



DOSSIER DE PRESSE  
Paris, 16 avril 2019

## HEALTH DATA HUB

### Annonce des lauréats du premier appel à projets

À la remise du rapport Villani, le Président de la République a indiqué que la santé serait un des secteurs prioritaires pour le développement de l'intelligence artificielle. Il a annoncé deux actions majeures : la **création d'un « Health Data Hub »** et **l'élargissement du système national de données de santé (SNDS)**. Le 16 mai 2018, Agnès Buzyn, Ministre des Solidarités et de la Santé a lancé une mission de préfiguration pour instruire ces deux actions.

Dans son rapport remis le 12 octobre 2018, la mission - pilotée par trois experts : Dominique Polton, présidente de l'Institut national des données en santé (INDS), Marc Cuggia, professeur d'informatique médicale et praticien hospitalier au CHU de Rennes et Gilles Wainrib, président fondateur de la start-up Owkin - propose une feuille de route pour la mise en œuvre opérationnelle de cette plateforme d'exploitation des données de santé. La ministre a confié cette action au directeur de la DREES (Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques du Ministère des Solidarités et de la Santé).

Une telle plateforme, facilitant les interactions entre producteurs des données de santé, utilisateurs et citoyens, dans des conditions élevées de sécurité, est essentielle à l'émergence d'innovations dans le domaine de la recherche. Elle est également requise pour la bonne réalisation de leurs missions par les acteurs de l'écosystème : ministères, agences et opérateurs. Il est en effet indispensable de mobiliser les données de santé pour éclairer le décideur, pour mener des missions de sécurité sanitaire et éviter des crises, ou encore pour développer une médecine de plus en plus personnalisée et accessible du plus grand nombre. Ne pas utiliser ce patrimoine collectif peut entraîner des pertes de chance pour le patient.

Conformément à l'ambition affichée dans le rapport de préfiguration, le Health Data Hub va être mis en place de façon progressive en s'appuyant sur l'écosystème de la santé. À cet effet, **un appel à projet a été lancé fin janvier dernier pour identifier les premières initiatives qui pourront bénéficier de l'accompagnement du Hub pour leur réalisation**. Ainsi, elles testeront l'offre de service du Hub en termes d'accompagnement technico-réglementaire et en termes de mise à disposition de technologies et de compétences. Elles contribueront également à la constitution du catalogue de données et d'outils du Hub.

L'accompagnement de « bout en bout » des projets sélectionnés via l'appel à projets permettra d'une part d'illustrer l'intérêt de l'utilisation de diverses sources de données pour améliorer la qualité des soins, mais aussi d'éprouver la pertinence de l'offre de service du Hub imaginée dans la lignée des recommandations du rapport de préfiguration. Enfin, il s'agira de préparer l'industrialisation de ces services en vue de l'ouverture plus large du Hub à la communauté au plus tard début d'année 2020.

La procédure était ouverte à tous : organismes de recherche, établissements et professionnels de santé, associations de patients, industriels, start-up... dès lors (1) qu'ils pouvaient faire état d'un usage innovant de la donnée de santé, (2) répondaient à une finalité d'intérêt public, (3) ne rentraient pas dans le champ des finalités interdites par le code de la santé publique. Conformément à la volonté de co-construction du Hub, les porteurs de projet devaient prévoir la mise à disposition de jeux de données, de documentation, d'outils ou d'algorithmes créés pouvant servir à l'ensemble de sa communauté.

Initié le 31 janvier, l'appel à projets a été clos le 9 mars 2019. La brièveté de la période de soumission laissait envisager une cinquantaine de candidatures maximum. Dans les faits, **ce sont 189 projets qui ont été présentés**. Ce nombre ainsi que la diversité des porteurs de projets (centres hospitaliers, agences, organismes de recherche, petits et gros industriels...) reflètent les attentes de l'écosystème vis à vis du Health Data Hub.

Afin de départager les candidats et ne retenir que les projets les plus pertinents à ce stade de la construction du Hub, la sélection s'est déroulée en 3 étapes. Dans la semaine du 11 au 15 mars, un comité de présélection composé de membres de l'équipe projet Health Data Hub a d'abord vérifié la recevabilité des dossiers : leur complétude et leur pertinence au regard de l'objectif de l'appel à projet. Du 18 au 22 mars, ces dossiers ont été évalués en fonction de **(1) leur maturité, (2) leur caractère innovant, (3) des bénéfices potentiels attendus, (4) leur contribution à la constitution du catalogue des données du Hub, (5) de leur**

**contribution au partage des données, des outils et des connaissances.** 21 projets ont été retenus en vue du jury conformément au dispositif annoncé lors de la publication de l'appel à projet.

Les 10 et 11 avril, les candidats ont pu défendre leur projet devant un jury composé de représentants qualifiés des domaines de la santé, de la recherche et de l'intelligence artificielle :

- Pr Bernard Nordlinger - Membre de l'Académie nationale de médecine, Service de chirurgie digestive générale et oncologique, Hôpital Ambroise Paré, Assistance Publique Hôpitaux de Paris, Président du jury
- Emmanuel Bacry - Directeur de Recherche au CNRS, Université de Paris Dauphine, Directeur scientifique du health data hub, Directeur projet santé/data Ecole Polytechnique
- Olivier Clatz - Directeur de programme grand défi "IA et diagnostics", Secrétariat général pour l'investissement, Services du Premier Ministre
- Stéphanie Combes - Cheffe de projet du Health Data Hub, Direction de la recherche, des études, de l'évaluation et des statistiques, Ministère des Solidarités et de la Santé
- Isabelle Gentil - Expert IA et santé, Ministère des Solidarités et de la Santé
- Pr. Didier Guillemot. Unité Mixte de recherche B2Phi U1181, Université Versailles Saint Quentin, Inserm, Institut Pasteur
- Yves Levy - Directeur exécutif du Vaccine Research Institute, INSERM, Conseiller d'Etat en Service Extraordinaire.
- Rosalie Maurisse - Responsable santé Direction innovation, Bpifrance
- Anne Paoletti, Directrice scientifique, secteur biologie santé, Direction générale pour la recherche et l'innovation, Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation
- Marie Seval, Conseiller information en santé, Direction générale de la santé, Ministère des Solidarités et de la Santé.

**À l'issue des deux jours d'auditions, le jury a finalement retenu dix projets**, qui, au-delà de leur qualité scientifique ou leurs retombées en termes d'amélioration du système de santé, vont permettre d'alimenter la première version du catalogue de données partagées par le Health Data Hub.

### **Deepsarc – Identifier les meilleurs schémas thérapeutiques pour le traitement du sarcome**

Cancer très complexe et varié, le sarcome est une catégorie de tumeurs rares pour laquelle l'efficacité des essais cliniques arrive à ses limites. Depuis près de 40 ans, des études randomisées ont été conduites sans pouvoir trancher sur l'intérêt de la chimiothérapie pour traiter les sarcomes, et sur la pertinence de l'utiliser avant ou après la chirurgie. Porté par le professeur Jean Yves Blay, Directeur Général du Centre Léon Bérard et directeur du réseau NetSarc, au côté de ses équipes et d'experts du Centre de lutte contre le cancer de Bordeaux et de l'université de Rennes, le projet Deepsarc vise à étudier l'impact des différents traitements sur données de vie réelle afin de déterminer le schéma thérapeutique le plus pertinent. Il mobilisera les données cliniques de la base NETSARC, base de référence sur le sarcome quasi exhaustive en France avec plus de 50 000 patients. Ces données, enrichies par les données médico administratives (Système National des Données de Santé), permettront de suivre les patients dans le temps.

### **Parcours IDM en IDF – Evaluer et améliorer les parcours de soin après un infarctus du myocarde**

Proposé par le docteur Axelle Menu-Branthomme, responsable de département à l'ARS Île-de-France, avec le soutien du SAMU 78 et du GCS Sesan, le projet prévoit de qualifier les parcours de soins des patients ayant eu un infarctus du myocarde aigu en Île-de-France. Cette pathologie, fréquente et mortelle, peut donner lieu à d'importantes complications (insuffisance cardiaque séquellaire par exemple). L'enrichissement du registre e-Must, qui recense finement l'évènement aigu lors de sa survenue, par des données médico-administratives du SNDS permettra d'évaluer différents types de parcours de soin du point de vue de la qualité, de la sécurité et de la pertinence.

## **PIMPON – Remonter aux prescripteurs des alertes pour les interactions médicamenteuses dangereuses**

Les outils de sécurisation de la prescription utilisés dans les logiciels d'aide à la prescription génèrent un nombre important d'alertes motivées par la détection quasi systématique d'interactions médicamenteuses à effet délétère éventuel. Cette multiplication des alertes sature la vigilance des prescripteurs qui leurs accordent de moins en moins d'attention. La gradation actuelle des alertes tient compte que de la gravité de leurs conséquences potentielles mais ignore la fréquence de ces complications. Porté par Jean-François Forget, directeur médical dans la société VIDAL, le projet vise à mobiliser les données du SNDS pour estimer la prévalence réelle des complications liées aux interactions médicamenteuses afin d'identifier des alertes nécessitant une mise en valeur particulière du fait de leur impact. L'attention des médecins pourra être restaurée vers les alertes les plus pertinentes, ce qui devrait aboutir à la réduction de l'incidence des complications graves, fréquentes et prévisibles.

## **Hydro – Prédire les crises d'insuffisance cardiaque pour les patients porteurs de pacemaker**

L'insuffisance cardiaque cause près de 5 % des hospitalisations en France, pour un coût de 1,8 milliard d'euros en 2016. La prédiction des épisodes d'aggravation constitue donc un véritable défi, que se propose de relever le docteur Arnaud Rosier, président de la startup Implicity. Son projet a vocation à développer une solution permettant de prévenir les crises d'insuffisance cardiaque pour les patients porteurs de pacemakers / défibrillateurs grâce à l'analyse des données de télé-suivi par intelligence artificielle. Le Health Data Hub permettra le croisement des données des dispositifs de plus de 8000 patients en soin courant avec celles du SNDS afin d'entraîner des modèles prédictifs sans avoir à labéliser manuellement les épisodes d'intérêt.

## **NS-Park – Prédire les trajectoires individuelles des patients Parkinsoniens**

Jean Christophe Corvol, Professeur de Neurologie à la Pitié-Salpêtrière, soutenu par l'ICM, l'Inserm et F-CRIN, porte ce projet dont l'objectif principal est de fournir aux neurologues un outils prédictif des trajectoires individuelles des patients parkinsoniens afin de mettre en œuvre les mesures préventives appropriées. Les données de cette très belle cohorte de 20 000 individus suivis dans les centres experts Parkinson seront chaînées à celles du SNDS pour pouvoir appréhender la prise en charge globale des patients. Des approches computationnelles et d'intelligence artificielle seront mises en œuvre pour modéliser la progression de la pathologie.

## **ARAC – Mesurer et comprendre les restes à charge réels des patients**

Porté par Laurent Borella, directeur santé à Malakoff Médéric Humanis, mutuelle à but non lucratif, le projet ARAC va permettre, en complétant les données de l'Assurance Maladie (SNDS) avec les des données de remboursements complémentaires, de calculer les restes à charge finaux réels des patients dans le modèle de protection sociale français. Aujourd'hui, ces restes à charge sont estimés à l'aide d'enquêtes ou imputés. Une mesure précise et à l'échelle individuelle de ces quantités constituera une ressource de grande valeur pour l'écosystème des économistes de la santé. Ils permettront notamment d'interroger les modes de financement des dépenses de santé, d'éclairer les questions de pouvoir d'achat et de solvabilisation du recours aux soins, etc.

## **ORDEI – Quantifier la proportion de patients touchés par un effet indésirable**

En 2017, près de 82 000 effets indésirables médicamenteux ont été déclarés pour 12 000 médicaments commercialisés. Les déclarations d'effets indésirables doivent s'interpréter au regard des indications du produit mais également du nombre de patients traités. Même s'il présente des limites, le taux de déclaration (nombre d'effets d'indésirables déclarés rapporté au nombre de patients exposés) est une estimation nécessaire pour les autorités, les professionnels de santé et les patients afin d'apprécier l'importance de chaque déclaration. Actuellement, son calcul se fait à la demande et « à la main ». Porté par Patrick Maison, conseiller scientifique à l'ANSM, l'objectif principal de ce projet est de développer un outil permettant d'automatiser ce calcul. Il s'appuiera à cette fin sur la base nationale de pharmacovigilance de l'ANSM, les données de prescriptions

remboursées en médecine de ville et en milieu hospitalier (SNDS) et les caractéristiques des produits dans CODEX (référentiel des AMM).

### **Oscour – Mobiliser les données d’urgences pour améliorer la surveillance sanitaire**

Porté par Yann Le Strat, directeur de la Direction Appui, Traitements et Analyses des données de l’Agence Nationale de Santé Publique, avec le concours de la FEDORU, des équipes du registre AVC de Brest et de l’équipe REPERES de l’université de Rennes, le projet soutenu propose de créer une base chaînant l’ensemble des données issues du réseau de surveillance Oscour® avec les données médico administratives du SNDS. Les données de la base Oscour® compilent les résumés de passage aux urgences depuis presque quinze ans, recensant plus de cent trente millions de passages aux urgences. Une telle cohorte permettra d’instruire quantité de questions en vue d’améliorer la surveillance sanitaire en France. Une première expérimentation sera menée sur les accidents vasculaires cérébraux, première cause de mortalité chez les femmes et troisième chez les hommes.

### **Deep.Piste – Evaluer l’apport de l’intelligence artificielle dans le dépistage organisé du cancer du sein**

Avec 11 883 décès estimés en 2017, le cancer du sein tue plus de femmes que tout autre cancer en France. Le dépistage précoce permet une baisse de 21% de la mortalité. Francisco Orchard, responsable unité data science au sein de la société Epiconcept, appuyé par son équipe, le Centre Régional de Coordination des Dépistages des Cancers en Occitanie et l’Institut Curie, propose ici d’évaluer l’impact d’insérer de l’intelligence artificielle dans le dispositif de dépistage organisé du cancer du sein. Il s’appuiera pour cela sur la base de données e-SIS de dépistage de Cancer du Sein des départements du Gard et de la Lozère, comportant plus de 250 000 images annotées, enrichies par les données médico administratives dans le but de réduire le nombre de faux négatifs. Cette base conséquente devrait permettre d’entraîner des modèles performants de réseaux de neurones convolutionnels sur des données jamais utilisées à ce jour dans cette optique.

### **REXETRIS – Mesurer l’impact à long terme de l’exposition aux médicaments immunosuppresseurs des patients greffé rénal**

Les informations concernant les relations entre expositions des patients aux médicaments et effets à long terme sont aujourd’hui très lacunaires, dans le cas général comme en transplantation d’organe. Dans le cadre d’un traitement à vie, la connaissance de ces relations permettrait d’optimiser les stratégies thérapeutiques, les doses mais aussi les formules de ces médicaments. Porté par le docteur Pierre Marquet du CHU de Limoges, avec l’appui de l’Inserm et de la société Optim’Care, le projet Rexetris étudiera les relations entre l’exposition aux médicaments immunosuppresseurs et le devenir à long terme du patient greffé rénal et du greffon. Ces travaux mobiliseront les données de la base Cristal de l’Agence de la Biomédecine et de la base ABIS du CHU de Limoges.