

## Poste d'Apprenti Data Scientist SNDS (f/h)

Poste ouvert aux personnes en situation de handicap

Référence de l'annonce : DATA-Appr-2026-02

### Poste

|                          |   |
|--------------------------|---|
| <b>Poste</b>             | Apprenti Data scientist SNDS (f/h)  |
| <b>Type de contrat</b>   | Contrat d'apprentissage   |
| <b>Rémunération</b>      | Selon les textes en vigueur   |
| <b>Prise de fonction</b> | Dès que possible  |
| <b>Renseignements</b>    | Sur le poste : Delphine SERRA Direction Appui, Traitements et Analyses des données – tél : +33 (0)1 55 12 54 09<br>Sur le recrutement : Sadia EL MOUDEN gestionnaire de formation à la Direction des Ressources Humaines <a href="mailto:sadia.elmouden@santepubliquefrance.fr">sadia.elmouden@santepubliquefrance.fr</a> |

### Localisation géographique

|                |  |
|----------------|--|
| <b>Adresse</b> | Le siège de l'agence est situé à Saint-Maurice (94).<br>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.<br>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

Les missions de l'unité " ABISS " s'organisent autour de plusieurs pôles d'activités :

- Le soutien aux directions métiers pour accès, exploitation et analyse des bases de données transverses issues de systèmes de surveillance spécifiques (pour des maladies à notification obligatoire ou non), des données du système national des données de santé (SNDS), de la surveillance syndromique, de référentiels, etc...)
- Le soutien aux directions métiers pour le développement métier d'applications informatiques : outils de recueil, d'analyse et/ou de restitution de données
- Le pilotage du programme de surveillance syndromique (système de surveillance SurSaUD)
- La mise à disposition des données de l'agence.
- L'unité assure une veille méthodologique et technique permettant un soutien de qualité à l'agence, une offre de formation en interne à l'agence ainsi que des permanences hebdomadaires pour assurer la continuité d'activité.
- En ce qui concerne le SNDS, Santé publique France dispose d'un accès permanent établi par le décret no 2016-1871 du 26 décembre 2016 et assure la présidence du comité des organismes et structures ayant un accès permanent au SNDS (COMAP). L'unité ABISS comprend une équipe SNDS de plusieurs collaborateurs disposant de compétences approfondies sur le traitement et l'analyse de ces données et qui intervient en soutien des directions scientifiques de l'agence (expertise, formation, documentation, animation de réseau).

## 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 pilote des projets structurants pour moderniser les chaînes de traitement des données médico-administratives, avec une attention particulière portée sur le Système National des Données de Santé (SNDS). Ces initiatives mobilisent des méthodes innovantes en statistique et en ingénierie des données, afin de répondre aux défis majeurs de santé

publique par l'analyse avancée de données massives, la modélisation prédictive et l'intégration de sources hétérogènes. Trois axes prioritaires illustrent cette démarche :

- Reconstituer les trajectoires des patients au sein du SNDS pour mieux évaluer l'impact des expositions environnementales et des inégalités sociales et territoriales sur la santé.
- Construire des scores pronostiques à partir du SNDS en déployant une infrastructure facilitant la création et la validation des modèles (sélection et ingénierie de données, modélisation, validation et évaluation). L'objectif est de standardiser la production de scores robustes et reproductibles, tout en accélérant leur mise à disposition pour la recherche et la décision.
- Démocratiser l'accès et l'analyse des données médico-administratives complémentaires, telles que l'EDP-santé et faciliter leur appariement avec des données complémentaires de santé, socio-économiques et/ou environnementales pour favoriser des analyses plus intégrées et transversales (One Health, approche vie entière, ...)

## Activités

Au sein de l'unité ABISS de la direction DATA et sous la responsabilité du responsable de l'unité, l'apprenti-e Data Engineer/Scientist SNDS participera activement à la conception, au développement et à l'optimisation des chaînes de traitement et d'analyse des données du Système National des Données de Santé (SNDS). Il-elle contribuera à des projets structurants visant à moderniser l'exploitation des données médico-administratives, en mobilisant des méthodes avancées en ingénierie des données, en statistique et en intelligence artificielle.

Ses activités s'articuleront autour des missions suivantes :

- **Contribution à la modernisation des chaînes de traitement des données SNDS** : Participation à la conception et à l'amélioration des extractions / traitements de données médico-administratives, en collaboration avec les équipes de la direction DATA et les directions métiers de l'agence. Cela inclut l'automatisation des processus d'extraction et de transformation, l'optimisation des requêtes et des traitements et leur mise à disposition (direction métiers, open source).
- **Analyse et valorisation des données médico-administratives** : participation à la reconstitution de trajectoires résidentielles et à l'appariement des données SNDS avec des sources complémentaires (socio-économiques, environnementales, etc.).
- **Développement de modèles prédictifs** : Appui à la construction et à la validation de modèles avancés, depuis la sélection et l'ingénierie de descripteurs jusqu'à leur évaluation. L'apprenti-e contribuera à la mise en place d'une infrastructure automatisée pour standardiser la production de modèles robustes et reproductibles, en vue de leur déploiement pour la recherche et l'aide à la décision.
- **Soutien à la veille méthodologique et technique** : Contribution à la documentation, à la formation et à l'animation du réseau interne autour des bonnes pratiques d'exploitation du SNDS. L'apprenti-e pourra également participer à la mise à disposition des données via des outils d'analyse et de visualisation.
- **Collaboration avec les partenaires institutionnels et académiques** : Implication dans des projets collaboratifs avec des acteurs externes, afin de renforcer la qualité scientifique des travaux et d'intégrer les innovations méthodologiques dans les pratiques de l'agence.
- **Veille technologique et formation continue** : Participation active aux formations internes et aux permanences techniques organisées par l'unité ABISS, afin de maintenir une expertise à jour sur les outils et méthodes émergents en data science et en santé publique.

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 : R, Python et SAS

- Environnement collaboratif : GitLab (versionning, intégration continue, gestion des issues)
- Formats et bases de données : PostgreSQL, DuckDB, fichiers Parquet, CSV
- Visualisation : Quarto, Shiny (R et Python)
- Environnements de développement : VS Code, RStudio, IA Mistral

## Profil recherché

**Niveau** Niveau Master 2, idéalement en statistique, data science, mathématiques appliquées ou  
**Diplôme** disciplines connexes.

**Aptitudes et compétences**

- Maîtrise de la programmation sous R et SAS
- Manipulation et nettoyage de données
- 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
- Une connaissance avec des outils de versioning comme GitLab 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](mailto:sadia.elmouden@santepubliquefrance.fr)