

Health Data Hub

9 Rue Georges Pitard
75015 Paris

Stagiaire data management / biostatistique

Equipe Data

Contrat – Stage

Rémunération – à définir

Durée – 6 mois

Niveau – Master/ Licence

Date de début – Juillet ou Septembre 2021

LE HEALTH DATA HUB

Le Health Data Hub (HDH) ou « plateforme des données de santé » est un Groupement d'Intérêt Public dont l'objectif est de faciliter les travaux de recherche sur le patrimoine des données de santé. Il interagit avec de nombreux acteurs tels que les Ministères, la Caisse Nationale d'Assurance Maladie, les hôpitaux. La transparence et la participation des citoyens sont des engagements clés du HDH.

Le Health Data Hub c'est :

- Une plateforme technologique à l'état de l'art disponible pour héberger les premiers projets de recherche
- Une équipe pluridisciplinaire d'une cinquantaine personnes impliquées dans un projet ambitieux au service de l'intérêt général
- Près de 1 pétaoctet de données de santé traitées par an lorsque le HDH aura atteint son régime de croisière
- 28 projets pilotes accompagnés - Exemples :
 - Analyse du non recours au soin pendant le confinement à partir des résumés de passages aux urgences
 - Etude de l'impact et des traitements sur des données de vie réelle afin d'identifier les meilleurs schémas thérapeutiques pour le traitement du sarcome
 - Evaluation de l'apport de l'IA dans le dépistage organisé du cancer du sein

Le HDH porte une mission de facilitation de la recherche médicale, et de ce fait est un vecteur d'amélioration de la vie des patients. Au cœur des enjeux du numérique et de la santé, il est porté par une équipe dynamique qui contribue au partage de la connaissance à travers l'open source.

Plus d'informations sur le Health Data Hub et son équipe sur notre [site](#).

DESCRIPTIF DU STAGE

Au sein de l'équipe Data et en lien étroit avec la Fabrique, vous interviendrez principalement sur l'établissement du catalogue de métadonnées pour les bases qui seront disponibles via le Health Data Hub. Ce catalogue sera accessible à tous et s'inscrit dans la démarche science ouverte du Health Data Hub.

L'accès aux données de santé est réglementé en France et l'instruction d'une demande dure au minimum 6 mois. Comprendre le contenu d'une base de données est donc primordial afin de s'assurer qu'elle contient bien les informations nécessaires à la réalisation du projet de recherche.

Dans le cadre de sa démarche science ouverte et open source, le HDH propose en accès libre des informations, de la documentation, et des formations afin d'utiliser et comprendre les données plus facilement. En particulier, le HDH et ses partenaires (DREES, Santé publique France, CNAM, ANSM, ARS, ATIH, etc.) mettent à disposition une documentation ouverte collaborative comprenant des notebooks pour se former au fonctionnement du Système National des Données de Santé (SNDS), un générateur de données de synthèse, de nombreuses fiches thématiques regroupant les spécificités du SNDS, une section open data listant toutes les ressources accessibles sur le thème des données de santé, un dictionnaire interactif, ainsi que des programmes utilisés par les data scientist et statisticiens. Cette initiative de partage des connaissances est une première en France dans le domaine de la santé et est un succès puisque certaines ressources ont été reprises et adaptées pour d'autres objectifs par des institutions en santé françaises.

En plus de la documentation collaborative et du dictionnaire interactif, le HDH est en train de travailler sur la description du contenu d'une base de données: les métadonnées. La structuration des métadonnées pour des bases de données en Santé est un enjeu majeur dans la réutilisation des bases de données.

Le catalogue de métadonnées du HDH contient trois niveaux de métadonnées :

- le niveau base qui apporte des informations générales sur la base : description, le responsable de la collecte des données, le nombre de tables, le nombre de variables, la profondeur des données, les protocoles de qualité effectués sur les données etc.
- le niveau table qui apporte des informations pour chaque table de la base : le nom de la table, le nombre de lignes, la description de la table, les clés permettant de joindre deux tables entre elles, etc.
- le niveau variable qui apporte des informations pour chaque variable de chaque table : description, format, statistiques, indicateurs de qualités, etc.

Le stage aura pour objectif d'accompagner les responsables de données dans la mise en place du catalogue de métadonnées.

Plus particulièrement :

- compréhension de la structure des données
- alimentation du catalogue de métadonnées

- représentations graphiques d'indicateurs (distribution de l'âge, distribution homme-femme, etc.)
- calcul d'indicateurs de qualité (moyenne, écart-type, médiane, bornes, complétudes, plausibilité, etc.)

Vous contribuerez également sur les autres chantiers science ouverte de l'équipe Data :

- alimentation de la documentation collaborative sur le SNDS
- alimentation du dictionnaire interactif sur le SNDS
- développement de la démarche opensource du Health Data Hub

PROFIL RECHERCHÉ

Compétences techniques

- Maîtrise en Data management;
- Maîtrise du calcul de statistiques;
- Connaissance en Python ou R souhaitée;
- Capacité à naviguer et appréhender une base de données particulièrement complexe;
- Appétence pour le secteur médical;
- (optionnel) Maîtrise de Git.

Qualités humaines et relationnelles

- Motivation;
- Qualités d'organisation, de planification et de rigueur;
- Sens du service et du travail en équipe;
- Bon relationnel;
- Curiosité et capacité d'adaptation.

Durée du stage

Stage de 6 mois

Pour postuler

Contact : jobs.data@health-data-hub.fr