

# STAGE - DATA SCIENTIST/ENGINEER - CARTOGRAPHIE DE L'UTILISATION DES BASES DE DONNÉES DE SANTÉ NATIONALES H/F

Direction Data du HDH

Contrat – Stage

Rémunération – 800€ / mois

Niveau Master

Date de début – dès que possible

Localisation - Paris 15ème (télétravail partiel possible)

## LE HEALTH DATA HUB

L'**Intelligence Artificielle** et les **données de santé** sont des outils incontournables pour la **recherche et l'innovation en santé**.

Malheureusement ces données sont souvent sous exploitées car dispersées. 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. Pour cela, nous avons 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. Notre offre technologique, en constante évolution, peut être consultée [ici](#).

Au travers de notre collaboration étroite avec l'écosystème de recherche français et international, nous avons aussi pour mission de relever les défis de l'open source pour les données de santé, de l'harmonisation et de l'interopérabilité de ces données à une échelle nationale et européenne.

## DESCRIPTIF DU POSTE

Vous serez rattaché à 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.

Dans le cadre de ce stage, vous aurez pour mission d'enrichir la [cartographie de l'écosystème](#) utilisant les bases de données françaises en santé. Cette cartographie est un **outil open source** d'exploration des publications scientifiques utilisant les bases de données de santé nationales. Elle vise à donner une vue d'ensemble de l'utilisation de ces bases par l'écosystème de recherche, via des dataviz concernant les publications, structures et chercheurs associés.

Un MVP (Minimum Viable Product) a été construit sur la base des publications utilisant uniquement **la base principale du SNDS (Système National des Données de Santé)**. Cette base est une base de données médico-administrative comportant principalement les remboursements de soins de ville et les hospitalisations pour la quasi-totalité de la population française depuis 2006.

Le but du stage est d'enrichir ce premier prototype au travers de divers axes. Pour cela, vous serez en charge :

- d'évaluer puis affiner la stratégie de récupération des publications utilisant le SNDS.
- de faire de la veille, harmoniser et enrichir les métadonnées des articles en utilisant des données Open source liées à l'écosystème de recherche récupérées via des datasets ou des APIs
- améliorer le moteur de recherche des publications
- d'étoffer le tableau de bord existant avec des indicateurs clés liés à l'enrichissement des métadonnées
- de généraliser la cartographie - pour le moment restreinte au SNDS - aux autres bases de données hébergées par le HDH

## PROFIL RECHERCHÉ

Pour ce stage, nous recherchons un.e candidat.e disposant d'une formation en data en science et/ou engineering, avec des compétences en data analysis et en NLP (traitement du langage naturel). Une connaissance ou une expérience dans le milieu de la recherche est un avantage.

### Technical skills


- Manipulation de données en Python
- Utilisation d'APIs
- Compétences de bases en NLP (pas d'utilisation de deep learning))
- Webscraping (BeautifulSoup, Selenium, ...)
- Dataviz (Plotly, Dash, ...)
- Notions de développement applicatif (React, JavaScript)
- Git
- Bibliométrie (optionnel)


### Soft skills

- Esprit d'analyse rigoureux

- Ouverture à l'autre : dialogue, écoute, appétence pour le travail en équipe pluridisciplinaire
- Curiosité intellectuelle
- Appétence pour le domaine de la recherche

## **POURQUOI CHOISIR LE HEALTH DATA HUB ?**

 Rejoindre le HDH c'est surtout participer à un projet ambitieux et 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 de notre première entrevue à envoyer par email
- Une rencontre RH

### **Pour postuler**

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