RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

Profil de poste

Santé publique France

Poste d'Apprenti Data Engineer/Scientist (f/h)

Poste ouvert aux personnes en situation de handicap

Référence de l'annonce : DATA-Appr-2026-01

Poste

Poste	Apprenti Data Engineer/scientist (f/h)
Type de contrat	Contrat d'apprentissage
Rémunération	Selon les textes en vigueur
Prise de fonction	Dès que possible
Renseignements	Sur le poste : Johnny PLATON Data Scientist / Data Engineer à la Direction Appui, Traitements et Analyses des données) – tél : 01 41 79 68 30 Sur le recrutement : Sadia EL MOUDEN gestionnaire de formation à la Direction des Ressources Humaines sadia.elmouden@santepubliquefrance.fr

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 direction DATA mène plusieurs projets stratégiques visant à moderniser les chaînes de traitement des données issues des systèmes de surveillance de Santé publique France. Ces projets s'appuient sur des approches innovantes en data ingénierie/science pour répondre aux enjeux de santé publique, notamment à travers la modélisation prédictive, l'intelligence

artificielle et l'analyse avancée de données massives. Trois systèmes majeurs illustrent cette dynamique :

- Le programme SurSaUD assure une surveillance syndromique en temps réel en exploitant les données des urgences hospitalières, de SOS Médecins et des certificats de décès.
- Les Maladies à Signalement obligatoire (MSO) suit en temps réel des pathologies à fort impact sanitaire grâce à la collecte systématique et à l'analyse des déclarations transmises par les professionnels de santé.
- Le projet Orchidée déploie une surveillance épidémiologique multi-thématique à partir des données hospitalières

Activités

Ces données permettent de générer un volume important de séries temporelles, décrivant l'évolution d'indicateurs de santé à différentes échelles spatio-temporelles. Leur structuration et leur analyse représentent un enjeu stratégique pour renforcer les capacités de surveillance, de modélisation et d'alerte.

Dans ce contexte, la Direction DATA a engagé un projet visant à construire une chaîne de traitement automatisée, fiable et évolutive, permettant de valoriser ces données au moyen de méthodes avancées de traitement et d'analyse.

L'apprenti interviendra sur l'ensemble de la chaîne de traitement des données, depuis la collecte jusqu'à la production et la mise à disposition d'indicateurs. Ce dernier participera entre autres aux activités suivantes :

- 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.
- 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

Ces activités s'intègrent dans un environnement technique dynamique et collaboratif, mobilisant des outils de développement modernes, des langages adaptés à la science des données, et des infrastructures de calcul performantes. L'apprenti évoluera au sein d'une équipe pluridisciplinaire, en interaction étroite avec des épidémiologistes, data scientists, statisticiens, ingénieurs et membres de la DSI ainsi que le RSSI.
Les principaux outils et technologies mobilisés incluent :

- Langages : Python, R
- Environnement collaboratif : GitLab (versionning, intégration continue, gestion des issues)
- Automatisation et orchestration : Apache Airflow (déploiement, supervision des workflows), Docker
- Formats et bases de données : PostgreSQL, DuckDB, fichiers Parquet, CSV

- Visualisation : Quarto, Shiny (R et Python)
- Environnements de développement : VS Code, RStudio, IA Mistral
- Traitement intensif : Apache Spark, via les serveurs de calcul internes de Santé publique France

Profil recherché

Niveau Diplôme

Niveau Master 1 ou Master 2, idéalement en informatique, data science, statistique, mathématiques appliquées ou disciplines connexes.

Aptitudes et compétences

- Maîtrise de la programmation sous R et/ou Python
- Feature engineering / Data cleaning
- Connaissances en développement logiciel, notamment sur les bonnes pratiques telles que les tests unitaires, le versionning et l'intégration continue.
- Connaissances en bases de données relationnelles et en programmation SQL (Postrgresql, DuckDB)
- Connaissance des modèles statistiques usuels en épidémiologie (Modèles linéaires généralisés et ses dérivés …)
- 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
- Connaissance de l'environnement LINUX (Centos/Ubuntu).
- Une expérience avec des outils comme GitLab, Airflow ou PostgreSQL serait un plus.

Le·la candidat·e devra faire preuve de bonnes capacités d'analyse, d'adaptation et d'organisation, ainsi que d'un réel esprit d'équipe pour évoluer au sein d'un environnement pluridisciplinaire. Ce poste en alternance offre l'opportunité de participer à des projets techniques innovants, au service de la surveillance de la santé publique, en étroite collaboration avec des experts en épidémiologie, data science et ingénierie, dans un cadre de travail stimulant, bienveillant et collaboratif.

Pour postuler

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