
RECRUTE :

DATA INGÉNIEUR - ENTREPÔTS DE DONNÉES DE SANTÉ ET DONNÉES DE VIE RÉELLE
H/F – CDI 100%

Service Data Factory & Analytics – Direction de la Recherche et des Data - Site St Herblain ou Angers

Recrutement n° 3251

MOTIF

Création de poste

L'Institut de Cancérologie de l'Ouest (ICO) Nantes-Angers est un centre d'excellence dans la prise en charge du cancer et la recherche. Labelisé *Comprehensive Cancer Center (CCC)* par l'Organisation of European Cancer Institutes (OECI), l'ICO est le 1^{er} centre de lutte contre le cancer de province en France et le 3^{ème} au niveau national après l'Institut Gustave Roussy (Villejuif) et l'Institut Curie (Paris) en termes de file active de patients et d'inclusion dans les essais cliniques. Son activité de radiothérapie est par ailleurs la plus importante de France avec 90.000 séances par an.

Depuis 5 ans, l'ICO a placé les **données de vie réelle** (Real-World Data / RWD) et l'IA au centre de sa stratégie de recherche et d'innovation. Son service Data Factory & Analytics est une structure pluridisciplinaire, intégrée, dédiée à la collecte, à la qualification, à l'analyse et à la valorisation des données générées à l'occasion du soin de nos patients. Le service assure la maîtrise complète du cycle de vie des données depuis leur production lors de la prise en charge des patients au sein de l'ICO jusqu'à leur analyse et leur valorisation. Il pilote la construction, l'alimentation et la qualification de plusieurs **entrepôts de données (EDS)**, infrastructures et programmes nationaux : l'EDS local de l'ICO ayant vocation à permettre la réutilisation des RWD à des fins de recherche et de pilotage de l'activité de l'organisation ; le programme national CONSORe (Continuum Soins Recherche) en lien avec la fédération des 19 centres de lutte contre le cancer en France (Unicancer) ; l'EDS national OncoDS (Onco Data Share), lauréat de l'appel à projet du Health Data Hub 2023 dédié à l'accompagnement et soutien à la constitution d'EDS hospitaliers) ; l'EDS inter-régional, le Ouest Data Hub en lien avec les CHU de Rennes, Brest, Nantes, Angers, Tours et Orléans (GCS HUGO – Hôpitaux Universitaires du Grand Ouest) ; et OPTIMA, première plateforme européenne de recherche en oncologie utilisant l'IA.

Le service est actuellement constitué d'une équipe de 13 personnes ayant des profils complémentaires : 2 data analysts RWD, 1 biostatisticien RWD, 3 data scientists RWD, 1 data manager RWD, 2 attachés de recherche sur données RWD, 1 interne en pharmacie, 1 chef de projet RWD, 1 responsable des opérations et 1 chef de service pharmacien PU-PH.

MISSIONS

Sous la responsabilité du chef de service Data Factory & Analytics qui est une des composantes de la Direction Recherche et Data de l'ICO et en lien fonctionnel avec le responsable des opérations, l'ingénieur data jouera un rôle central dans la **gestion, le traitement, la transformation, l'exploitation, l'analyse et la valorisation des données de santé du Système d'Information (SI) de l'ICO**. .

PRINCIPALES ACTIVITÉS

Le Data Ingénieur assure les missions au sein de l'équipe du service Data Factory & Analytics et en collaboration les autres services de la Direction Recherche et Data (Recherche Clinique, Recherche Translationnelle, Unité IA, ...) ainsi que la Direction des Systèmes d'Information (DSI), la Direction du Développement et de l'Innovation (DDI), les experts métiers et cliniciens de l'ICO :

- **Entrepôts de données de santé (EDS) de l'ICO :**

- ✓ Participer activement au mapping et à la maîtrise du patrimoine de données de l'ICO : connaissance (modèles) des bases de données sources existantes, répertorier les tables et champs d'intérêts, en lien avec les éditeurs et les référents applicatifs et métiers.
- ✓ Participer à la constitution de l'EDS en lien avec la DSI et des experts métiers de la Direction Recherche et Data :
 - Participer à l'élaboration des prérequis sur l'architecture du SI Data de l'ICO (gouvernance, référentiels, modèles métiers).

- Concevoir l'architecture fonctionnelle de l'EDS (intégration, qualité, sécurité, transformation, persistance, consommation) en lien avec la DSI.
- Définir avec la DSI, les traitements d'ingestion de données à mettre en place.
- Automatiser les processus de contrôle qualité (avec orchestrateur) et de validation des données (gestion des incohérences, exhaustivité, format, ...), rédaction des rapports de qualification et documentation associée.
- Suivre les applications dans le temps (paramétrage, maintenance corrective et évolutive)
- Concevoir l'architecture technique de l'EDS (choix des composants au regard des contraintes de la Direction Recherche et Data).
- ✓ Produire la documentation et les méta-données associées à l'EDS (catalogue de données, spécifications, nomenclature, guides méthodologiques et d'utilisation).
- ✓ Garantir la qualité, la sécurité des données recueillies selon les obligations législatives et réglementaires en vigueur.
- ✓ Garantir l'intégrité et la cohérence des données issues de sources variées (applications métier, bases externes et internes).
- ✓ Développer un dashboard Key Performance Indicators (KPI) pour l'EDS (graphiques, tendances temporelles et filtres interactifs pour explorer les données de vie réelle).
- **Autres projets Data : expérimentations d'innovations, IA, outils d'exploitations des données**
 - ✓ Définir / préciser les besoins exprimés par les acteurs métiers.
 - ✓ Extraire, traiter et transformer des données complexes.
 - ✓ Concevoir, intégrer et maintenir des pipelines d'exploitation de données non-structurées utilisant des techniques hybrides de type : RegEx, NLP ou encore LLM (Mistral).
 - ✓ Contribuer avec la DSI à la définition des besoins d'infrastructures avec les acteurs métiers pour rédiger les spécifications.
 - ✓ Maintenir les développements réalisés par l'équipe.
 - ✓ Garantir la documentation associée aux outils et fonctionnalités développés
- **Autres**
 - ✓ Être force de proposition en termes d'évolution des environnements IT & Data de l'ICO.
 - ✓ Assurer une veille technologique : actualité, nouveauté, applicatifs, outils, progiciels, matériel, etc.
 - ✓ D'autres missions peuvent être confiées selon le niveau d'autonomie et l'actualité de l'hôpital.

DIPLÔMES ET COMPÉTENCES REQUISES

- **Formation et qualifications :**
 - ✓ Bac +5, diplôme d'ingénieur en Data Engineering, data science ou équivalent.
 - ✓ 5 ans d'expérience minimum dans le domaine des bases de données, du data engineering, des entrepôts de données structurées et non structurées.

- **Compétences techniques :**

- ✓ Expert en langage : Python, R, SQL.
- ✓ Connaissance des systèmes de gestion de bases de données (PostgreSQL, Oracle, MySQL, etc.).
- ✓ Bonne compréhension des architectures informatiques (infrastructure, réseaux, sécurité)
- ✓ Maîtrise des outils afférents au poste : ex. Git, , Airflow, Docker, elastic, etc.
- ✓ Solides connaissances des réglementations sur les données de santé (RGPD, CNIL, HDS).
- ✓ Seraient un plus :
 - Connaissance et utilisation de LLM,
 - Expérience dans le traitement de données médicales et connaissance des standards de données (FHIR, OMOP, SNOMED, ICD...).
 - Connaissance des logiciels de gestion de données (RedCap, Ennov Clinical, ...)
 - Notions en cancérologie (pathologies cancéreuses, terminologie, médicaments).
- ✓ Maîtrise de l'anglais écrit et oral.

- **Compétences personnelles :**

- ✓ Rigueur et esprit analytique.
- ✓ Autonomie, rigueur, sens de l'organisation et respect des délais.
- ✓ Capacité à travailler en équipe multidisciplinaire et à communiquer avec des profils non techniques.

CONTRAT ET STATUT

- Prise de poste : au plus tôt.
- Type de contrat : CDI 100 %.
- Poste sur St Herblain ou Angers.

CONTACT

Merci d'adresser, **au plus tôt**, votre candidature à
La Direction des Ressources Humaines – INSTITUT CANCEROLOGIE DE L'OUEST
par mail : srh.recrutement@ico.unicancer.fr