Offre de stage

Offre de stage à Santé publique France
Année universitaire 2026-2027

Intitulé du stage : Stage de Data engineering/Science pour la surveillance en santé publique (EUROMOMO)

Référence de l'offre : DATA-STA-2026-10

▪ 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 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.
Le siège de l'agence est situé à Saint-Maurice (94). L'agence dispose d'implantations régionales auprès des agences régionales de la santé.

Localisation du stage :
☒ Saint-Maurice
☐ En région, indiquer la région et la ville concernées :

▪ 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 Odyssée 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 de l'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.

L'unité AMETIS travaille également en collaboration étroite avec l'unité ABISS en charge du pilotage du système SurSaUD, qui collecte quotidiennement les informations démographiques, administratives et médicales :

- des passages dans 700 services d'urgences participant au réseau de surveillance coordonnée des urgences (OSCOUR®) ;
- des actes médicaux effectués par 62 associations SOS Médecins (données de médecine d'urgences de ville) ;
- les décès enregistrés par plus de 16 000 bureaux d'état-civil, avec une transmission automatisée vers l'Institut national de la statistique et des études économiques (Insee) ;
- des volets médicaux des certificats de décès transmis par voie électronique, à l'Inserm-CépiDc.

Sujet du stage :

Santé publique France surveille chaque semaine la mortalité toutes causes à partir des données d'état civil transmises par l'Insee, en s'appuyant sur le modèle européen EuroMOMO pour estimer le nombre attendu de décès et détecter d'éventuels excès. Une application R Shiny développée en 2015 permet de visualiser ces analyses à un pas de temps quotidien et hebdomadaire, tous âges et par classe d'âge, aux échelles nationale, régionale et départementale. Elle repose sur un échantillon d'environ 5 000 communes (84 % de la mortalité nationale). Au 1er semestre 2027, un élargissement de l'échantillon à environ 12 000 communes va permettre l'usage d'un autre axe d'analyse majeur : le type de lieu de décès (domicile, établissement hospitalier, maison de retraite). Cette application est actualisée quotidiennement avec les nouvelles données reçues de l'Insee.

Le stage vise à moderniser cette application R-Shiny et à enrichir les analyses : intégration de nouvelles strates (sexe, type de lieu de décès), intégration des résultats d'une méthode de nowcasting pour tenir compte des délais de transmission, cartographie par type de lieu de décès, indicateurs de surmortalité plus fins que le simple dépassement de seuil. Il comprend également une refonte de l'architecture du code, afin que toutes les applications de Santé publique France soit harmonisées.

Le stage s'inscrit dans une volonté plus large de la direction DATA d'automatiser la création d'applications Shiny à l'aide de l'IA agentique : la construction et la modernisation de l'application s'appuieront sur des agents de code autonomes (par exemple Cline, connecté aux modèles de Mistral AI), appelés à devenir un mode de développement standard pour les applications de l'agence. Le stage comprendra ainsi la création et l'amélioration de « skills » : bibliothèques de consignes, contextes et procédures réutilisables qui guident ces agents IA pour la génération, la refactorisation et la documentation du code R/Shiny. Cette démarche vise à accélérer le développement, à en fiabiliser les sorties et à capitaliser l'expertise méthodologique de l'unité sous forme d'actifs réutilisables par les autres équipes de l'agence.

Missions du stage :

1. Prise en main et état des lieux

- Comprendre le dispositif de surveillance de la mortalité (données Insee, délais de transmission, principes d'EuroMOMO) et les sorties de modèles mises à disposition.

- Auditer l'application R Shiny existante (code, dépendances, performances, ergonomie) et proposer un plan d'évolution.
- Se familiariser avec l'environnement de développement assisté par IA (Cline, modèles Mistral AI) et son cadre d'usage (sécurité et confidentialité).

2. Refonte de l'architecture de l'application

- Adapter l'application aux nouvelles volumétries et au nouveau format des données, et optimiser ses performances.
- Structurer et documenter le code (projet reproductible, tests, gestion des versions), en l'harmonisant avec les autres applications de Santé publique France.

3. Développement assisté par IA et création de « skills »

- Utiliser l'agent de code (Cline / Mistral) pour accélérer la refactorisation, la génération de modules, les tests et la documentation, sous revue humaine systématique.
- Concevoir, documenter et fiabiliser des « skills » (instructions, contextes métier, patrons de code, procédures de validation) pour les tâches récurrentes du projet.
- Contribuer à un cadre de développement réutilisable pour les futures applications de l'agence, en mesurant l'effet des skills sur la qualité et la vitesse de développement et en capitalisant les bonnes pratiques pour les autres équipes de la direction DATA.

4. Intégration des nouvelles stratifications

- Ajouter les résultats stratifiés par sexe et par type de lieu de décès, aux niveaux géographiques définis avec l'équipe, avec des visualisations adaptées (courbes, tableaux, comparaisons entre strates).

5. Cartographie

- Développer des cartes interactives de la mortalité et de la surmortalité par type de lieu de décès et selon d'autres strates.

6. Indicateurs d'intensité de surmortalité

- Représenter des niveaux d'intensité de surmortalité, au-delà du simple dépassement d'un seuil (par exemple par classes d'intensité), en lien avec les statisticiens de l'équipe.

7. Validation, documentation et transfert

- Tester l'application avec les utilisateurs (équipe de surveillance, Cellules régionales) et intégrer leurs retours.
- Rédiger la documentation technique et un guide utilisateur, puis présenter les résultats lors de réunions de restitution.
- Rédiger un retour d'expérience sur l'usage de l'IA générative (apports, limites, recommandations pour l'agence).

Activités du stage : Le stage s'intègre dans un environnement technique dynamique et collaboratif, mobilisant des outils de développement modernes — dont des assistants de développement basés sur l'intelligence artificielle générative (agents de code type Cline, modèles Mistral AI) — 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, data scientists, statisticiens, ingénieurs.

A la fin du stage, le candidat retenu aura acquis les compétences suivantes :

Compétences techniques :

- Développement d'applications R Shiny avancées : architecture modulaire, optimisation des performances, gestion de gros volumes de données.
- Conception d'un système paramétrable et organisation de code reproductible (structuration en projet ou package, tests, gestion de versions avec Git).
- Développement assisté par intelligence artificielle : pilotage d'agents de code autonomes (type Cline / Mistral AI), revue critique des sorties, bonnes pratiques d'usage en environnement professionnel.
- Conception, itération et capitalisation de " skills " (consignes et contextes réutilisables) pour encadrer et fiabiliser le développement assisté par IA.
- Cartographie interactive et visualisation de données de santé (cartes, graphiques, tableaux de bord).

- Mise en forme et communication d'indicateurs statistiques complexes (intensité de surmortalité, intervalles de confiance) pour un public non spécialiste.

Connaissances métier

- Fonctionnement d'un système national de surveillance sanitaire et de la mortalité toutes causes.
- Principes de la modélisation EuroMOMO (nombre attendu de décès, excès de mortalité) et lecture critique des résultats, en lien avec les délais de transmission des données.
- Connaissance des données d'état civil et de leurs limites (couverture géographique, absence de causes médicales de décès).

Compétences transversales

- Travail en équipe pluridisciplinaire avec des épidémiologistes, des statisticiens et des chargés de surveillance.
- Recueil et prise en compte des besoins d'utilisateurs, puis conduite de tests de validation.
- Rédaction de documentation technique et de guides utilisateurs.
- Restitution orale de résultats lors de réunions de présentation.
- Gestion de projet : planification, priorisation des évolutions, autonomie.

Description :

Pour tout renseignement complémentaire, consulter le portail sur le site http://www.santépubliquefrance.fr

▪ Type de stage proposé

☐ Licence

☐ Master 1 ☐ Master 1 + sujet tutoré* ☐ Master 1 « observation »**

☒ Master 2 Professionnel ☐ Master 2 Recherche

* stage M1 + sujet tutoré (proposé par l'ISPED) : le stage complète un travail préalable de sujet tutoré de 2 mois et demi (revue de la littérature, analyse d'une base de données, etc.)

** stage M1 « observation » (proposé par l'ISPED) : le stagiaire contribue aux activités courantes de l'équipe sans être en charge d'un dossier particulier

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 : dès que possible

Durée minimum : 5-6 mois

Extension possible au-delà de la période obligatoire ☐ Oui ☒ Non

▪ Prérequis

☐ Aucun

☒ Compétences spécifiques : Etudiant(e) en Master 2 en data science, biostatistique, géomatique ou informatique appliquée à la santé, maîtrisant R (idéalement Shiny) et sensible à la visualisation de données. Une appétence pour les outils d'intelligence artificielle générative appliqués au développement (agents de code, LLM) serait un plus. Un intérêt pour la santé publique et la cartographie serait également un plus

.☒ Maîtrise d'un logiciel spécifique : Maîtrise de R. Expérience en R Shiny appréciée, Git. La connaissance ou la curiosité pour les assistants de développement IA (type Cline, Mistral AI) est appréciée.

☒ Autre (préciser) : Profils proactifs, curieux et force de proposition appréciés, avec une bonne aisance relationnelle et orale.

▪ Stage proposé par

Direction : Direction Appui, Traitements et Analyses des données (DATA)

Unité : AMETIS

Maître(s) de stage / personne contact :

	Nom : SOMMEN	Prénom : Cécile
	Fonction : Chargée d'expertise et d'analyse scientifique	
	Téléphone : 01 71 80 16 86	Adresse email : cecile.sommen@santepubliquefrance.fr
Co-encadrement	Nom : FOUILLET	Prénom : Anne
	Fonction : Coordinatrice SURSAUD	
	Téléphone :	Adresse email : anne.fouillet@santepubliquefrance.fr

- **Pour postuler**

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