

Intitulé du stage : Stage en Data Science - Modernisation des bulletins de surveillance épidémiologique SurSaUD

Référence de l'offre : DATA-STA-2026-05

■ **Présentation de Santé publique France**

Santé publique France est l'Agence nationale de santé publique française. Etablissement public de l'Etat sous tutelle du ministre chargé de la santé créé par l'ordonnance 2016-246 du 15 avril 2016, elle intervient au service de la santé des populations.

Agence scientifique et d'expertise du champ sanitaire, 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 quatre conseils (Conseil d'Administration, Conseil scientifique, Comité d'Ethique et de Déontologie et Comité d'orientation et de dialogue), de directions scientifiques et transversales, et de directions assurant le support et le soutien à l'activité. L'Agence dispose d'implantations régionales (Cellules régionales) auprès des agences régionales de la santé.

Son programme de travail, arrêté par son Conseil d'administration, s'articule autour de 6 enjeux : Anticipation, préparation et réponse aux menaces de santé publique, dont les épidémies ; Numérique en santé publique ; Santé environnementale, changement climatique et environnement de travail ; Fardeau des maladies et de leurs déterminants, efficacité des interventions et retour sur investissement de la prévention ; Stratégie de prévention, marketing social et approche par populations ; Inégalités sociales, vulnérabilités territoriales.

Son siège est situé à Saint-Maurice (94).

Localisation du stage :

☒ Saint-Maurice

☐ En région, indiquer la région, le département et la ville concernés :

■ **Le stage**

Contexte :

Description de la Direction : Le stage s'inscrit au sein de la Direction Appui, Traitements et Analyses des données (DATA) de Santé publique France, qui 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 », unité « Appui et méthodes pour les études et investigations dans le domaine de la surveillance », 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 Odissée 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, interne, doctorants et chercheurs, dans une dynamique d'innovation continue au service de la surveillance en santé publique.

Description du projet : Le stage s'inscrit dans les travaux menés autour du système de surveillance syndromique SurSaUD, piloté par Santé publique France. Mis en place en 2004 à la suite de la canicule de 2003, ce dispositif permet de suivre en temps quasi réel l'état de santé de la population, en facilitant la détection précoce d'événements sanitaires, qu'ils soient attendus ou inhabituels. SurSaUD repose sur l'intégration quotidienne et automatisée de données individuelles et anonymisées issues de quatre sources complémentaires :

- Les passages aux urgences hospitalières (réseau OSCOUR),
- Les interventions des associations SOS Médecins,
- Les données de mortalité issues des bulletins d'état civil transmis à l'Insee,
- Les certificats électroniques de décès transmis à l'Inserm-CépiDc.

Chaque semaine, les données collectées alimentent les bulletins SurSaUD, publiés au format PDF. Ces bulletins analysent l'évolution de multiples indicateurs syndromiques à différentes échelles spatio-temporelles. Diffusés aux partenaires internes et externes, ils constituent un pilier du dispositif de surveillance de Santé publique France, en éclairant la compréhension en temps réel de la situation sanitaire et en jouant un rôle clé dans l'anticipation et la gestion des risques. Une synthèse générale complète ces bulletins, et l'ensemble des documents est mis en ligne chaque semaine sur le site de l'agence. Dans ce cadre, la Direction DATA a lancé un projet de refonte et de modernisation de son infrastructure SurSaUD, incluant une évolution des bulletins pour en renforcer la pertinence, l'accessibilité et l'automatisation. Le stage proposé s'inscrit pleinement dans cette dynamique.

Sujet de stage : Le ou la stagiaire contribuera à la modernisation des bulletins SurSaUD, en prenant en charge l'ensemble du processus : de la conception, à l'automatisation et au déploiement des nouveaux supports. Il ou elle explorera également des solutions innovantes pour la visualisation des données (quarto, Shiny, Superset) et l'automatisation des synthèses (via des modèles de langage avancé de type LLM).

Missions du stage : Le stage inclura la participation aux missions suivantes, organisées en deux axes complémentaires :

1. Conception et innovation

- Analyse des bulletins existants, définition de critères statistiques et de comparaison de dynamiques temporelles justifiant une remontée des indicateurs
- Maquetter les nouveaux bulletins pour garantir une clarté, une pertinence et une accessibilité optimales en intégrant la charte de l'agence.
- Expérimenter l'automatisation de la génération des synthèses à l'aide de modèles de langage avancés (Mistral).
- Mobiliser les différentes sources du système SurSaUD et leurs référentiels pour constituer les jeux de données nécessaires à la réalisation des bulletins.
- Développer les bulletins en utilisant des technologies modernes sous R pour une production reproductible et dynamique.

2. Déploiement et valorisation

- Participer à la mise en production des nouveaux bulletins, en assurant leur intégration fluide dans les processus existants.
- Contribuer à la création d'une version en ligne sous forme de tableau de bord, pour une visualisation interactive et actualisée des indicateurs syndromiques.
- Créer la documentation technique complète (guides, scripts, structuration des jeux de données) et mettre en œuvre des pratiques garantissant la reproductibilité

Environnement de travail : Le stage s'intègre 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. Le-la stagiaire évoluera au sein d'une équipe pluridisciplinaire, en interaction étroite avec des épidémiologistes, médecins, data scientists, statisticiens, ingénieurs et membres de la DSI.

Les principaux outils et technologies mobilisés incluent :

- Langages : R, Python et SAS
- Environnement collaboratif : GitLab
- Automatisation et orchestration : Apache Airflow, Docker
- Formats et bases de données : PostgreSQL, fichiers Parquet, CSV
- Environnements de développement : R/Python, Rstudio, VS Code, IA Mistral
- Outils de publication : Quarto, Shiny, Superset ...

Profil recherché : Nous recherchons un·e stagiaire de niveau Master 2, issu·e d'une formation en, statistique ou disciplines connexes en data science. Le·la candidat·e devra en particulier maîtriser les langages R et posséder des connaissances solides en traitement de données de santé publique ou épidémiologiques. Une expérience avec Quarto, le système de publication scientifique et technique, serait un atout majeur pour ce stage.

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 stage offre l'opportunité de participer à des projets 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.

Description : Pour tout renseignement complémentaire, consulter le portail sur le site <http://www.santepubliquefrance.fr>

▪ **Type de stage proposé**

- ☐ Licence
- ☐ Master 1 ☐ Master 1 + sujet tutoré ☐ Master 1 « observation »
- ☒ Master 2 Professionnel ☐ Master 2 Recherche

Gratification du stage : selon la réglementation en vigueur dans les établissements publics

▪ **Date proposée pour le stage et durée :**

- ☒ Sans contrainte de date
- ☐ A partir de : (indiquer une date)
- Durée minimum : 6 mois
- Extension possible au-delà de la période obligatoire ☒ Oui ☐ Non

▪ **Prérequis :**

- ☐ Aucun
- ☒ Compétences spécifiques (préciser) : Traitement de données (santé publique)
- ☒ Maîtrise d'un logiciel spécifique (préciser) : Langage R, système Quarto
- ☒ Autre (préciser) : Connaissance sur les outils de versionning de code tels que GitHub/GitLab

▪ **Stage proposé par :**

Direction : DATA

Unité : ABISS

Maître(s) de stage / personne contact :

Nom : METHY

Prénom : Nicolas

Fonction : Epidémiologiste

Téléphone :

Adresse email : nicolas.methy@santepubliquefrance.fr

Co-encadrement

Nom : RAHALI

Prénom : Sofiane

Fonction : Data Scientist

Téléphone :

Adresse email : sofiane.rahali@santepubliquefrance.fr

▪ **Pour postuler**

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