# PROFIL DE POSTE

## *Ingénieur Data*

## *Entrepôt de données de santé*

- ✓ **CDI à temps plein**

**DIPLÔME REQUIS :**

- ✓ **Bac +5 en informatique**

- ✓ **Qualités relationnelles et travail équipe pluridisciplinaire**
- ✓ **Rigueur et organisation**
- ✓ **Capacité d'adaptation et prise d'initiative**

- ✓ **Participation à un projet structurant d'exploitation des données médicales à des fins de recherche**
- ✓ **Travail en réseau avec les autres CHU impliqués**


**PERSONNES À CONTACTER :**

**M. William BIENVENU**
william.bienvenu@chu-st-etienne.fr

**Mme Sarah DEVIDAL**
sarah.devidal@chu-st-etienne.fr


## Présentation de l'établissement

Le CHU de Saint-Étienne est situé au cœur de région Auvergne Rhône Alpes, deuxième région de France. Il est l'établissement support du GHT Loire, un des plus grands GHT de France en nombre d'établissements. Le CHU irrigue son territoire au travers un budget de plus de 730 millions d'euros. Il est le 20ème CHU des 32 CHRU de France.

Les activités du CHU sont articulées autour de 11 pôles et 68 services, disposant de 1.920 lits et places, répartis sur ses 3 principaux sites. Les 8.350 professionnels de ces services prennent en charge annuellement plus de 170.000 patients en hospitalisation et séances et plus de 500.000 en consultations.

## Enjeux du poste

Le CHU de Saint-Etienne, dans le cadre d'un groupement avec les 3 autres CHU de la région Auvergne-Rhône-Alpes, déploie un entrepôt de données de santé qui vise à valoriser les données cliniques, biologiques et médico-administratives issues des systèmes d'information hospitaliers pour soutenir la recherche, l'évaluation des pratiques et l'innovation en santé.

L'équipe dédiée au sein de la DSN sera responsable de concevoir, mettre en place et maintenir les flux d'alimentation de l'EDS, afin de garantir des données fiables, accessibles et directement exploitables par les Data Scientists du pôle Recherche et Innovation.

## Activité et profil recherché

**Dans ce cadre, nous recherchons un ingénieur Data qui intégrera l'équipe "Interopérabilité et Développement transversaux". Cette personne sera responsable de la conception, du développement et de la maintenance des "pipelines de données" de l'EDS.**

**En collaboration avec les Data Scientists, elle devra identifier les différentes données à collecter, les extraire, les nettoyer, les transformer et les stocker dans l'EDS en veillant à leur qualité et à leur intégrité.**

**Missions:**

- **Analyser les différentes sources de données métier** afin d'en comprendre les spécificités et d'identifier les contraintes d'intégration à l'EDS: format, volumétrie, fréquence d'alimentation, usages possibles
- **Collaborer avec les équipes projets de recherche** pour comprendre les besoins en jeux de données et apporter une expertise technique sur les sources de données et la chaine de traitement
- **Concevoir, développer et documenter les pipelines de traitement** (collecte, modélisation, transformation, stockage) en s'appuyant sur des outils ELT adaptés aux contraintes du SI hospitalier
- **Maintenir et optimiser les chaines de traitement:** optimisation des requêtes SQL, amélioration des performances, gestion de l'orchestration des traitements
- **Garantir la sécurité et la confidentialité des données** en appliquant les réglementations en vigueur pour les données de santé (RGPD, HDS)

## Profil recherché

### *Formation*

Diplôme d'ingénieur ou Master 2 en informatique avec spécialisation Big Data

### *Compétences techniques*

- Maîtrise d'une solution ELT (idéalement **DBT**) et d'un outil d'orchestration de worflows (**Dagster**)
- Excellente connaissance des bases de données SQL (**SQL server, Oracle**), des bases colonnaires (**ClickHouse**) et des moteurs de recherche textuelle (**ElasticSearch, OpenSearch**)
- Maîtrise de **Python** et du **Shell** pour le développement de script et d'outils de traitement
- Compréhension des standards de données de santé (**OMOP, FHIR, CDISC** ...)
- Notions sur les environnements **Kubernetes** et l'exploitation de clusters
- Pratique d'autres langages de programmations (**R, Java, Scala**)
- Maîtrise des outils de gestion de version (**GitLab, GitHub**)
- Connaissance de la réglementation sur les données personnelles et les principes de cybersécurité
- Maîtrise de l'anglais technique

## Expérience souhaitée

Expérience minimum de 2 ans sur un poste d'ingénieur Data, idéalement dans le domaine de la santé ou de la recherche clinique.

Une expérience de participation à des projets d'entrepôt de santé est un atout.

## Conditions d'exercice

- Poste à temps plein, basé au CHU de Saint Etienne (SLAT Bellevue)
- Rattachement hiérarchique à la Direction des Systèmes Numériques
- Collaborations régulières avec des chercheurs hospitaliers, des partenaires extérieurs, les ingénieurs informatiques et les juristes de l'établissement
- Possibilité de télétravail selon la politique de l'établissement
- Rémunération selon l'expérience et la grille de rémunération de l'établissement

**PERSONNES À CONTACTER :**

**M. William BIENVENU**
william.bienvenu@chu-st-etienne.fr

**Mme Sarah DEVIDAL**
sarah.devidal@chu-st-etienne.fr