RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

Offre d'emploi

Santé publique France

Data Engineer / Scientist

Poste ouvert aux personnes en situation de handicap

Référence de l'annonce : DATA-CDI-2026-03

Poste

Poste	Data Engineer / Scientist (f/h)
Catégorie d'emploi	Catégorie 1
Emploi repère	Chargé de projet et d'expertise scientifique en santé publique
Type de contrat	CDI de droit public à temps complet Mise à disposition, détachement fonction publique d'Etat, Territoriale ou Hospitalière possible.
Rémunération	Selon l'expérience et le niveau de formation par référence aux grilles indiciaires des agences sanitaires, en application du décret n° 2003-224 du 7 mars 2003, ou selon statut particulier si fonctionnaire.
Motif	Redéploiement
Prise de fonction	Dès que possible
Renseignements	Sur le poste : Christophe Bonaldi, responsable de l'unité AMETIS, Direction Appui, Traitements et Analyses des données – tél : 01 55 12 53 89 Sur le recrutement : Marie-Christine Morel, Chargée du recrutement à la DRH – tél : 01 41 79 68 28

Localisation géographique

Adresse	Le siège de l'agence est situé à Saint-Maurice (94). Elle dispose également de 16 implantations régionales auprès des agences régionales de la santé, incluant l'outre-mer, de 4 pôles d'aide à distance en santé. Localisation du poste : 12 rue du Val d'Osne à Saint-Maurice (94)

Présentation de l'agence

Santé publique France est l'agence nationale de santé publique française. Etablissement public de l'Etat sous tutelle du ministre chargé de la santé, issu de la fusion de plusieurs établissements publics, créé par l'ordonnance 2016-246 du 15 avril 2016, l'agence intervient au service de la santé des populations. Agence scientifique, d'expertise et de sécurité sanitaires, elle a pour missions :

1° L'observation épidémiologique et la surveillance de l'état de santé des populations ;
2° La veille sur les risques sanitaires menaçant les populations ;
3° La promotion de la santé et la réduction des risques pour la santé ;
4° Le développement de la prévention et de l'éducation pour la santé ;
5° La préparation et la réponse aux menaces, alertes et crises sanitaires ;
6° Le lancement de l'alerte sanitaire.

L'agence est organisée autour de 12 directions scientifiques, transversales ou assurant le soutien à l'activité.

Les orientations stratégiques de l'agence et son programme de travail, arrêtés par son Conseil d'administration, se déclinent en trois axes : Consolider la capacité d'anticipation et de réponse réactive pour faire face aux menaces sanitaires ; Mesurer et évaluer l'ampleur des maladies et des facteurs de risques pour guider leur prévention et leur contrôle ; Renforcer l'impact sur la santé dans toutes les politiques publiques et la prévention et promotion de la santé.

Affectation

Direction

Direction Appui, Traitements et Analyses des données (DATA)
La DATA assure un appui transversal à l'ensemble de l'agence pour le traitement, l'analyse et la valorisation des données. Forte d'une cinquantaine d'agents, la direction est structurée en trois unités spécialisées (unité « Applications, big data et surveillance syndromique » (ABISS), unité « Appui et méthodes pour les études et investigations dans le domaine de la surveillance » (AMETIS), et unité « Enquêtes »). Elle intervient sur l'ensemble du cycle de vie des données de santé.
Ses missions couvrent notamment la gestion de données, l'analyse statistique, la géomatique, la métrologie, ainsi que le développement d'outils informatiques d'analyse et de visualisation. Elle pilote ou soutient plusieurs dispositifs structurants, notamment le système de surveillance syndromique SurSaUD, l'enquête Baromètre santé, le site open-DATA Odissé ainsi que l'exploitation de bases médico-administratives comme le SNDS. La direction DATA développe une expertise reconnue en modélisation spatio-temporelle, détection automatique de signaux, et intelligence artificielle. Soucieuse de renforcer la qualité scientifique de ses travaux, elle collabore activement avec des partenaires institutionnels et académiques, et accueille régulièrement stagiaires, internes, doctorants et chercheurs, dans une dynamique d'innovation continue au service de la surveillance en santé publique.

Unité

L'unité AMETIS (Appui et Méthodes pour les Études et Investigations dans le domaine de la Surveillance) met en œuvre des méthodes quantitatives avancées et des outils d'ingénierie / science des données pour accompagner les études et dispositifs de surveillance de Santé publique France. Elle couvre l'intégralité du cycle des données, avec une expertise structurée autour de cinq axes principaux :

- Conception d'études : élaboration des designs méthodologiques et des protocoles de recueil de données.
- Ingénierie des données : nettoyage, structuration et automatisation de pipelines, ainsi que gestion de bases de données complexes.
- Modélisation et science des données : mise en œuvre d'analyses statistiques avancées et de méthodes de data science pour produire des indicateurs épidémiologiques fiables et innovants.
- Visualisation et outils interactifs : Développement de cartographies dynamiques, de tableaux de bord interactifs et de rapports automatisés.
- Veille et innovation : Intégration de méthodes émergentes, IA et collaborations académiques.

En parallèle, AMETIS contribue aux dispositifs nationaux (ex : portail open-DATA Odissé) en assurant la qualité, la traçabilité et la reproductibilité des données et analyses produites.

Description du poste

Pour ne pas alourdir le texte, nous nous conformons à la règle qui permet d'utiliser le masculin avec la valeur de neutre.

Missions

La personne recrutée aura pour mission de traiter et analyser des données issues des systèmes de surveillance épidémiologique de Santé publique France, en couvrant l'ensemble des étapes : gestion et structuration des bases de données, développement et optimisation de scripts, construction d'indicateurs épidémiologiques à l'aide de méthodes statistiques ou de data science, ainsi que la restitution de ces indicateurs sous forme de

bulletins ou de tableaux de bord. Les données traitées pourront provenir de systèmes de surveillance spécifiques ou non spécifiques et concerner des thématiques variées, qu'il s'agisse de populations générales ou spécifiques, de pathologies ou de déterminants de santé. Une attention particulière sera portée sur le développement et le maintien d'outils et de processus permettant de produire, en temps réel et de manière réactive, des indicateurs essentiels en situation d'alerte ou de crise sanitaire. Le titulaire du poste contribuera également à l'innovation méthodologique en participant à la conception et à l'implémentation de méthodes et d'outils avancés pour le traitement des données de santé. Ce rôle implique une collaboration étroite avec des équipes pluridisciplinaires (épidémiologistes, biostatisticiens, etc.) et une contribution active à la veille technologique et scientifique, afin de renforcer la capacité de l'agence à répondre aux enjeux actuels et émergents en santé publique. La restitution des résultats, sous forme de rapports, de visualisations ou de présentations, sera un aspect clé du poste, tout comme la formation et l'accompagnement des équipes à l'utilisation des outils développés.

Activités

Sous l'autorité du responsable d'unité et en lien étroit avec la personne en charge du pilotage de l'activité Data Science & Data Ingénierie de la direction, le data engineer/scientist aura pour mission d'assurer la gestion, le traitement et la restitution des données.

Ses activités consisteront notamment à :

- Concevoir, développer et maintenir des systèmes de gestion de données et des pipelines automatisés couvrant l'ensemble du cycle de traitement et de restitution des données.
- Nettoyer, structurer et préparer des données, en garantissant leur qualité, leur fiabilité, leur traçabilité et leur conformité aux exigences réglementaires.
- Explorer et proposer des solutions technologiques pour faire progresser la qualité et la fiabilité des données.
- Identifier les possibilités d'acquisition et d'intégration de nouvelles sources de données pour enrichir les connaissances au service de la santé des populations.
- Définir, implémenter et interpréter des indicateurs de surveillance adaptés aux enjeux métiers.
- Concevoir, développer et optimiser des méthodes statistiques et d'apprentissage automatique et profond pour la construction d'indicateurs épidémiologiques, la détection de signaux et la modélisation prédictive.
- Développer des outils de monitoring et de restitution (API, tableaux de bord interactifs, rapports automatisés) pour répondre aux besoins opérationnels et stratégiques de l'agence.
- Identifier et intégrer de nouvelles sources de données, ainsi que des approches technologiques émergentes (IA, traitement distribué, etc.) pour enrichir les analyses et renforcer la réactivité en situation de crise.
- Travailler en étroite interaction avec des épidémiologistes, biostatisticiens et ingénieurs pour traduire les besoins métiers en solutions techniques robustes.
- Rédiger des notes méthodologiques, contribuer à la valorisation des résultats (bulletins, rapports d'études et articles scientifiques) et former les équipes à l'utilisation des outils développés.

Ce poste s'inscrit dans un environnement technique évolutif et collaboratif, mobilisant des outils de développement modernes et des langages adaptés à la science des données. À terme, il bénéficiera d'une infrastructure de calcul performante, actuellement en cours de déploiement. La personne recrutée évoluera au sein d'une équipe pluridisciplinaire, en interaction étroite avec des épidémiologistes, data scientistes, biostatisticiens, ingénieurs, ainsi que les membres de la Direction des Systèmes d'Information (DSI) et le RSSI (Référent sécurité des systèmes d'information).

Les principaux outils et technologies mobilisés incluent :

- Langages de programmation : Python / R / SAS ;
- Rapports automatisés : Rmarkdown / Quarto ;
- Environnement collaboratif : GitLab (versioning, intégration continue, gestion des issues) ;

- Automatisation et orchestration : Apache Airflow (déploiement, supervision des workflows), Docker ;
- Formats et bases de données : PostgreSQL, fichiers Parquet, CSV ...
- Environnements de développement : VS Code, R Studio, IA Mistral ;
- Traitement intensif : Apache Spark via serveurs sécurisés de calcul haute performance (en cours de construction).

La personne recrutée pourra être amenée à représenter Santé publique France dans des instances nationales et internationales, lors d'échanges scientifiques ou institutionnels. Elle encadrera également des stagiaires de niveau Master 2, des internes de santé publique et des doctorants sur des thématiques liées à l'ingénierie et à la science des données. Enfin, elle participera à la continuité d'activité par le biais de permanences hebdomadaires et d'astreintes de week-end, et pourra être sollicitée pour contribuer à toute activité relevant des missions de l'Agence, notamment en situation d'urgence ou de crise.

Profil recherché

Niveau Diplôme

Ingénieur/Master ou Doctorat de Science dans le domaine de la data sciences, du traitement des données ou des biostatistiques.

Expérience

Une expérience professionnelle d'au moins 3 ans en tant que Data Engineer / Scientist ainsi qu'une bonne connaissance du secteur de la santé publique/épidémiologie est souhaitable.

Aptitudes et compétences

Il est attendu pour ce poste :

- Très bonne maîtrise de la programmation sous Python et/ou R.
- Maîtrise des pipelines de données, de l'automatisation des traitements et des méthodes de modélisation de données avancés.
- Connaissance des modèles statistiques usuels en épidémiologie : Modèles linéaires généralisés (GLM) et ses dérivés (GAM, GLMM, GAMM…), modèles marginaux (GEE), modèles de survie …
- Connaissances en Machine Learning, incluant la conception, l'entraînement et l'évaluation de modèles prédictifs ou descriptifs (supervisés et non supervisés), ainsi que la capacité à adapter ces méthodes aux spécificités et besoins métiers.
- Connaissances en bases de données relationnelles et en programmation SQL (PostrgreSQL).
- Bonnes Connaissances des outils de reporting et de DataViz modernes (Shiny, Quarto, Superset …).
- Connaissances des outils d'automatisation et orchestration (Apache Airflow, Gitlab).
- Connaissance de l'environnement LINUX.

Il est attendu pour ce poste les compétences suivantes :

- Conduite de projet
- Autonomie et force de proposition
- Qualités relationnelles, aptitudes à la discussion et au travail d'équipe
- Réactivité, résistance à la pression
- Curiosité intellectuelle et esprit d'initiative
- Capacité d'anticipation, d'organisation, de réactivité et de décision
- Expression écrite et orale
- Rigueur scientifique

Pour postuler

Adresser les candidatures (lettre de motivation + cv) en indiquant la référence de l'annonce par courriel :
recrut@santepubliquefrance.fr