

Fiche de poste — Ingénieur(e) d'étude

Projet	ALLERGOPEST — Allergies, asthme et pesticides (Grand défi IA & Santé — Data Challenge Health Data Hub 2023-2025)
Porteur du projet	Société Française d'Allergologie (SFA), en partenariat avec le Health Data Hub
Financement	Bpifrance / Conseil de l'innovation, dans le cadre de l'appel à projets Grand défi IA & Santé
Type de contrat	CDD dédié au projet de recherche
Durée du contrat	18 mois (début de contrat idéalement entre le 15 sept et 15 oct)
Lieu(x) d'exercice	Reims (Laboratoire CRIEG - équipe REGARDS, Université de Reims Champagne-Ardenne / Hôpitaux Champagne Sud), avec déplacements ponctuels à Montpellier (IDESP)
Rattachement hiérarchique	Dr Maxime THORIGNY (MCF Économie-Santé, Université de Reims Champagne-Ardenne, Laboratoire REGARDS), en lien avec le Comité de Coordination du projet (Pr Joana VITTE, coordonnatrice)
Budget maximal alloué au poste	≈ 90 000 € (selon profil)

Contexte

ALLERGOPEST vise à créer une base de données nationale ouverte associant des données de santé (SNDS/PMSI) et des données d'exposition non professionnelle aux pesticides, afin de développer, via un data challenge en santé, un outil d'intelligence artificielle prédisant le risque et la sévérité de l'asthme et des allergies respiratoires associés à cette exposition. L'ingénieur(e) d'étude est le pivot technique du projet : il/elle assure la production, le contrôle qualité et la structuration de la base de données géo-économique, puis, en lien étroit avec les équipes médicales et statistiques du consortium, l'appariement des données géo-économiques et de santé.

Missions principales

- Mettre en œuvre la chaîne de traitement géomatique du projet (découpage cadastral, sélection des zones d'usage à partir des données OpenStreetMap, création de buffers autour des bâtiments selon les 10 seuils de distance retenus, croisement avec le Registre Parcellaire Graphique - RPG) sous QGIS.
- Produire, pour chaque commune française (n = 34 806) et chaque année (2014-2023), les fichiers de surfaces agricoles par type et groupe de culture, en vue de l'estimation de l'exposition aux pesticides.
- Participer à la construction de la grille de collecte des données de santé (PMSI et SNDS) en lien avec les médecins de santé publique et épidémiologistes de Reims et Montpellier, et contribuer à l'appariement des données de santé et environnementales.
- Assurer le contrôle qualité, la vérification de la cohérence et la non-réidentifiabilité de la base finale, en lien avec le Délégué à la Protection des Données (DPO) du projet.
- Préparer les jeux de données d'apprentissage, d'évaluation et de test destinés aux compétiteurs du data challenge, en coordination avec la plateforme d'hébergement retenue et le Health Data Hub.
- Contribuer à l'analyse des performances des algorithmes soumis lors du data challenge (Lot 3).
- Participer à la préparation de la base de données et de la documentation associée en vue de sa mise en open data (licence Ouverte / Open Licence Etalab 2.0) et à la valorisation scientifique du projet (Lot 4).

Compétences requises

- Maîtrise des systèmes d'information géographique, en particulier QGIS (version 3.30 ou supérieure) et manipulation de données géoréférencées (RPG, cadastre, OpenStreetMap, IGN).
- Compétences en traitement et gestion de grands volumes de données (le projet prévoit plus de 382 millions de données numériques), idéalement avec des langages de script (Python et/ou R).

- Connaissance des bases de données de santé françaises (SNDS, PMSI) et de leurs contraintes d'accès et de confidentialité appréciée.
- Rigueur méthodologique, sens de l'organisation et capacité à documenter et fiabiliser des protocoles de traitement de données reproductibles.
- Aptitude à travailler en équipe pluridisciplinaire (cliniciens, épidémiologistes, économistes, géographes, data scientists) et à communiquer des résultats techniques à des non-spécialistes.

Profil recherché

Formation Bac+5 (diplôme d'ingénieur ou Master 2) en géomatique, sciences des données, statistique appliquée ou épidémiologie, avec une première expérience (stage, alternance ou emploi) en traitement de données géographiques et/ou de santé. Une sensibilité aux enjeux de santé environnementale et d'exposomique constitue un atout.

Contacts / candidatures

Professeur Joana VITTE, joana.vitte@inserm.fr

M. Maxime THORIGNY, maxime.thorigny@univ-reims.fr