



Proposition de stage de Master 2 pour 2023

Nom de l'entreprise ou du laboratoire : URGI, INRAE

Adresse où se déroulera le stage : INRAE, RD10 Route de Saint-Cyr, 78000 Versailles

Responsable du stage (personne qui sera contactée par les candidats):

- Nom, Prénom : Michotey Célia, Pommier Cyril
- Statut (Ing, chercheur, DR, MCF, Pr, autre) : Ingénieur
- Coordonnées (mél, tél) : celia.michotey@inrae.fr, cyril.pommier@inrae.fr

Titre du stage : Développement et amélioration d'une boîte à outils pour la publication de données dans le portail national Recherche Data Gouv

Mots clés résumant les méthodes et techniques à utiliser au cours du stage :

- Data management, qualité des données, FAIR
- Python, web services REST, Dataverse, JSON, GNU/Linux, Shell.

Résumé du projet de stage :

L'URGI est une unité de recherche INRAE dédiée aux plantes et à leur bioagresseurs. Elle héberge une plateforme de bioinformatique dont l'une des activités majeure est d'assurer le partage, la standardisation et la structuration des données de recherches de projets nationaux et internationaux. Nous utilisons donc le portail national [Recherche Data Gouv](#) (RDG) pour publier [ces jeux de données](#) hétérogènes (génétique, génomique, phénomique). Ils sont ensuite intégrés et rendus accessibles dans les systèmes d'informations à portée internationale que nous développons ([GnpIS](#), [FAIDARE](#)). Notre but est de mettre à disposition des données de qualité suivant les principes FAIR (Findable, Accessible, Interoperable, Reusable) et ainsi participer aux efforts pour la science ouverte.

L'URGI propose un stage de Master en développement de workflow de data management. Il vise à : 1) créer un outil de validation des jeux de données à soumettre dans RDG, 2) intégrer cet outil avec les outils de qualité de données existants, 3) connecter ce workflow au workflow de soumission de données de RDG et 4) publier dans les formats standards de la communauté les datasets validés. Si possible, le/la stagiaire s'intéressera également à l'indexation des données de RDG dans le portail fédératif FAIDARE.

Pour répondre à ces besoins, la personne recrutée utilisera des outils existants (scripts Python, web services REST, Dataverse) et développera des solutions innovantes (utilisation de Frictionless Data, workflow Python). De vrais jeux de données seront également disponibles pour valider le travail réalisé et assurer sa pérennité.

Montant des indemnités de stage

Approximativement 500€.

Modalités de candidature

Les candidatures (CV + lettre de motivation) seront évaluées régulièrement et doivent être adressées au plus tard 2 mois avant le début du stage par courriel à celia.michotey@inrae.fr et cyril.pommier@inrae.fr avec l'objet suivant : **[stage 2023 - data-publication]**