

## Ingénieur.e en science des données

 CDD 1 an renouvelable  Début : Juillet 2021  Villejuif  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 des 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

<b>Catégorie</b>	A
<b>Corps</b>	IE
<b>Emploi-Type</b>	Ingénieur statisticien (E2D46)

### Structure d'accueil

**Unité** France Cohortes US47 – Cellule CESP

#### A propos de la Structure d'accueil

L'Unité mixte de service (UMS) France Cohortes vise à apporter des services mutualisés aux cohortes épidémiologiques sur les aspects réglementaires, de système d'information, et sur les outils, méthodes et standards adéquats de production, de diffusion et d'exploitation de données, en suivant une démarche de science ouverte. France Cohortes associe plusieurs tutelles académiques, et est structurée sous forme d'une cellule centrale coordonnant plusieurs cellules de site. L'une d'elles est hébergée par le CESP.

Le [CESP](#) est l'un des principaux acteurs de la recherche épidémiologique en France. Il accueille 11 équipes de recherche dont les thématiques vont des biostatistiques les plus fondamentales aux sciences humaines et sociales en passant par la recherche clinique et les études en population générale.

Le CESP est doté d'un pôle mutualisé de recherche qui conseille les chercheurs en matière de méthodologie statistique, au sein duquel se situera le poste proposé.

**Adresse** Bat. 15-16 Hôpital Paul brousse, 16 av. Paul Vaillant Couturier 94800 Villejuif

**Délégation Régionale** Paris 11

## Description du poste

### Mission principale

Dans un contexte d'incitation forte à la science ouverte, il devient paradoxalement de plus en plus difficile de travailler collectivement sur des jeux de données du fait d'une sécurisation de plus en plus grande de l'accès aux données (RGPD, référentiel de sécurité SNDS).

Pour surmonter cette difficulté, une solution consiste à utiliser un outil technologique visant à "cloner" ses jeux de données en des jeux de données artificiels ayant la même distribution conjointe que les jeux de données initiaux. L'objectif est de proposer un accès totalement libre à ces jeux de données artificiels dans le but de stimuler des collaborations les plus larges possibles (une analyse des données réelles serait décidée dans un second temps, quand les pistes soulevées lors de l'analyse des données artificielles semblent prometteuses).

La personne recrutée aura pour mission de faire l'interface avec les équipes de recherche pour collecter les jeux de données susceptibles d'être clonés, de réaliser les traitements statistiques de ces jeux afin de rendre ce clonage possible, et de réaliser cette opération de clonage, soit directement, soit en lien avec des prestataires. Il ou elle devra également mettre ces jeux à dispositions en maintenant à jour un site web dédié.

La démarche et les méthodes mises en œuvre sur les données du CESP devront ensuite pouvoir s'appliquer dans l'environnement des données prise en charge plus largement par France Cohortes.

### Activités principales

- Mettre en place et optimiser les procédures de recueil et de contrôle des données
- Réaliser le clonage des jeux de données
- Participer au développement de logiciels spécifiques avec des prestataires
- Vérifier que les caractéristiques de la distribution statistique des données sont conservées
- Diffuser et valoriser des résultats sous forme de rapports techniques ou d'études
- Adapter les applications informatiques aux besoins du projet
- Gérer et maintenir des outils informatiques partagés
- Conseiller et former aux techniques et outils développés
- Assurer une veille scientifique et technologique dans le domaine du clonage et du floutage des données
- Participer à des réseaux professionnels
- Effectuer des présentations et des formations pour assurer un transfert de compétences
- Veiller à la reproductibilité des traitements réalisés et à la conservation de la propriété de la distribution

### Spécificités du poste

- Obligation de respecter le secret statistique et professionnel dans le cadre législatif existant
- Contraintes de calendrier à respecter

### Connaissances

- Maîtrise des outils et méthodes de génération de données
- Maîtrise des logiciels R, SAS, et Python
- Connaissance des langages de programmation
- Connaissance approfondie des structures habituelles de bases de données utilisées en épidémiologie ou en recherche clinique
- Connaissance approfondie de l'analyse et traitement des données
- Langue anglaise : B1 à B2 (cadre européen commun de référence pour les langues)

### Savoir-faire

- Découvrir et s'adapter à des méthodes statistiques innovantes y compris de data sciences
- Interagir avec des épidémiologistes et des informaticiens
- Garantir la qualité et la pertinence des outils d'analyse et des résultats
- Traiter des données, construire et gérer une base de données
- Programmer dans différents environnements informatiques
- Appliquer un dispositif de contrôle
- Assurer une veille
- Rédiger la documentation pour les utilisateurs
- Travailler en équipe y compris avec des industriels
- Transmettre des connaissances
- Utiliser les techniques de présentation

### Aptitudes

- Curiosité intellectuelle
- Autonomie

- Capacité de conceptualisation et de raisonnement analytique
- Sens du compromis

### Expérience souhaitée

- Une première expérience de 2 à 3 ans serait appréciée

### Diplôme et formation

- Master spécialité bio statistiques, ou biomathématiques, ou bio-informatique, ou mathématiques appliquées

## Informations Générales

### Date de prise de fonction

Juillet 2021

### Durée (CDD et détachements)

12 mois

Renouvelable :  OUI  NON

### Temps de travail

- Temps plein
- 38h30 hebdomadaires
- 32 jours Congés Annuels et 13 jours RTT

### Activités télétravaillables

OUI \*  NON

\* 2 jours par semaine hors contexte sanitaire.

### Rémunération

- **Contractuels** : A partir de 2500€ (brut mensuel) en fonction de l'expérience professionnelle sur des postes de niveau équivalent.

## Modalités de candidature

### Date limite de candidature

23 avril 2021

### Information / Contact

bruno.falissard@inserm.fr

### Modalités candidature

- Envoyer CV et lettre de motivation à **sabine.barbut@inserm.fr**
- Précisez vos prétentions salariales.

### Pour en savoir +

- Sur l'Inserm : <https://www.inserm.fr/>
- site RH : <https://rh.inserm.fr/Pages/default.aspx>
- Sur la politique handicap de l'Inserm et sur la mise en place d'aménagements de poste de travail, contactez la Mission Handicap : [emploi.handicap@inserm.fr](mailto:emploi.handicap@inserm.fr)