



Contexte d'activités et organisation de la cohorte :

La cohorte Constances est un échantillon de 220 000 volontaires âgés de 18 à 69 ans à l'inclusion (www.constances.fr/). L'inclusion des volontaires se fait dans des Centres d'Examens de Santé. Le suivi longitudinal, sans limitation de durée, s'effectue : (1) par un auto-questionnaire annuel et une invitation régulière dans les Centres pour des examens biomédicaux et (2) dans les bases de données administratives nationales socioprofessionnelles et de santé. En plus de ce suivi, des données d'enrichissement via des questionnaires spécifiques (problèmes dermatologiques, expositions aux rayonnements ionisants...) et des appariements à des données d'expositions professionnelles ou environnementales.

La cohorte est structurée en fonctions, soutenues par un cadre de gouvernance des données. Un comité de projet (COPROJ) supervise le projet, garantissant une approche unifiée et stratégique. Les composantes clé de l'organisation comprennent : i) le pôle stratégie, animé par la direction scientifique et technique, établit les principales orientations et la vision stratégique. Il repose sur des fonctions supports, la communication et la gestion administrative et financière. ii) le pôle gisement de données comprend trois fonctions axées sur la collecte de données provenant de trois sources principales : questionnaires, centres d'examen de santé et liens avec les bases de données nationales. iii) le pôle méthode et traitement des données est dédié, au maintien de la qualité des données, à l'enrichissement continu de la base de données avec de nouvelles données, au calcul des scores de pondération et à la mise en œuvre d'algorithmes d'identification. iv) le pôle accompagnement des projets est spécifiquement destiné à organiser l'ouverture des données aux utilisateurs externes. v) le pôle Infrastructure numérique et gestion des flux de données.

La base de données de Constances comprend plus de 4000 variables, des centaines de tables, plusieurs espaces de stockage et l'alimentation de ces bases est orchestrée par des flux automatisés ou semi-automatisés selon le périmètre des données traitées.

Missions et responsabilités :

Au sein du pôle Infrastructure numérique et gestion des flux de données, il (elle) aura pour missions principales la gestion de la base de données Constances sur son périmètre, le maintien et le développement de l'interface d'analyse et de pilotage de la cohorte.

Activités principales :

1. Gestion de la base de données :

- Responsable de l'intégration des données de son périmètre
 - Mise en place des outils pour l'intégration des données ;
 - Suivi des rejets ;
 - Documentation des programmes et process mis en place ;
 - Amélioration continue des process et programmes.



Cohorte Constances

Fiche de poste

DATA ENGINEER

- Evolutions / corrections de la base de données en fonction des demandes métiers (nouvelles variables, nouvelles tables...)
- Participation à la migration des programmes d'intégration de données de SAS en Python
- Pilotage de l'automatisation des pipelines sous Prefect

2. Responsable de l'interface d'analyse et de pilotage de la cohorte (la DataApp)

Cette interface, développée sous Streamlit, centralise les flux de données issus du DatalaLake (MariaDB) et d'APIs diverses.

- Maintenir la fluidité de l'outil et corriger les bugs
- Intégrer des nouvelles sources de données
- Être force de proposition sur de nouvelles utilisations

3. Participation à des projets divers incluant le SI : la gouvernance de la donnée, projet de documentation de métadonnées...

Compétences :

- Maîtrise de **Python** et de son écosystème data (*Pandas, NumPy*), la connaissance de **SAS** serait un plus
- Bonne maîtrise des systèmes de gestion de bases de données (*MySQL, MariaDB*)
- Expérience dans la gestion de notre outil d'orchestration de workflows (**Prefect**)
- Compétences en **data visualisation** (conception de tableaux de bord, restitution claire des données)
- Maîtrise des outils de **versioning** (Git)
- Bonne connaissance des environnements **Windows et Linux**
- Sensibilité aux enjeux de **sécurité des données** et des systèmes

Savoir-être :

- Respect strict de la **confidentialité des données** et de leur bonne utilisation
- Esprit d'initiative, **force de proposition** et proactivité
- Excellentes capacités de **communication**, y compris avec des interlocuteurs non techniques
- Rigueur, sens de l'organisation et méthodologie
- Bon relationnel et esprit d'équipe

Nature du contrat :

- Contrat de la fonction publique
- 3 à 5 ans d'expérience
- Salaire selon expérience
- CDD d'une durée de 24 mois renouvelable à temps plein
- Lieu de travail : UMS 11, Hôpital Paul Brousse, Villejuif
- Engagement de confidentialité à prévoir