

## Data engineer/Data manager - plateforme de données en cancérologie de l'INCa

### Structure d'accueil

L'activité s'exerce au sein de l'UMR1252 SESSTIM, Sciences Economiques et Sociales de la Santé & Traitement de l'Information Médicale, à la Faculté des Sciences Médicales et Paramédicales de la Timone, à Marseille. L'UMR SESSTIM s'efforce de produire une recherche d'excellence, pluridisciplinaire et interdisciplinaire, en sciences humaines et sociale et en santé publique pouvant conduire à des changements dans les différents champs de la médecine prédictive, personnalisée, préemptive et participative. Les chercheurs du SESSTIM développent, ou sont associés à des travaux de recherche tentant d'apporter des réponses pour faire face à des enjeux actuels pour la société et ses populations et contribuent à des développements méthodologiques et à des avancées méthodologiques. En termes de pathologies d'intérêts, nos travaux de recherche portent principalement sur la maladie cancéreuse et les maladies infectieuses et transmissibles. Nos questionnements portent sur des facteurs individuels, populationnels ou contextuels. Nos travaux ciblent des populations de France, ou plus largement du Nord, mais aussi du Sud, que ce soit au niveau du bassin méditerranéen ou de l'Afrique subsaharienne, principalement.

### Missions principales

Le/La candidat·e travaillera au sein de l'équipe multidisciplinaire « Méthodes Quantitatives et Traitement de l'Information Médicale ([QuantIM](#)) », composée de chercheur·se·s en épidémiologie et santé publique, statisticien·ne·s, biostatisticien·ne·s, informaticien·ne·s et data scientists. Il/Elle sera plus spécifiquement affecté·e au projet AREMAMA (Accès à la reconstruction mammaire après mastectomie totale pour un cancer du sein : disparités spatiales) dont l'objectif principal est d'étudier les disparités spatiales associées à la reconstruction mammaire (RM) après mastectomie totale pour cancer du sein. Ce projet financé par l'Institut National du Cancer (INCa SHS-E-SP-RISP 2022) implique des membres de l'équipe « Cancer, Biomédecine & Société ([CanBIOS](#)) », ainsi qu'un médecin du Département d'oncologie chirurgicale de l'Institut Paoli Calmettes. Il s'appuiera sur les données de la plateforme de cancérologie de l'INCa, qui est une extraction du Système National des Données de Santé (SNDS).

### Activités

L'INCa a mis en place entrepôt de données intitulé « Plateforme de données en cancérologie ». Cette base médico-administrative est accompagnée d'algorithmes déjà développés et a donné lieu à des publications innovantes en raison de la quasi-exhaustivité des données. Selon la provenance des données ([DCIR ou PMSI](#)), il peut y avoir peu de tables mais très volumineuses ou beaucoup de tables mais peu volumineuses. Le/La candidat·e devra acquérir les connaissances de cette base. A cette fin, des formations sont prévues, notamment pour découvrir l'architecture de ces données. Le/La candidat·e bénéficiera de la présence d'une documentation interne à la plateforme pour comprendre les tables supplémentaires créées par l'INCa. Il/Elle travaillera à distance sur la plateforme sécurisée dans une sphère SAS en lien direct avec une sphère Oracle. Il/Elle sera en charge de la mise en place de requêtes pour l'extraction des variables en collaboration avec les chercheur·se·s impliqué·e·s dans le projet. A l'écoute des divers intervenant·e·s pouvant le/la guider dans le choix de ces variables, il/elle devra examiner les données pour déceler d'éventuelles incohérences ou anomalies, et s'assurer que les données collectées sont complètes et précises. Il/Elle réalisera des analyses menant à une description détaillée de la RM en France, selon le type de chirurgie. Une attention particulière sera portée sur les déplacements que les patients doivent faire pour subir la RM et, de façon plus générale, à d'éventuelles disparités spatiales. Ces analyses descriptives donneront lieu à la rédaction d'un article à laquelle le/la candidat·e sera associé·e. Enfin, il/elle interviendra en support de statisticien·ne·s en charge de faire d'autres analyses sur ces données.

## Compétences

### Savoir-faire

- Gestion de bases de données ;
- Capacité à travailler avec de grands ensembles de données ;
- Bonnes connaissances opérationnelles en SQL ;
- Maîtrise de SAS ;
- Concepts statistiques ;
- Connaissance du langage R (un atout).

### Savoir-être nécessaire

- Excellent sens de l'organisation ;
- Capacité de travail en équipe ;
- Rigueur ;
- Capacités d'analyse et de synthèse ;
- Sens de l'écoute et de la pédagogie pour dialoguer avec les différents interlocuteurs ;
- Être force de proposition.

### Compétences linguistiques

- Anglais : niveau scientifique

### Niveau de diplôme et expérience

- Bac+3 ou plus en biostatistiques, data management, data science, santé publique, ou épidémiologie.
- Minimum 1 an d'expérience dans la gestion de bases de données.

## Contrat

**Début** : Dès que possible, selon les délais administratifs de recrutement.

**Durée** : 36 mois.

**Rémunération** : Niveau ingénieur·e d'étude avec expérience ou ingénieur·e de recherche ; grille de salaires d'Aix-Marseille Université.

**Candidature** : SVP envoyer à [roch.giorgi@univ-amu.fr](mailto:roch.giorgi@univ-amu.fr) et [nathalie.graffeo@univ-amu.fr](mailto:nathalie.graffeo@univ-amu.fr) le dossier constitué :

- D'une lettre de motivation expliquant comment le/la candidat·e pense qu'il/elle pourra contribuer à ce projet ;
- D'un curriculum vitae.

**Date de fin de candidature** : 30 Mai 2023.

*Les entretiens seront réalisés par visio ou à Marseille.*