie médicale).

**12 000** femmes

meurent chaque année d'un cancer du sein

**60 000**

nouveaux cas sont diagnostiqués chaque année



## REXETRIS

Optimiser la prise en charge des patients greffés rénaux

### PARTENAIRES

Le **CHU de Limoges**, avec son service de pharmacologie, toxicologie et pharmacovigilance et avec l'**Unité Mixte de Recherche "Pharmacologie des immunosuppresseurs en transplantation" (UMR 1248) de l'Inserm** propose aux professionnels de santé transplantateurs des outils d'ajustement de posologie des traitements immunosuppresseurs, qui ont été utilisés pour près de la moitié des transplantés rénaux en France.

L'**Agence de la biomédecine (ABM)** est une agence publique nationale de l'État. Elle exerce ses missions dans les domaines du prélèvement et de la greffe d'organes, de tissus et de cellules, ainsi que dans les domaines de la procréation, de l'embryologie et de la génétique humaines. Par son expertise, elle est l'autorité de référence sur les aspects médicaux, scientifiques et éthiques relatifs à ces questions.

■ Pour les patients souffrant d'insuffisance rénale, la dialyse et la greffe rénale sont les traitements principalement proposés. Lorsque la greffe est possible, elle améliore considérablement la survie des patients par rapport à la dialyse. Néanmoins, les listes d'attente s'allongent, retardant pour certains patients l'accès à la greffe, et pouvant parfois être la cause de décès. L'une des raisons expliquant la saturation de ces listes est la fréquence des rejets de greffe.

Bien que la survie des greffons se soit améliorée, elle reste limitée dans le temps et de nombreux patients doivent recevoir une deuxième, voire une troisième transplantation. La survie du greffon dépend en grande partie des particularités du patient (système immunitaire propre à chacun), et du traitement immunosuppresseur (qui diminue l'activité du système immunitaire).

■ Le projet REXETRIS cherche à définir comment mieux utiliser dans la durée les traitements existants, pour prolonger le fonctionnement des reins greffés et éviter (ou retarder) les re-transplantations ou le retour en dialyse. Dans le cadre d'un traitement à vie, la connaissance issue de ce projet permettrait

d'optimiser les stratégies thérapeutiques, les doses, mais aussi les formules de ces médicaments.

À long terme, une plus large diffusion de ces modèles permettrait de désencombrer la liste d'attente des demandeurs de greffe, en prolongeant la durée de vie des greffons.

■ Le projet REXETRIS repose sur un croisement de plusieurs sources de données pour établir des modèles de risque "cliniques" (perte de fonction du greffon, décès), des modèles prédictifs de réponse aux traitements immunosuppresseurs (score de prédiction individuel) et valider une méthode innovante d'adaptation de posologie par méthode pharmacocinétique.

Ces sources sont au nombre de trois : CRISTAL de l'Agence de la Biomédecine, la base principale du SNDS de l'Assurance Maladie et ABIS du CHU de Limoges.

---

# 14 ans

est la durée de vie moyenne d'un greffon rénal

---





---

**40 000**

patients portent un greffon rénal en France

---

Le projet REXETRIS a été lauréat du premier appel à projets du Health Data Hub lancé en 2019. Dans ce cadre, l'apport du Health Data Hub a notamment consisté à fournir un appui en expertise pour la réalisation de l'appariement avec la base principale du SNDS qui reposait sur une approche probabiliste. Ces espaces étaient notamment nécessaires pour réaliser la première phase de croisement entre les deux sources de données. L'équipe du projet REXETRIS a aussi pu bénéficier de l'appui de l'Assurance Maladie pour la mise à disposition des données issues de la base principale.

Les deux appariements ont été réalisés afin de constituer une base pseudonymisée réunissant l'ensemble des patients transplantés rénaux suivis dans les centres de transplantation français depuis 2005.

Dans l'objectif d'une optimisation de la prise en charge des patients greffés rénaux, les porteurs de projet proposeront notamment une amélioration du score de prédiction du risque de perte du greffon rénal, développé, breveté et publié par l'UMR 1248 en 2017. Ce score de prédiction individuel du risque sera rendu disponible pour tous les médecins en charge des transplantés rénaux, à travers le système expert ABIS. Les porteurs du projet REXETRIS émettront également des recommandations de personnalisation des traitements immunosuppresseurs des patients transplantés.

---

**500**

patients meurent chaque année  
faute d'accès à une greffe rénale

---





## E-CORE

Décrire l'utilisation des glucocorticoïdes par voie systémique en cas de Covid-19

### PARTENAIRES

**PELyon** est un bureau d'études spécialisé dans les études en vie réelle dont le but est d'aider les acteurs de la santé, publics ou privés, à répondre à leurs questions de recherche et de stratégie.

**IQVIA** est un acteur de la recherche clinique et de la donnée de santé. En connectant données, informations, technologies et expertises, IQVIA accompagne ses clients et ses partenaires pour leur permettre de développer et de commercialiser des thérapies innovantes.

■ Environ 10 à 20 % des patients positifs à la Covid-19, en particulier les plus âgés ou ceux présentant des comorbidités, développent une pneumopathie suivie parfois d'un syndrome de détresse respiratoire aiguë, qui nécessite une hospitalisation et une assistance ventilatoire. Cette population présente également un risque élevé de troubles inflammatoires induits par le système immunitaire.

L'utilisation systémique de glucocorticoïdes, comme la dexaméthasone, a été associée à une réduction de la mortalité chez les patients ventilés ou sous oxygénothérapie d'appoint. Cependant, il manque encore des informations sur l'utilisation de corticostéroïdes chez les patients atteints de Covid-19, notamment : le type de molécule, la posologie, le moment et la durée de l'administration, ainsi que l'identification des groupes de patients chez lesquels ces traitements auront le plus de bénéfices, etc.

■ Afin de répondre à ces enjeux, l'Agence européenne des médicaments (EMA) a initié, en 2020, une étude au niveau européen dont l'objectif principal est de décrire l'utilisation des glucocorticoïdes par voie systémique en cas de Covid-19.

**210 Go**

de données avant transformation

**33 Go**

après standardisation au format OMOP-CDM

■ En juin 2020, l'EMA a confié à IQVIA un projet visant à conduire plusieurs études de cohortes multicentriques sur l'utilisation des médicaments chez les patients atteints de Covid-19. Il avait pour objectif d'étudier rapidement dans plusieurs pays l'utilisation, l'efficacité et la sécurité des thérapies pour le traitement de la Covid-19. Pour répondre en partie à ces questions, le protocole demandé par l'EMA prévoit d'explorer les schémas d'utilisation et d'administration systémiques de glucocorticoïdes chez les patients ayant un test PCR positif pour le SRAS-CoV-2 (Covid-19) et les cas suspects de Covid-19 (sCovid-19) à partir de bases de données de santé de neuf pays : sept pays européens (Allemagne, Belgique, Espagne, France, Italie, Pays-Bas et Serbie) ainsi que des États-Unis et du Royaume-Uni. Un travail de mise au format OMOP-CDM était requis pour que les bases de données soient interoperables.

Pour la France, c'est le bureau d'études PELyon qui a réalisé les analyses pour le compte d'IQVIA. Il a utilisé les données issues des soins en ambulatoire (soins reçus en dehors de l'hôpital) et dans le cadre hospitalier (soins reçus lors d'une hospitalisation) pour 300 000 patients hospitalisés pour Covid-19 en 2020 et 2021 contenues dans la base principale du SNDS. Les données ont été transmises au Health Data Hub grâce à l'aide des équipes de l'Assurance Maladie pour qu'il les transforme au format OMOP-CDM et en teste la qualité.

**300 000**

données de patients atteints de la Covid-19  
mises au format OMOP-CDM



**180 tables** et  
**4500 variables**

avant transformation

Ces résultats ont fait l'objet d'un article scientifique décrivant l'administration de glucocorticoïdes chez les patients atteints de la Covid-19 et ses conséquences. Première étude de la Covid-19 mobilisant des données de vie réelle à l'échelle européenne, elle révèle une grande diversité de protocoles au sein des pays étudiés.

› [Layton D., Reich C., Duarte Salles T., Rjinbeek P., Prieto-Alhambra D., Salek S. \(2021\), \*Systemic glucocorticoids use in the treatment of COVID-19 and risks of adverse outcomes in COVID-19 patients in the primary and secondary care setting.\*](#)

**37 tables** et  
**396 variables**

après transformation

› [Leclerc, J., Brodovicz, K. G., Chang, S., Assimon, M. M., Urushihara, H., & Girman, C. J. \(2021, Août\). \*Believability of real-world evidence \(RWE\) for informed regulatory decisions.\* In \*PHARMACOEPIDEMIOLOGY AND DRUG SAFETY\* \(Vol. 30, pp. 3-3\). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.](#)

Si cette diversité ne permet pas de faire ressortir de conclusion arrêtée, l'étude a permis la constitution d'une cohorte inédite à l'échelle de neuf pays qui pourra être mobilisée pour de prochaines recherches internationales.

Il s'agit par ailleurs du premier projet conduit sur les données de la base principale du SNDS transformées vers un format international et interopérable. Il a été présenté dans différents congrès tels que le Medical Informatics Europe 2022 et l'OHDSI Europe 2022.



## EHDEN

Standardiser les données pour mener des études de cohortes plus larges

### PARTENAIRE

L'**European Health Data & Evidence Network (EHDEN)** a été lancé en novembre 2018 pour relever les défis relatifs à l'utilisation de données de vie réelle à grande échelle. EHDEN rassemble 25 partenaires de 12 pays : 11 partenaires publics, dirigés par l'Erasmus University Medical Centre au Pays-Bas (coordinateur), et 14 partenaires pharmaceutiques, dirigés par Janssen Pharmaceutica NV en Belgique (chef de projet).

La base principale du SNDS est une base médico-administrative très riche, mais complexe et qui reste encore sous exploitée. Une des difficultés à son utilisation est le temps nécessaire pour monter en compétences sur le schéma de données et la compréhension des biais et limites de ces données. Standardiser la base principale du SNDS dans un modèle de données commun, en s'appuyant sur des normes de représentation et de stockage, permet de simplifier et d'élargir son utilisation pour des projets de recherche nationaux et européens. Cette standardisation est par exemple très bénéfique dans le cas de maladies rares, puisque les pays disposent individuellement d'un volume de données trop réduit pour mener des recherches. Standardiser les données permet d'agréger les résultats de plusieurs bases de données et ainsi mener des études sur des cohortes plus larges.

En 2020, le Health Data Hub a été lauréat d'un appel à projets porté par EHDEN. Le projet retenu visait à transformer le SNDS Fast-Track (base relative aux patients atteints de la Covid-19 sur 2019-2020) au format OMOP-CDM (Observational medical outcomes partnership - Common Data Model). Ce format, centré sur le patient, permet de passer d'une base comportant plus de 180 tables pour la base principale du SNDS, à moins de 20 tables métiers (moins de 40 tables en comptant les terminologies et métadonnées). Cela permet d'analyser plus facilement les parcours de soins, quelle que soit la provenance géographique ou administrative de la donnée.

La standardisation de la base principale du SNDS s'est déclinée en trois étapes principales.

Tout d'abord, l'alignement syntaxique consiste à homogénéiser les structures des bases de données et trouver les correspondances entre les schémas de données source et cible. Par exemple, dans le cadre d'une standardisation au format OMOP-CDM, cela consiste à identifier où sont stockées les informations du patient dans la base de données source, et à les reporter dans la table "PERSON" du format cible. L'alignement syntaxique nécessite une connaissance très fine de la base principale du SNDS et plusieurs experts du Health Data Hub ont été mobilisés sur ce projet.

La seconde étape consiste à aligner les terminologies françaises vers des terminologies standards. Par exemple, il s'agira de trouver l'équivalent dans une terminologie internationale, telle que la SNOMED-CT, du code "chirurgie de la hanche" selon une nomenclature française. À titre d'illustration, il y a plus de 70 000 codes sur les procédures, sur les médicaments et sur la biologie en France. Ce travail d'alignement de terminologie a été effectué avec l'aide d'internes en médecine. Il a été choisi pour les codes de procédures médicales d'aligner prioritairement les procédures les plus fréquemment conduites en ville et à l'hôpital sur les années 2019 et 2020. Cela représente près de 700 codes sur 9 000. Les codes procédures restants ont été alignés au niveau du chapitre supérieur. Chaque alignement de codes ou chapitres a été relu et corrigé par un second interne en médecine.



Enfin, une vérification de la qualité de la standardisation a été nécessaire. OHDSI propose plus de 3 000 tests de qualité permettant de vérifier, par exemple, qu'une date de fin de soins est bien postérieure à la date de début ou encore si le nombre de femmes ayant un cancer de la prostate ne dépasse pas un seuil défini. Dans le cas d'une valeur trop élevée, soit il y a une erreur dans l'alignement syntaxique ou sémantique, soit l'erreur provient de la base de données source. Ces résultats de qualité ont été transmis et discutés avec EHDEN afin d'améliorer la transformation.

Grâce à ce travail, les données de la base principale du SNDS des années 2019 et 2020 ont été mises au standard OMOP-CDM, tandis que la documentation décrivant le processus de standardisation des données ainsi que les scripts créés ont été ouverts en open source.

Les résultats ont notamment été dévoilés lors de l'événement du **14 décembre 2022** organisé par le Health Data Hub, dont le replay est disponible [ici](#). La documentation de la transformation se trouve sur la [documentation collaborative SNDS](#) et les scripts de transformation sont sur [Gitlab](#). Ces travaux ont également été présentés en 2022 à des congrès internationaux et européens tels que les congrès [OHDSI](#) (Rotterdam et Washington DC).

À nouveau lauréat de l'appel à projets du réseau EHDEN en 2021, le Health Data Hub a poursuivi les travaux de standardisation de la base principale du SNDS, permettant de couvrir une profondeur de données plus importante (2015-2021). Ces résultats pourront également contribuer au projet visant à créer et tester une version beta de l'Espace européen des données de santé (Health Data@EU Pilot), qui s'intéresse entre autres à la standardisation des données de santé.

À noter que ce travail a été rendu possible avec l'aide de l'Assurance Maladie qui a extrait les données.

Passage de 180 tables à moins de

**20 tables**

sur la base principale du SNDS



## PROJET MICI - AFA CROHN RCH FRANCE

Mieux connaître les populations atteintes de maladies inflammatoires chroniques de l'intestin (MICI)

### PARTENAIRE

L'association **Afa Crohn RCH France** est dédiée aux personnes atteintes de maladies inflammatoires chroniques de l'intestin, telles que la maladie de Crohn et la rectocolite hémorragique. Engagée depuis plus de 40 ans, l'Afa est un soutien pour les patients et leurs proches, offrant écoute, informations, ressources et accompagnement. Elle construit une communauté et encourage la recherche médicale pour améliorer la vie des personnes touchées par les MICI.

Les MICI sont les maladies inflammatoires chroniques de l'intestin ; elles comprennent la maladie de Crohn et la rectocolite hémorragique (RCH). Cela se caractérise par une inflammation de la paroi du tube digestif. Au quotidien, les patients peuvent avoir des douleurs abdominales, des diarrhées, des douleurs anales, des pertes d'appétit, des nausées et vomissements, etc. Ils vivent chaque jour avec ces symptômes qui handicapent leur quotidien. Les MICI évoluent par périodes de poussées, avec des durées et intensités variables. Le diagnostic de ces pathologies peut prendre plusieurs années. Les facteurs qui en sont responsables demeurent encore mal compris à ce jour.

En Europe, 3 millions de personnes seraient concernées, bien que ces chiffres varient selon les publications scientifiques. Il existe actuellement des données en open data pour caractériser les personnes atteintes de cette pathologie (nombre de personnes concernées selon le sexe, les lieux de vie, etc.) mais elles sont souvent partielles.

Le projet vise à construire un tableau de bord recensant différents indicateurs et caractéristiques

des personnes atteintes de MICI, ainsi que leur prise en charge. Ces données permettront à l'Afa d'en savoir davantage sur les patients concernés, d'apprécier l'évolution de la pathologie dans le temps et de conduire différents types d'actions, dont de la prévention.

Ce projet s'inscrit dans une démarche open science, avec pour objectif la diffusion d'une information accessible à tous. Les codes utilisés pour l'estimation des indicateurs et les éléments de chiffrage obtenus seront en accès libre et pourront compléter la documentation ouverte relative à la base principale du SNDS.

L'apport au Health Data Hub a consisté à construire ce tableau de bord. Des experts ont notamment utilisé l'algorithme de repérage de l'Assurance Maladie pour identifier les patients atteints de MICI, dont la description est disponible en ligne dans la cartographie des pathologies et des dépenses. Des indicateurs sur les patients, notamment leurs caractéristiques démographiques (âge, sexe, localisation), leurs traitements et leur prise en charge hospitalière ont été calculés et restitués. Des

Trouver des pistes de recherche, établir des facteurs de risques, améliorer la vie des malades

ACCUEIL BIBLIOTHÈQUE DE DONNÉES ÉTUDES EN COURS SONDAGES EN LIGNE QUI SOMMES-NOUS ? afa

**COMBIEN DE MALADES DE CROHN ET DE RCH EN FRANCE ?**  
Les dernières données de l'assurance-maladie sur le régime des ALD  
VOIR L'INFOGRAPHIE



échanges entre l'association, des cliniciens et les experts du Health Data Hub ont permis d'affiner ces indicateurs.

Le premier tableau de bord a été restitué à l'Afa. Il comprend notamment une déclinaison du nombre de patients pris en charge pour des MICI en 2019, par sexe, classe d'âge, répartition géographique ; du nombre d'hospitalisations longues et de jours ; du volume des principales chirurgies digestives et de la part des personnes ayant reçu au moins un acte en lien avec une pose de stomie digestive ; du nombre moyen de jours d'arrêt dans l'année (même s'il n'est pas possible de relier ces arrêts à leur cause) ; et du recours aux principales classes de médicaments.

Un support pédagogique comprenant les clés de lecture a été fourni. Il permettra de faciliter la réu-

tilisation et la valorisation de ces éléments. Il est convenu de mettre ces indicateurs à jour tous les trois ans. Ces données viennent compléter des indicateurs déjà recueillis dans le cadre de l'observatoire de l'Afa et issus principalement de l'Assurance Maladie.

En 2019, près de

**273 000**

personnes étaient prises en charge dans le cadre d'une MICI

**44 %** d'entre elles (soit 120 000 personnes) souffrent de la maladie de Crohn



## PROJET MALADIES NEUROMUSCULAIRES - AFM TÉLÉTHON

Dénombrer et décrire les patients atteints de maladies neuromusculaires bénéficiant de certains traitements

### PARTENAIRE

L'AFM Téléthon est une association de malades atteints de maladies neuromusculaires et de proches de malades. Elle mène un travail de lutte contre des maladies rares, évolutives et lourdement invalidantes.

Les maladies neuromusculaires sont des pathologies du muscle ou de sa commande nerveuse. L'unité motrice (constituée par le muscle, la jonction nerf/muscle et le nerf) fonctionne mal et le muscle ne peut pas se contracter normalement. Le symptôme le plus courant est une faiblesse musculaire, qui peut varier en termes de moment d'apparition, d'étendue sur les muscles (un ou plusieurs groupes musculaires), de durée (temporaire ou permanente) et de progression (évolutive ou stable). Ces maladies peuvent causer d'autres symptômes : des problèmes cardiaques, des difficultés d'apprentissage, une atteinte de l'œil et/ou de l'audition, etc. Suivant leur nature, ces maladies se déclenchent à des âges différents de la vie, mais nombre d'entre elles sont infantiles.

Le projet vise à construire un tableau de bord identifiant les médicaments prescrits aux personnes atteintes de certaines maladies neuromusculaires. Celui-ci permettra à l'AFM de produire de la connaissance sur l'accès aux médicaments intervenant dans ces maladies.

Ce projet s'inscrit dans une démarche open science, avec pour objectif la diffusion d'une information accessible à tous. Les codes utilisés pour l'estimation des indicateurs et les éléments de chiffrage obtenus, seront en accès libre et pourront compléter la documentation ouverte relative à la base principale du SNDS.

L'apport du Health Data Hub a consisté à construire ce tableau de bord. Dans un premier temps, la preuve de concept a été développée sur trois médicaments. Pour chacun d'eux, le travail a consisté à identifier le nombre de patients sur la période 2018-2021 selon l'âge, le sexe, la répartition géographique et la comédication (ou changement de traitement). Les codes de médicaments utilisés pour l'identification des traitements ont été récupérés grâce au serveur multi-terminologies de l'Agence du Numérique en Santé (ANS). Cette approche est en cours de généralisation aux autres médicaments intervenant dans les traitements des pathologies d'intérêt.

La preuve de concept a permis d'estimer le nombre de patients pour trois traitements (Zolgensma, Evrysdi et Spinraza) suivant les caractéristiques démographiques de ces derniers, dans l'amyotrophie spinale. Les résultats peuvent être corroborés par l'expérience terrain de l'association et l'existence d'un registre de qualité dans cette maladie. Le travail entrepris permettra d'affiner la méthodologie et les éléments pertinents pour l'AFM Téléthon. Une fois cette étape validée, les indicateurs pourront être développés pour plusieurs autres maladies d'intérêt pour l'association.





## PROJET VASCULARITES - FRANCE VASCULARITES

Mieux connaître les patients atteints de vascularites pour améliorer leur prise en charge

### PARTENAIRE

**France Vascularites** est une association dédiée au soutien, à l'information et à l'accompagnement des personnes atteintes de vascularites en France. Fondée par des patients et leurs proches, elle vise à sensibiliser sur les différentes formes de vascularites, favoriser l'échange d'expériences entre les patients, promouvoir la recherche et améliorer la prise en charge médicale de ces maladies.

Les vascularites sont des maladies rares auto-immunes, ce qui signifie que le système immunitaire du corps attaque par erreur ses propres tissus et organes. Elles se traduisent par une inflammation des vaisseaux sanguins, ou de la paroi de ces vaisseaux. L'inflammation va endommager les organes et les tissus irrigués par ces vaisseaux pouvant ainsi entraîner des atteintes rénales, pulmonaires, cardiaques, digestives, ORL, cutanées et cérébrales. Elles peuvent provoquer différents symptômes tels que de la fièvre, de la fatigue, une perte de poids et d'appétit, des douleurs articulaires et musculaires, ou encore des neuropathies. Les vascularites regroupent une quinzaine de maladies, toutes rares, classifiées en fonction de la taille des vaisseaux atteints. Elle peuvent atteindre homme, femme, enfant de tout âge et de toute origine.

Le projet vise à construire quelques tableaux recensant différents indicateurs et caractéristiques des personnes atteintes de vascularites, ainsi que leur prise en charge. Ces données permettront à France vascularites d'en savoir davantage sur les patients concernés.

Ce projet s'inscrit dans une démarche open science, avec pour objectif la diffusion d'une information accessible à tous. Les codes utilisés pour l'estimation des indicateurs et les éléments de chiffrage obtenus seront en accès libre et pourront compléter la documentation ouverte relative à la base principale du SNDS.

L'apport du Health Data Hub a consisté à construire ce tableau de bord. À partir des échanges entre l'association, des experts cliniques et des experts du Health Data Hub, une méthode d'identification des maladies en lien avec la vascularite a été identifiée, puis des indicateurs pertinents pour l'association ont été formalisés. Ont ainsi été sélectionnées les personnes ayant eu au moins un séjour en établissement public ou privé en 2020 en médecine, chirurgie, obstétrique et odontologie (MCO) avec un diagnostic de vascularite (diagnostic principal, relié ou associé) identifié par des codes CIM-10.

Des tableaux ont été restitués à France Vascularites comprenant notamment une déclinaison du nombre de patients pris en charge pour vascularite en 2020, par sexe, région, classe d'âge et statut vital.

En 2020,

**6 815**

personnes étaient suivies dans le cadre d'une prise en charge de vascularite dont

**2 661** dans le cadre d'une granulomatose avec polyangéite

Documentation du SNDS & SNDS OMOP

Le tableau suivant présente le nombre de patients pris en charge pour vascularite en 2020 dans le cadre d'une hospitalisation

Diagnostic	Nombre de patients en 2020
Granulomatose de Wegener (M313)	2061
Micro-polyangéites (M317)	1443
Syndrôme de la crose aortique [Takayasu] (M314)	545
Périartérite noueuse (M300)	437
Périartérite juvénile (M302)	<50
Autres affections apparentées à la périartérite noueuse (M308)	130
Syndrôme de Behçet (M352)	1832
Sous-ensemble des patients (M313 ou M317)	3916
Ensemble des patients (M313, M317, M300, M302, M308, M352)	6815

\* A ce jour, cinq vascularites ont pu être identifiées, validées et font l'objet d'une extraction.



## LE DATA CHALLENGE TISSUENET

Optimiser la phase de diagnostic en cas de cancer du col de l'utérus

### PARTENAIRE

La **Société Française de Pathologie (SFP)** réunit les pathologistes de tout mode d'exercice. Elle est régie par la loi de 1901, reconnue d'utilité publique et compte 913 membres. Elle est vouée à l'étude, la promotion et l'enseignement de l'anatomie et cytologie pathologiques.

Actuellement, les diagnostics de lésions cancéreuses et précancéreuses du col de l'utérus sont réalisés à partir de l'observation de lames d'anatomopathologie au microscope optique. Les outils de l'intelligence artificielle, appliqués à l'analyse de lames virtuelles – c'est-à-dire issues des lames numérisées – sont susceptibles d'apporter une aide au diagnostic aux professionnels de santé.

Ainsi, pour favoriser le développement d'outils d'intelligence artificielle, un premier Data Challenge a été lancé par la Société Française de Pathologie en partenariat avec le Health Data Hub, le Grand Défi "amélioration des diagnostics médicaux par l'intelligence artificielle" et Bpifrance.

Le but d'un Data Challenge est, pour ses participants, de répondre en un temps imparti et à partir de données spécifiques, à une question posée. Il s'adresse aux data scientists (chercheurs, indus-

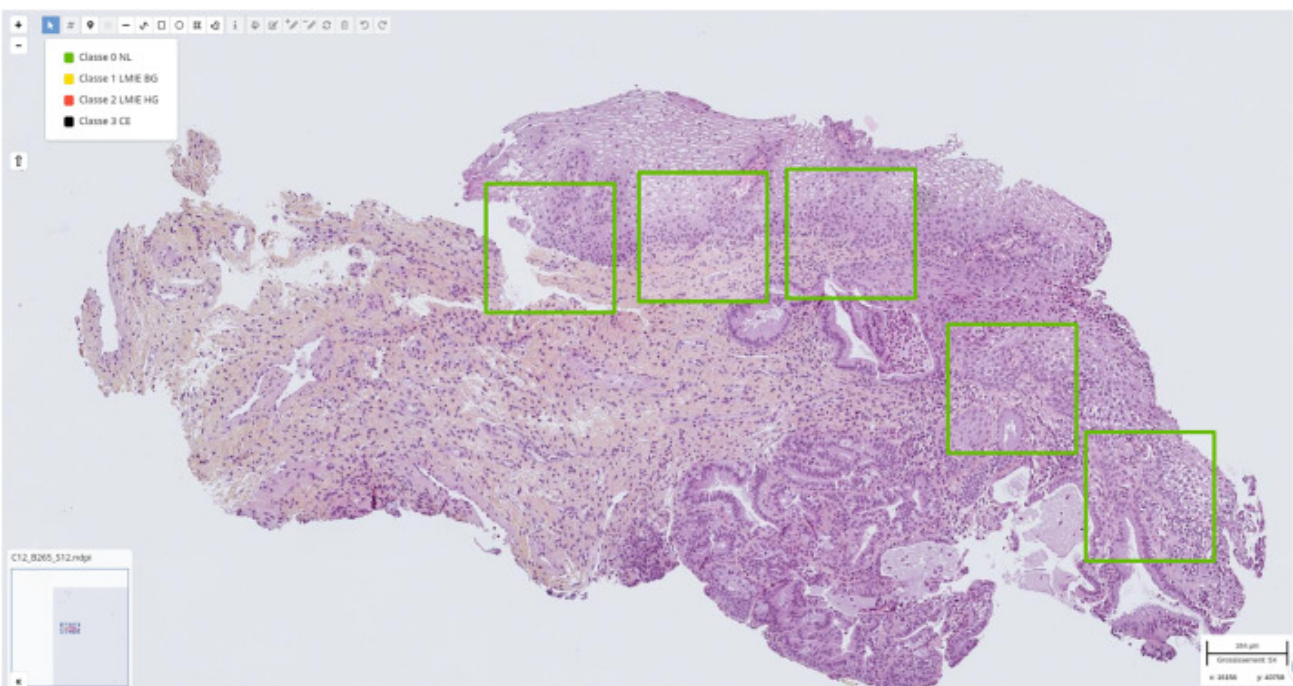
triels, étudiants...) de toutes nationalités. Les compétiteurs développent des algorithmes à partir d'un jeu de données anonymisées. De surcroît, l'objectif du projet s'inscrit dans une démarche de partage et d'ouverture des algorithmes à l'écosystème, au service de la recherche. Dans une ultime étape, l'accessibilité des données et des algorithmes issus du Data Challenge est encouragée afin de permettre à la recherche de poursuivre sa progression.

Les étapes majeures de ce projet ont impliqué la formulation de la problématique médicale, la collecte et la préparation des données. Cette préparation a inclus une phase d'annotation par des médecins

Près de

# 5 000

lames de biopsie numérisées





Scores finaux allant jusqu'à

**93 %**

spécialistes, couvrant plus de 5 000 images de lames histologiques de biopsies du col de l'utérus numérisées. Une fois les images annotées et rendues anonymes, elles ont été mises à la disposition des participants sur une plateforme dédiée. C'est sur cette plateforme de Data Challenge que les participants ont pu accéder aux données et élaborer des algorithmes capables de classifier les lésions en fonction de leur gravité.

Après la clôture du Data Challenge, les algorithmes proposés ont fait l'objet d'une évaluation selon une métrique préalablement définie en collaboration avec des médecins et des experts en science des données. Les équipes ayant développé les solutions les plus performantes ont été récompensées par un prix global de 25 000 € à se partager.

Le Health Data Hub a joué un rôle essentiel tout au long du processus de mise en œuvre du projet TissueNet en fournissant un soutien organisationnel, logistique et technique. En effet, le Health Data Hub a mobilisé ses experts en data science et en apprentissage automatique pour contribuer à la conception du Data Challenge. Sur le plan juridique, les bonnes pratiques ont été partagées pour garantir un traitement des données en conformité avec le Règlement général sur la protection des données (RGPD). Parallèlement, un appui logistique a été fourni lors de la collecte des données auprès d'une vingtaine de centres fournisseurs de données, ainsi que dans l'information des patients concernés.

**20**

centres fournisseurs de données volontaires

Les lauréats de cette compétition ont réussi à détecter les lésions du col de l'utérus avec une grande précision. Ils ont obtenu des scores supérieurs à 90 %, ce qui est remarquable pour une compétition de deux mois seulement. Les algorithmes des lauréats ayant accepté l'ouverture de leur solution ont été rendus disponibles en open source sur GitHub, pour bénéficier à l'ensemble de la communauté scientifique.

Par ailleurs, deux articles scientifiques ont été publiés à l'issue de ce Data Challenge :

› [Loménie, N., Bertrand, C., Fick, R. H., Hadj, S. B., Tayart, B., Tilmant, C., ... & Bertheau, P.\(2022\). Can AI predict epithelial lesion categories via automated analysis of cervical biopsies: The TissueNet challenge?. Journal of Pathology Informatics,13,100149.](#)

› [Delaune, A., Valmary-Degano, S., Loménie, N., Zryouil, K., Benyahia, N., Trassard, O., ... & Bertheau, P. \(2022, March\). Le premier data challenge organisé par la Société Française de Pathologie: une compétition internationale en 2020, un outil de recherche en intelligence artificielle pour l'avenir?. In Annales de Pathologie \(Vol. 42, No. 2, pp. 119-128\). Elsevier Masson.](#)

De plus, le Health Data Hub a mis en ligne un **MOOC** gratuit pour partager les bonnes pratiques à mettre en œuvre pour lancer un Data Challenge en santé.

**574**

participants du monde entier



## LE DATA CHALLENGE VISIOMEL

Contribuer à la recherche de nouveaux marqueurs prédictifs de la récurrence pour les mélanomes primitifs non métastatiques

### PARTENAIRES

**La Société Française de Pathologie (SFP)** réunit les pathologistes de tout mode d'exercice. Elle est régie par la loi de 1901, reconnue d'utilité publique et compte 913 membres. Elle est vouée à l'étude, la promotion et l'enseignement de l'anatomie et cytologie pathologiques.

**La Société Française de Dermatologie (SFD)** a pour mission d'assurer la promotion de la dermatologie française dans la communauté médicale et scientifique internationale. Elle s'inscrit dans le projet à travers **le Groupe de Cancérologie Cutanée (GGC)** qui regroupe des spécialistes impliqués dans la prise en charge des cancers de la peau. Il coordonne notamment l'accès aux données des bases nationales collaboratives.

**Le Conseil National Professionnel des Pathologistes (CNPath)** est le conseil national professionnel des médecins dont la spécialité est l'anatomie et la cytologie pathologique. En tant qu'association loi 1901, il a vocation à représenter tous les membres de cette spécialité médicale.

Les mélanomes représentent 10 % des cancers de la peau mais se distinguent par leur haut potentiel d'évolution métastatique. Ce potentiel métastatique est particulièrement complexe à évaluer concernant les mélanomes cutanés de stades précoces (mélanomes localisés et peu épais). En effet, les mélanomes fins (moins de 1 mm d'épaisseur) associés pour la majorité à un bon pronostic, sont cependant à l'origine d'environ un tiers de rechutes et de décès, sans que la pratique clinique n'arrive à les expliquer et les prédire. De même, pour les mélanomes d'épaisseur intermédiaire (compris entre 1 et 4 mm) à plus haut risque de récurrence, peu de facteurs prédictifs de cette potentielle évolution métastatique existent.

De nos jours, certains traitements permettent de limiter le risque de rechute pour des mélanomes de stade III-IV opérables. Cependant, au-delà de leur coût élevé, ces traitements exposent les patients à des toxicités médicamenteuses importantes. C'est pourquoi il devient urgent de cibler les patients dont le risque de rechute est élevé, et les distinguer de ceux qui, a contrario, n'obtiendraient pas de bénéfice clinique de ces thérapies.

Près de

**2 000**

lames d'exérèse et données cliniques

Le Data Challenge VisioMel, porté par la SFP, la SFD, le GGC et le CNPath, a pour objectif de contribuer à la recherche de nouveaux marqueurs prédictifs de la récurrence pour les mélanomes primitifs non métastatiques. Pour cela, les organisateurs donnent accès aux participants, le temps de la compétition, à une base de données cliniques anonymes associées aux lames histologiques numérisées de près de 2 000 patients. Les lauréats, c'est-à-dire ceux dont les algorithmes ont démontré la meilleure performance, bénéficient d'une récompense.

Dans le cadre du projet VisioMel, des données cliniques ont été associées aux données histologiques afin d'augmenter les chances d'identification de marqueurs pronostiques et d'aller au-delà de la segmentation obtenue avec les approches actuelles. Ces données cliniques sont issues de la base de données **RIC-Mel**, rassemblant à ce jour les données d'environ 40 000 patients atteints de mélanomes.

La préparation de la base de données – qui est probablement la plus grande base de données cliniques et histologiques sur le thème du mélanome – a demandé un travail collaboratif et minutieux de près d'un an. Une fois cette phase terminée, le Data Challenge a pu être lancé au début de l'année.





**541**

participants  
du monde entier

**637**

algorithmes  
proposés

**26**

centres hospitaliers  
mobilisés

**40**

laboratoires français  
d'anatomie pathologique

Le projet VisioMel a été lauréat de l'appel à projets du Health Data Hub, lancé en 2021 en partenariat avec le Grand Défi "amélioration des diagnostics médicaux par l'intelligence artificielle" et Bpifrance. Il est l'un des six lauréats et constitue le deuxième Data Challenge pour la Société Française de Pathologie. Dans ce contexte, l'équipe Data Challenge du Health Data Hub a accompagné les porteurs de projet tout au long de l'organisation.

D'après les premières analyses des prédictions soumises lors du Data Challenge, les algorithmes les plus performants ont correctement détecté en moyenne 73 % des cas de rechutes et 75% des cas "sans rechute". Ces résultats sont prometteurs. Des analyses complémentaires sont actuellement en cours pour déterminer si les algorithmes se révèlent plus performants pour certaines catégories de diagnostics.

Ces résultats préliminaires ont permis de démarquer trois lauréats qui ont été récompensés à l'occasion

d'une remise des prix organisée conjointement par la SFP et le Health Data Hub, le 25 mai 2023 à Parisanté Campus, dont le replay est disponible [ici](#). Les algorithmes des lauréats ayant accepté l'ouverture de leur solution ont été rendus disponibles en open source, sur [Github](#), pour permettre la poursuite des travaux.

Le projet a également permis de constituer une base de données riche et de qualité vouée à être partagée en open data.

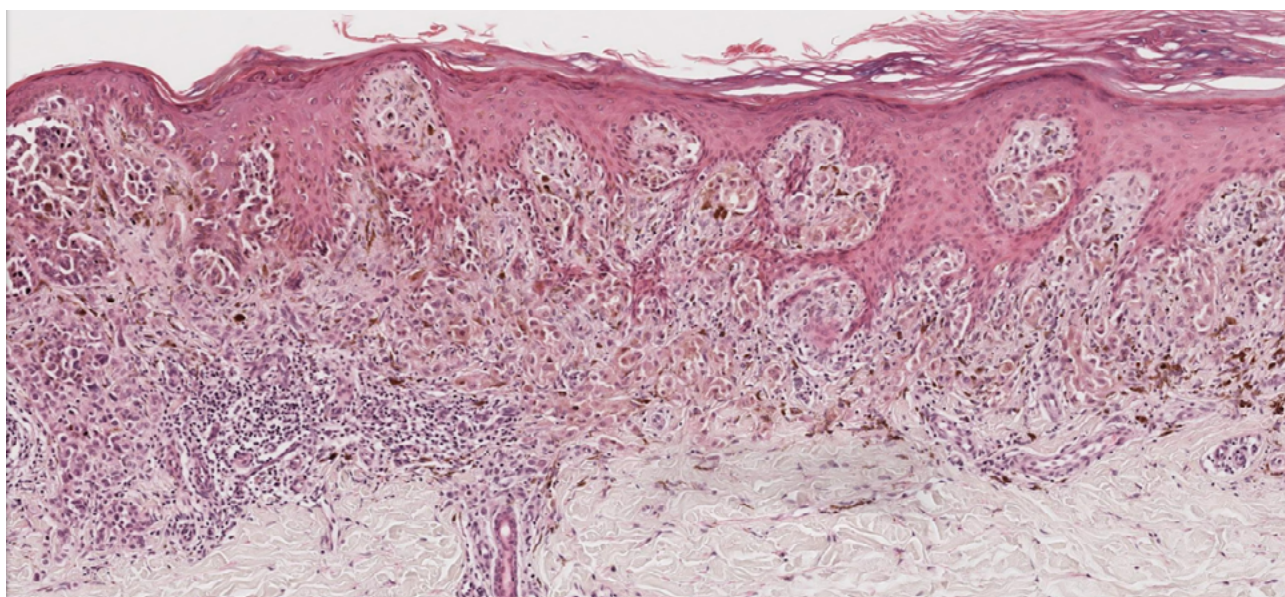
Détection correcte pour

**73 %**

des cas de rechute et pour

**75 %**

des cas de non rechute pour  
les trois algorithmes les plus performants





L'accès aisé et unifié, transparent et sécurisé,  
aux données de santé pour améliorer la qualité  
des soins et l'accompagnement des patients.

**9 rue Georges Pitard, 75015 Paris**  
**E-mail : [hdh@health-data-hub.fr](mailto:hdh@health-data-hub.fr)**

**Édition : Health Data Hub**  
**Conception et réalisation : Parimage**  
**Septembre 2023**

