

## Fiche de poste

**Fonctions : Ingénieur-e en science des données (data scientist) (H/F)**

**Métier ou emploi type\* : ingénieur d'études**

E2D46 - Ingénieur-e statisticien-ne

\* REME, REFERENS, BIBLIOPHILE

### Fiche descriptive du poste

**Catégorie : A**

**Corps :** Ingénieur d'études

### Affectation

**Géographique :** Centre de Recherche en Epidémiologie et Santé des Populations (CESP) UMR Inserm 1018, Equipe, Epidémiologie Clinique, Bâtiment 15/16, 16 avenue Paul Vaillant Couturier, Hôpital Paul Brousse, Villejuif (94).

**Employeur :** Université Paris Saclay – Faculté de médecine - 63 rue Gabriel Péri – 94 276 Le Kremlin Bicêtre.

### Contexte

L'équipe « Epidémiologie Clinique » du CESP (Unité Inserm U1018) recherche un(e) ingénieur-e data-scientist, pour une durée d'1 an renouvelable. La personne recrutée intégrera le groupe de coordination de la cohorte clinique CKD-REIN <https://francecohortes.org/cohortes-associees/ckd-rein-chronic-kidney-disease-rein>

Lancée en 2012, l'étude CKD-REIN a inclus 3 033 patients atteints de tous types de maladie rénale chronique qui ont été suivis pendant 5 ans. L'objectif de cette étude, la première de cette ampleur en France, est de répondre aux nombreuses questions des patients, des médecins et des scientifiques sur les causes, les facteurs de progression et les conséquences de cette maladie afin d'identifier et de promouvoir les pratiques médicales et l'organisation des soins les plus efficaces pour en ralentir la progression et prévenir ses complications. Elle a aussi pour ambition de comprendre ce qui est important du point de vue des patients.

*Le CESP dirigé par le Pr Bruno Falissard, regroupe près de 600 chercheurs, ingénieurs, techniciens et doctorants travaillant dans la plupart des domaines de la santé, comme les maladies chroniques, le cancer et la santé mentale, répartis dans 11 équipes de recherche. L'équipe Epidémiologie Clinique dirigée par le Pr Laurence Meyer mène des recherches centrées sur l'épidémiologie et l'évaluation de stratégies de prévention et de traitement appliquées aux maladies rénales et cardiovasculaires, à l'infection à VIH, et aux maladies chroniques de l'enfant avec pour objectif la réduction de leur impact sur l'espérance de vie, la qualité de vie et les ressources de santé.*

### Missions / Activités

Vous travaillerez en étroite collaboration avec le groupe projet CKD-REIN et le service support aux cohortes de l'Inserm (UMS France Cohorte).

#### Missions :

- Etre responsable de l'intégration des données structurées ou non structurées dans l'entrepôt de France Cohorte ;
- Assurer la mise en qualité des données et leur mise à disposition des utilisateurs finaux (statisticiens et chercheurs) ;
- Documenter les variables et participer à la création des métadonnées nécessaires à la mise en conformité de la cohorte avec les principes FAIR.

#### Activités principales :

- Effectuer la transformation et le nettoyage des données sources (ETL) dans l'entrepôt en prévision des analyses statistiques ;

<ul style="list-style-type: none"> <li>• Référencer et documenter les traitements réalisés dans des rapports ;</li> <li>• Elaborer des programmes permettant d'automatiser au maximum les traitements.</li> </ul> <p><b>Activités associées :</b></p> <ul style="list-style-type: none"> <li>• Communiquer directement avec les utilisateurs / destinataires finaux des données ;</li> <li>• Produire des tableaux de bords et des rapports de synthèse incluant statistiques de base et graphiques à destination des non-experts ;</li> <li>• Participer aux réunions avec le groupe projet constituée de chercheurs en épidémiologie, de cliniciens, de statisticiens, d'informaticien, etc.</li> </ul> <p>A noter que les données recueillies pour l'étude CKD REIN sont de nature très variée : données cliniques et biologiques, données -omiques, données issues d'auto-questionnaire patient, données alimentaires, environnementales ou encore médico-administratives, données textuelles. La cohorte est chaînée avec le Système National des Données de Santé (SNDS).</p> <p>Le poste offre des possibilités d'évolution intéressante avec une participation possible aux projets de recherche incluant par exemple : l'analyse statistique des données du SNDS, le développement d'algorithmes prédictifs à partir de données massives (dont données textuelles), la génération de données synthétiques, l'intégration de pipeline de bio-informatique.</p> <p><b>Conditions particulières d'exercice (Logement, Horaires spécifiques, NBI, PFI, etc...) :</b> Aucune  <b>Encadrement :</b> NON  <b>Conduite de projet :</b> NON</p>
---

Compétences*
<p><b>Connaissances, savoir :</b></p> <ul style="list-style-type: none"> <li>• Maîtrise du langage SQL et d'au moins un système de gestion de base de données relationnelle</li> <li>• Maîtrise d'un des langages de scripts utilisés en biostatistique (R et/ou Python)</li> <li>• Connaissance souhaitée du langage de programmation SAS : base, macro-langage, proc SQL (possibilité de formation)</li> <li>• Connaissance appréciée d'un ou plusieurs outils de recueil de données cliniques (REDCap ou équivalent)</li> <li>• Appétence / intérêt pour les bases de données en santé et les principes du FAIR</li> <li>• Anglais : niveau B1 minimum requis</li> </ul>
<p><b>Aptitudes, savoir-être :</b></p> <ul style="list-style-type: none"> <li>• Sens de la communication, du travail en équipe et des relations partenariales ;</li> <li>• Autonomie, capacité à organiser, rigueur ;</li> <li>• Esprit analytique et de synthèse</li> <li>• Capacité à restituer le travail réalisé</li> <li>• Obligation de respecter le secret statistique et professionnel dans le cadre législatif existant</li> </ul>

\* Conformément à l'annexe de l'arrêté du 18 mars 2013 (NOR : MENH130559A)

CONDITIONS
<p><b>Prise de fonction :</b> Dès que possible</p> <p><b>Durée et type du contrat :</b> CDD Université Paris-Saclay de 1 an, renouvelable</p> <p><b>Formation / Expérience :</b> De formation supérieure (de type Master) en science des données, informatique décisionnelle ou gestion de base de données cliniques ou équivalent / Débutant-e accepté-e</p>

**Rémunération :**

*Rémunération selon diplômes et expérience.*

**Avantages :**

**Quotité du temps de travail :** 100%.

**Télétravail possible :** OUI, maximum 2 jours par semaine après période d'intégration et hors contexte sanitaire.

**Nombre de postes à pourvoir :** 1

Toute candidature doit comprendre un CV détaillé et d'une lettre de motivation. Elle devra être adressée à [ckdrein@inserm.fr](mailto:ckdrein@inserm.fr)

**Date limite de candidature :** 30 juin 2024

**Au moment de la prise de fonction :**

Date et Visa du Supérieur hiérarchique :

*(nom-prénom – Qualité)*

Date et visa de l'agent :

*(avec mention : « vu et pris connaissance »)*