

STAGE - CONSTRUCTION D'UNE LIBRAIRIE PERMETTANT D'EXPLOITER LA BASE PRINCIPALE DU SNDS H/F

Direction Data du HDH

Contrat – Stage (à partir de 6 mois)

Rémunération – 800€ / mois

Date de début – dès que possible

Localisation - Paris 15ème (télétravail partiel possible)

LE HEALTH DATA HUB

Comment **améliorer les dépistages** et faire en sorte que les patients soient pris en charge le plus tôt possible ?

Comment leur **proposer les meilleurs traitements** sur le long cours ?

Comment **appuyer les professionnels de santé** dans un contexte clinique qui se complexifie ou en cas de crise sanitaire ?

L'**Intelligence Artificielle** et les **données de santé** font partie de la réponse. Elles sont incontournables pour la **recherche et l'innovation en santé**, par exemple, pour prévenir des insuffisances cardiaques à partir de données issues d'appareils connectés, pour accélérer le dépistage du cancer du sein à partir d'analyses automatiques des examens de mammographies ou même pour réunir assez d'informations afin **d'améliorer la prise en charge** des maladies rares.

Et pour ça, la France a la chance de disposer de bases de données extrêmement riches !

Mais ces données sont souvent sous exploitées car dispersées. Grâce à des **solutions innovantes** telles que l'IA, l'objectif du Health Data Hub est justement de permettre **d'accéder de manière facilitée, unifiée, transparente et sécurisée** à un catalogue de bases de données de santé françaises.

Comment ?

Le Health Data Hub a mis en place une **plateforme technologique** qui met à disposition des porteurs de **projets d'intérêt public**, dans un environnement technologique sécurisé et à l'état de l'art, les données de santé **pseudonymisées** des français. Ces porteurs de projets vont mobiliser des sources de données très **volumineuses**, les croiser entre elles, et utiliser une puissance de calcul pour faire tourner des algorithmes de recherche complexes. Il s'agit par exemple de projets de start-up pour améliorer des logiciels d'aide au professionnel de santé, de projets permettant **d'améliorer la prise en charge des patients** en comparant l'efficacité de prise en charge, de projets portés par les administrations pour éclairer les politiques publiques.

Notre offre technologique, en constante évolution, peut être consultée [ici](#).

Les défis sont de taille pour traiter ces données de santé sensibles, volumineuses de natures et formats variables. La plateforme doit être un **levier d'innovation** dans l'écosystème de la donnée de santé français.

En résumé, avec le Hub, nous **accompagnons des porteurs de projets innovants** qui contribuent à trouver les solutions de demain pour améliorer la santé de tous les citoyens.

DESCRIPTIF DU POSTE

Direction des données :

Pour mener à bien les missions qui lui ont été confiées, le Health Data Hub a formé la direction des données dont les objectifs principaux sont de :

1. **Soutenir les projets d'intérêt public** que le HDH accompagne, aussi bien sur la compréhension des données de santé que sur leur exploitation via des experts des données de santé, des data scientists et des data engineers ;
2. **Gérer et mettre à disposition les données** qui lui sont confiées aux porteurs de projet au sein de la plateforme technologique du Health Data Hub ;
3. **Partager et mutualiser les outils et les connaissances** nécessaires à l'analyse des données de santé, dans le cadre d'une démarche open source ;
4. **Définir des stratégies novatrices** sur la gestion, l'exploitation et le partage de données de santé, permettant de réaliser la vision du HDH.

Au sein de la Direction des données, vous serez rattaché au pôle de "gestion des données" et aurez la mission de développer une librairie permettant d'exploiter la base principale du SNDS.

Mission du stage :

Le pôle "**Gestion des données**", de la Direction des données, développe les librairies de manipulation de données sur la plateforme du HDH, à usage interne (équipes du HDH) ou externe (chercheurs, médecins ayant un projet sur la plateforme).

Dans le cadre de ce stage, vous serez amené à préparer en particulier l'ingestion de la base principale du Système National des Données de Santé (SNDS), qui contient les données relatives aux remboursements de soins de ville réalisés par l'Assurance Maladie, aux hospitalisations et aux causes médicales de décès pour la quasi-exhaustivité de la population française. Ces données sont d'un intérêt majeur pour de nombreux projets de recherche médicale (exemple : étude des comorbidités des patients hospitalisés atteints de Covid-19).

L'objectif sera de développer une librairie logicielle permettant d'effectuer deux types d'opérations sur la base principale du SNDS :

- **réception et vérification automatisée des données sources reçues de la Caisse Nationale d'Assurance-Maladie**
 - validation du contenu des tables
 - mise en qualité (changement de format, renommage, stockage, etc.)

- **mise à disposition pour les projets de recherche**
 - appariement / ciblage : identification des patients et de leurs identifiants dans la base
 - extraction : sélection de la sous-partie de la base correspondant aux patients identifiés

Vous contribuerez au développement de cette librairie en suivant les bonnes pratiques de développement :

- **gestion des tâches via un backlog** (type Scrum - Kanban : user stories / technical stories)
- **usage de Git et merge reviews**
- **écriture de tests unitaires** (+ intégration continue)
- **génération de rapports** (logs structurés) lors de l'exécution des process

Ce travail sera réalisé en étroite collaboration avec les experts SNDS qui définissent les règles métiers de manipulation du SNDS.

PROFIL RECHERCHÉ

Compétences techniques

- Python (orienté objet)
- Git
- CI/CD
- Spark
- SQL
- Microsoft Azure (Azure Storage)
- Algorithmes et structures de données

Qualités relationnelles et personnelles

- Rigueur, clarté de présentation
- Sens du travail en équipe
- Force de proposition
- Curiosité intellectuelle

POURQUOI CHOISIR LE HEALTH DATA HUB ?

Vous êtes motivé(e) à rejoindre une équipe impliquée dans un projet ambitieux, qui a du sens et une finalité d'intérêt public ? Rejoignez-nous !

Notre structure récente, d'une cinquantaine de collaborateurs/trices, a besoin de talents créatifs, autonomes et proactifs pour continuer de grandir !

Ensemble, nous nous sommes engagés à :

- Accompagner les porteurs de projet visant à analyser les données de santé pour le bien commun.

- Construire et opérer une plateforme technologique pour leur offrir les meilleurs outils avec un très haut niveau de sécurité à respecter.
- Réunir et mettre en forme les données au plus grand potentiel pour la recherche et l'innovation.
- Promouvoir le partage des connaissances, des expertises et du savoir et diffuser une culture de la donnée de santé auprès de tous.

Bon à savoir:

💪 Rejoindre le HDH c'est surtout participer à un projet enrichissant humainement qui a du sens, avec un fort impact sociétal

🏆 Au HDH on favorise la prise d'initiative, dans une ambiance de challenge perpétuel

😎 Ici la bonne humeur et l'esprit d'équipe règnent

PROCÉDURE DE RECRUTEMENT

Après avoir postulé, le recrutement se déroulera de la manière suivante :

- Un premier entretien avec un membre de l'équipe (45 min)
- Une mise en situation à réaliser chez soi et un compte rendu des échanges
- Un entretien avec le manager basé sur la mise en situation (45 min)
- Une rencontre RH

Pour postuler

Retrouvez nous [ICI](#) sur Welcome to the Jungle (CV et Lettre de motivation)