Stagiaire data engineer H/F
Direction Data du HDH

Equipe data
Contrat – stage
Rémunération – à définir
Date de début – entre janvier et mars 2022
Durée du stage - 6 mois

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 p**roposer 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, ou 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 ultra performante** permettant aux porteurs de projet d'intérêt public de

- mobiliser des sources de données très volumineuses,
- les croiser entre elles,
- d'utiliser une puissance de calcul assez grande pour faire tourner des algorithmes de recherche complexes.

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

DESCRIPTIF DU STAGE

Le stage se déroulera au sein de l'équipe data. En 2020, le Health Data Hub est un des lauréats de l'appel à projet Covid-19 organisé par EHDEN, le réseau européen des données de santé.
A travers ce nouveau partenariat, le Health Data Hub transforme des données médicales du SNDS (Système National des Données de Santé) vers un format qui a vocation à devenir une référence internationale : le format OMOP-CDM (Observational medical outcomes partnership - Common Data Model). La conversion des bases de données de santé vers le format OMOP-CDM renforcera leur interopérabilité et facilitera le partage d'outils, de méthodes, d'études et de programmes.

Le SNDS 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. De par son système de sécurité sociale, la France fait partie des quelques pays qui disposent de systèmes d'informations médico-administratives couvrant l'ensemble du territoire et de la population.

Aujourd'hui le SNDS, c'est plus de 180 tables, plus de 4500 variables, de nombreuses règles métiers, une évolution constante, 1.2 milliards de feuilles de soin par an, 11 millions de séjour hospitalier par an. Cela en fait une des plus grosses bases médico-administratives au monde : 250 To au total.

Le stage aura pour objectif d'accompagner l'équipe data dans ce projet de conversion des données SNDS sous le format OMOP-CDM à partir d'une première version du projet réalisée en 2021.
Plus particulièrement :

- Affinage de l'ETL
- Industrialisation de la transformation, en particulier gestion des mises à jour
- Contrôle qualité des résultats
- Documentation des données et métadonnées spécifiques à ces nouvelles tables
- Contribution ponctuelle aux outils spécifiques de l'équipe data (outil de création de données synthétiques, etc.)
- Aide à l'implémentation de standards et normes
- Harmonisation des transformations avec l'AP-HP, Bordeaux-BPE, etc.

PROFIL RECHERCHÉ

Compétences techniques

- Maîtrise en informatique et programmation ;
- Très bonne maîtrise du langage SQL ;
- Maîtrise du logiciel Python ;

- Connaissance des processus d'ETL et du data modeling (DBT)
- Connaissance du traitement des données massives et des technologies Big Data
- Capacité à naviguer et appréhender une base de données particulièrement complexe
- Appétence pour le secteur médical

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 ;

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 récente structure, 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

Pour postuler

Cv et lettre de motivation à jobs.data@health-data-hub.fr