

## Health Data Hub

9 Rue Georges Pitard  
75015 Paris

# Data Engineer

**Contrat – CDI**

**Rémunération – à définir**

**Date de début – dès que possible**

## 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 de plus de 60 personnes à haut potentiel, oeuvrant dans une structure récente 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 POSTE

En lien avec les producteurs de données, l'équipe data à la charge de la collecte, de la mise en qualité et de la gestion du catalogue de données. Elle prépare les jeux de données requis pour la réalisation des projets par leurs porteurs, requérant une expertise fine des données pour la création d'extractions pertinentes sur le plan médical. Elle réalise les chaînages lorsque plusieurs sources de données sont mobilisées et doit être capable d'en documenter la qualité. Pour industrialiser ces procédures, elle met en place une gouvernance de la donnée. Elle apporte son soutien aux producteurs dans le processus de préparation des données et notamment la déidentification et la création d'échantillons synthétiques. Elle élabore des services pour faciliter la compréhension des données disponibles au catalogue par les utilisateurs en mettant en place des outils de catalogue de métadonnées en ligne et de requêtage simple. Elle est force de proposition dans l'offre de formation et de développement d'outils à visée pédagogique sur le périmètre qui est le sien.

Au sein d'une équipe data d'une dizaine de personnes, le data engineer aura pour mission :

- d'opérer la plateforme technologique du HDH, tant pour la réception de bases au catalogue que pour la réalisation des extractions et appariements à destination des projets ayant accès à la plateforme. Cela inclut une forte composante opérationnelle, et notamment:
  - implémentation, mise en production et exécution de programmes de contrôle de la qualité, de l'intégrité et de la sensibilité des données réceptionnées
  - réalisation des extractions et des appariements entre différentes sources de données
  - appui, suivi et documentation des projets accédants ou contribuant au catalogue de bases
- de contribuer aux divers chantiers internes de l'équipe, entre autres :
  - implémentation d'un catalogue de métadonnées, à destination des producteurs et consommateurs des données du catalogue
  - implémentation d'un pipeline de transformation des bases vers un format international (OMOP)
  - implémentation et travail de fond sur des problématiques de déidentification, d'anonymisation et de génération synthétique de données
  - définition de l'offre de service en terme d'outils à déployer sur la plateforme technologique
- de supporter et d'appuyer les autres équipes sur des problématiques transverses: cadrage technique des projets accédant à plateforme, participation et restitution des travaux du HDH lors de meetup, soutien à l'organisation de data challenge, etc.

## PROFIL RECHERCHE

### Expérience et Compétences requises

- Confortable dans un ou plusieurs langages de programmation et de scripting
- Bonne connaissance opérationnelle en SQL ou librairie de traitement de données
- Capacité à concilier des compétences en analyse fonctionnelle des données avec une expertise et flexibilité technique
- Capacités rédactionnelles
- Capacité à travailler en équipe multidisciplinaire
- Bon relationnel - apte à gérer une relation avec des partenaires du HDH
- Bonne capacité d'autonomie et force de proposition - apte à évoluer dans un environnement en forte évolution et croissance
- Expérience dans un framework de traitement distribué (e.g. spark) est un plus
- Expérience en informatique ou recherche médicale est un plus

### Pour postuler

Contact : [jobs@health-data-hub.fr](mailto:jobs@health-data-hub.fr)