

Ingénieur.e en science des données

 **Début : Dès que possible**

 **Paris**

 **Télétravail partiel**

 **Bac +5**

L'Inserm est le seul organisme public français entièrement dédié à la recherche biologique, médicale et en santé des populations. Il dispose de laboratoires de recherche sur l'ensemble du territoire, regroupés en 12 Délégations Régionales. Notre institut réunit 15 000 chercheurs, ingénieurs, techniciens et personnels administratifs, avec un objectif commun : améliorer la santé de tous par le progrès des connaissances sur le vivant et sur les maladies, l'innovation dans les traitements et la recherche en santé publique.

Rejoindre l'Inserm, c'est intégrer un institut engagé pour la parité et l'égalité professionnelle, la diversité et l'accompagnement de ses agents en situation de handicap, dès le recrutement et tout au long de la carrière. Afin de préserver le bien-être au travail, l'Inserm mène une politique active en matière de conditions de travail, reposant notamment sur un juste équilibre entre vie personnelle et vie professionnelle.

L'Inserm a reçu en 2016 le label européen HR Excellence in Research et s'est engagé à faire évoluer ses pratiques de recrutement et d'évaluation des chercheurs.

Emploi

Poste ouvert aux candidats

- Agents fonctionnaires de l'Inserm par voie de mobilité interne
- Agents fonctionnaires non Inserm par voie de détachement
- CDD agents contractuels

Catégorie

A

Corps

IE-Ingénieur d'études

Emploi-Type

Ingénieur.e en traitement de données (data scientist)

Structure d'accueil

Unité

UMS 47 France Cohortes

A propos de la Structure

L'Unité mixte de service (UMS) France Cohortes associe plusieurs tutelles académiques dont l'Inserm, et s'inscrit dans un contexte évolutif de déploiement de services, au bénéfice des cohortes épidémiologiques. Ces services mutualisés sont proposés aux cohortes épidémiologiques et aux équipes de recherches de l'Inserm dans leurs projets d'études sur les données du Système national des données de santé (SNDS) sur les aspects réglementaires, de système d'information, les outils de production, de diffusion et d'exploitation de données.

Dans une démarche de science ouverte, les données produites doivent suivre les recommandations FAIR, et en particulier faire l'objet d'une documentation permettant la meilleure information auprès des utilisateurs potentiels pour valoriser au mieux les données, dans le respect de l'intégrité scientifique des études en santé publique. Cette démarche s'inscrit également dans le cadre de projets européens en cours visant à faire des recommandations de bonnes pratiques en la matière.

Directeur

Grégoire Rey

Adresse

48-50, rue Albert 75013 Paris

Description du poste

Mission principale

Vous aurez pour rôle principal d'accompagner les data manager des cohortes et de mettre en œuvre la documentation des données des cohortes prises en charge par France Cohortes, pour permettre une valorisation optimale de ses données via son catalogue, et tous les catalogues associés.

Vous aurez également pour mission de faire l'interface avec les équipes de recherche pour collecter les jeux de données susceptibles d'être clonées, de réaliser les traitements statistiques de ces jeux afin de générer des données synthétiques, et de réaliser cette opération, soit directement, soit en lien avec des prestataires. Vous devrez également mettre ces jeux de données synthétiques à dispositions sur un site web dédié.

Vous contribuez ainsi à la stratégie de science ouverte portée par France Cohortes.

Activités principales

- Documenter le contenu des variables des bases (attribution de labels, recodage des valeurs manquantes)
- Collecter les métadonnées sur les enquêtes de cohorte (objectifs, champ, méthodologie, taux de réponses, etc.) et les documents associés
- Evaluer les moyens nécessaires pour faire évoluer les documentations de chaque cohorte participante vers la cible retenue et les accompagner dans cette démarche
- Structurer cette information selon la norme DDI (Data Documentation Initiative), le format requis pour la mise à disposition des données
- Assurer la publication des méta-données sur le catalogue de France Cohortes
- Rédiger des notes et présenter le travail effectué
- Mettre en place et optimiser les procédures de contrôle des données
- Participer à la conception et au choix du générateur de données synthétiques
- Générer des jeux de données synthétiques des cohortes pour l'open data
- Participer au développement de logiciels spécifiques avec des prestataires
- Vérifier que les caractéristiques de la distribution statistique des données sont conservées

Spécificité(s) et environnement du poste

- Obligation de respecter le secret statistique et professionnel dans le cadre législatif existant
- Relations externes fréquentes et nombreuses (établissements de recherche, porteurs de projets, experts métier)

Connaissances

- Connaissance approfondie en épidémiologie, notamment dans la production de données, savoir reconnaître et expliquer les différents protocoles d'étude
- Connaissances des principes FAIR data
- Des connaissances sur l'utilisation des métadonnées et de leur représentation seraient un plus : normes et standards DDI, DCAT, adaptés à la production, à l'exploitation, au stockage et à la diffusion des données médicale, scientifiques et/ou statistiques
- Maîtrise des langages et schémas RDF, XML et/ou Json
- Connaissance du traitement de données et du data management
- Connaissance d'un ou plusieurs logiciels de traitement de données (SAS, R ou Python)
- Connaissance des langages de programmation
- Maîtrise des outils et méthodes de génération de données

Savoir-faire

- Découvrir et s'adapter à des concepts et méthodes statistiques innovants y compris de data sciences
- Interagir avec des épidémiologistes, des statisticiens et des informaticiens
- Traiter des données, construire et gérer une base de données
- Rédaction de documentation technique et méthodologique (rapport, article scientifique)
- Capacité à travailler en anglais indispensable

Aptitudes

- Curiosité intellectuelle
- Capacité de conceptualisation et de raisonnement analytique

- Aisance relationnelle
- Capacités d'organisation, de planification et de rigueur
- Discrétion et confidentialité

Expérience(s) souhaité(s)

- Une expérience dans le domaine de la manipulation et la documentation de données serait appréciée

Niveau de diplôme et formation(s)

- Bac + 5 de type Master en science des données, statistique, mathématiques appliquées ou épidémiologie

Informations Générales

Date de prise de fonction

Janvier 2024

Durée

CDD de 12 mois

Renouvelable : OUI NON

Temps de travail

- Temps plein
- 38h30 par semaine

Activités télétravaillables

OUI * NON

* 2 à 3 jours/semaine maximum sur accord du responsable hiérarchique.

Rémunération

- Entre 2 494 € et 3 640 € bruts mensuels (négociable) en fonction de votre profil et expérience professionnelle sur des postes de niveau équivalent

Avantages

- Congés Annuels et RTT : 45 jours
- Restauration collective subventionnée sur place
- Comité d'action sociale (prestations sociales, culturelles, sportives)
- Transports publics remboursés partiellement

Modalités de candidature

Date limite de candidature

Modalités de candidature

- Envoyer CV et lettre de motivation à : **Recrutement.FC@inserm.fr** avec la référence IE-SDD-102023
- Préciser vos prétentions salariales

Pour en savoir +

- Sur France Cohortes : <https://francecohortes.org>
- Sur l'Inserm : <https://www.inserm.fr/> ; site RH : <https://rh.inserm.fr/Pages/default.aspx>