

Stage - Cartographie de l'écosystème SNDS

Direction Data du HDH

Contrat – stage de 4 à 6 mois, temps plein

Date de début – dès que possible

Niveau : Master

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 DE POSTE

Le stage se déroulera au sein de l'équipe data du HDH.

Le [SNDS](#) (Système National des Données de Santé) historique 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. De par son système de sécurité sociale, la France fait partie des quelques pays qui disposent de systèmes d'information médico-administrative couvrant l'ensemble du territoire et de la population, ce qui en fait une base extrêmement riche pour la recherche médicale.

La cartographie de l'écosystème SNDS est un outil [open source](#) d'exploration des publications scientifiques utilisant le SNDS. L'objet du stage est de construire une étude bibliométrique avancée de ces publications scientifiques, en particulier de produire des indicateurs reflétant l'état de la recherche sur le SNDS et son potentiel de réutilisation pour des projets de recherche. Plus précisément, le stage comportera les phases suivantes qui pourront faire éventuellement l'objet d'une publication scientifique:

- Design mathématique d'indicateurs (analyse sémantique, par thématiques, longitudinales, par types d'acteurs, etc)
- Implémentation des indicateurs
- Intégration des indicateurs au sein du [dashboard](#) déjà existant
- Participation à une étude menée par la CNIL et permettant de dresser un portrait global de la recherche utilisant des données de santé

PROFIL RECHERCHE

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

Compétences techniques

- Econométrie, statistiques
- Bibliométrie (optionnel)
- Manipulation de données en Python (préférentiellement) ou en R
- Notions en NLP (bibliothèques python tensorflow, pytorch, nltk, transformers, spacy)
- Webscraping (BeautifulSoup, Selenium,...)
- Dataviz (Plotly, Dash,...)
- Git
- Appétence pour le domaine de la santé publique

Qualités humaines et relationnelles

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

Postuler: CV et lettre de motivation sur welcome to the jungle [ICI](#)