



## Stage Master 2 Intégration connaissances en graphe chez le blé

### Stage Master 2 Intégration de connaissances liées à des ressources génétiques et génomiques, dans une base de données graphe

Partager   Transférer   Imprimer

#### Stage Master 2 Intégration connaissances en graphe chez le blé

#### Stage M2 - Intégration de connaissances liées à des ressources génétiques et génomiques, dans une base de données graphe

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

- Graphe de connaissances, intégration de données, ontologies biologiques et environnementales
- Langages et outil de développement : Neo4j, Python, RDF, conteneur Docker ou Singularity/Apptainer, GNU/Linux, Shell.

#### Résumé du projet de stage

Un des enjeux de la biologie végétale est d'étudier, voire de prédire la capacité à s'adapter à

des stress à partir de données génétiques, génomiques et environnementales. Pour y parvenir les chercheurs étudient des collections de ressources génétiques représentant la diversité existante d'une espèce et souhaitent disposer de connaissances intégrées aussi exhaustives que possible sur chacune des accessions de ces collections.

L'URGI est une unité de recherche basée sur le centre INRAE de Versailles-Saclay, dont un agent est missionné au sein du CNRGV sur le centre INRAE de Toulouse. Elle développe des approches basées sur les graphes de connaissances permettant d'intégrer des données hétérogènes dans l'optique d'apporter un appui efficace et rapide en termes d'exploration de données à la communauté scientifique.

L'objectif du stage proposé sera de compléter le graphe de données existant qui permet pour le moment de collecter et interroger les données et métadonnées génomiques et génétiques d'un large panel d'espèces de plantes d'intérêt agronomique ou représentatives de leur famille botanique. Cette extension se fera sous deux angles :

- L'introduction dans le graphe de données d'informations sur les accessions des collections chez le blé (variétés et accessions<sup>[1]</sup> de plantes, synonymie, caractérisation primaire et phénotypique sur la base d'ontologie ou du thésaurus INRAE<sup>[2]</sup> à faire évoluer, provenance des données) ainsi que l'intégration avec d'autres données disponibles dans le système d'information GnpIS<sup>[3][4]</sup> (génotypage, phénotypage, génomes et annotations, données omiques) pour permettre une exploration de la diversité intra-spécifique. L'intérêt de représenter dans le graphe l'héritage ou l'apparentement des accessions sera étudié.
- La mise en place d'un processus semi-automatisé permettant de réaliser un catalogue de génomes d'intérêt le plus exhaustif possible et associé à des métadonnées riches (lien vers les séquences assemblées et annotations, leurs paramètres de qualité, auteurs, métadonnées de l'échantillon séquencé, ...). Cela passerait par l'évaluation et le choix de certains outils (BioCypher, Data2Service, Monarch Initiative) qui réalisent déjà partiellement ce travail, sur des critères de performance, prise en main, documentation, maintenabilité, éventail de services, communauté de développeurs et d'utilisateurs. Un effort particulier sera attendu sur l'automatisation des processus de collecte, notamment sur la mise à jour et l'ajout de nouvelles données. Nombre d'entre elles sont disponibles dans les bases de données de référence (EBI/NCBI, Phytozome-JGI...), d'autres le sont sous des formats tabulés, enfin des références croisées permettent de lier vers d'autres ressources externes, en partie de la bibliographie.

Quel que soit le lieu d'exercice, le ou la candidate aura sur place un encadrant expérimenté.

## **Adresse où se déroulera le stage (deux localisations possibles)**

INRAE-URGI, Centre INRAE de Versailles-Saclay, RD10, 78000 Versailles

ou

INRAE-CNRGV, Centre INRAE de Occitanie-Toulouse, 24 Chemin de Borde Rouge 31326  
Castanet Tolosan

## Responsable du stage

Nom, Prénom : FLORES Raphaël

Statut : Ingénieur

Coordonnées : [raphael.flores@inrae.fr](mailto:raphael.flores@inrae.fr)

## Montant des indemnités de stage

Indemnités de stage selon barème en vigueur (environ 550 euros net par mois)

## Modalités de candidature

Les candidatures (CV + lettre de motivation) doivent être adressées au plus tard fin 2023 par courriel à [raphael.flores@inrae.fr](mailto:raphael.flores@inrae.fr) avec l'objet suivant : [2024-stage-graphel]

[1] Une même variété peut être reproduite dans plusieurs Centres de Ressources Biologiques (CRB), avec leur propre numéro d'accession, le matériel biologique ayant évolué depuis leur diffusion dans chaque CRB

[2] Thésaurus INRAE : <https://consultation.vocabulaires-ouverts.inrae.fr/thesaurus-inrae/fr/>

[3] C. Pommier *et al.* Applying FAIR Principles to Plant Phenotypic Data Management in GnpIS. *Plant Phenomics*. 2019;2019:1671403. DOI:[10.34133/2019/1671403](https://doi.org/10.34133/2019/1671403)

[4] D. Steinbach *et al.*, GnpIS: an information system to integrate genetic and genomic data from plants and fungi, *Database*, Volume 2013, 2013, bat058, <https://doi.org/10.1093/database/bat058>