

Stage Master – Ingénierie des connaissances

Titre du stage :

Construction d'une ontologie dédiée à la métabolomique

Description du contexte :

INRAE¹ est un institut de recherche public œuvrant pour un développement cohérent et durable de l'agriculture, l'alimentation et l'environnement.

Pour répondre à des questions scientifiques de plus en plus complexes, il est souvent nécessaire de combiner des données provenant de banques de données publiques. La manière de représenter ces données peut varier en fonction des communautés scientifiques ou des outils utilisés pour les produire par exemple. Dans ce cas, il est nécessaire de formaliser la manière de représenter et de nommer les objets du domaine d'étude ainsi que leurs propriétés. Cette étape se concrétise généralement par la constitution de ressources sémantiques telles que les thésaurus et les ontologies.

L'infrastructure nationale de métabolomique MetaboHUB² a pour objectif de fournir des outils technologiques de pointe et des services en métabolomique et fluxomique aux équipes de recherche académiques et à des partenaires industriels dans les domaines de la santé, de la nutrition, de l'agriculture, de l'environnement et des biotechnologies. Dans le cadre de sa nouvelle feuille de route, un des objectifs principaux est de favoriser une meilleure intégration des données et des services MetaboHUB au sein du consortium et au-delà. Les études actuelles sur le métabolisme nécessitent l'utilisation d'un nombre croissant de ressources de données, complexes, produites ou stockées et au sein des plateformes de métabolomique et sur le Web.

C'est dans ce contexte que le projet du stage s'inscrit. Ainsi, le projet **MetaSaurus** réunit des experts en ingénierie des connaissances et en science des données du domaine de la métabolomique avec pour objectif de bâtir un modèle de connaissance adapté aux études métabolomiques en s'appuyant sur les ontologies, thésaurus et vocabulaires contrôlés publiés (Thésaurus INRAE, Ontologies OBO Foundry / BioPortal).

Objectifs du stage :

Produire une ontologie adaptée au domaine de la métabolomique et les supports méthodologiques d'accompagnement de la démarche.

Déroulement du stage :

A partir des spécifications du projet MetaSaurus et de la description des étapes d'intégration de données en métabolomique, le projet consistera à élaborer la première version d'une ontologie dédiée à la métabolomique. Vous étudierez les modalités de réutilisation des ressources sémantiques existantes en lien avec la maintenance de cette ontologie (mis en œuvre des

¹ <https://www.inrae.fr/>

² <https://www.metabohub.fr/>

principes OBO) et intégrerez l'environnement de gestion de l'ontologie dans une forge logicielle. L'implémentation de l'ontologie sera réalisée au format OWL.

Cette démarche fera l'objet d'une proposition de méthodologie générique qui pourra être transposée à d'autres cas d'études.

Compétences requises :

Étudiant de niveau master 2 avec :

- Connaissances générales des standards du web sémantique (OWL/RDF) et des outils : Protégé, Github.
- Aptitude à la rédaction de documentation technique
- Maîtrise de l'anglais technique (lu)

Modalités d'accueil :

Lieu du stage : Site INRAE - Le Rheu (35)

Durée : 6 mois

Gratification : selon le barème en vigueur au 1er janvier 2023 (environ 600 euros)

En rejoignant INRAE, vous bénéficierez (selon le type de contrat et sa durée) :

- d'une restauration collective ;
- d'activités sportives et culturelles.

Environnement scientifique :

Vous serez intégré dans l'équipe RCA (Rendement sous Contraintes Abiotiques) de l'unité IGEPP (Institut de Génétique, Environnement et Protection des Plantes) et vous travaillerez au sein de l'équipe informatique de la plateforme P2M2³ (plateau de profilage métabolique et de métabolomique) dédiée aux analyses de produits du métabolisme végétal. P2M2 dispose d'une infrastructure informatique associée à plusieurs équipements dédiés (systèmes de spectrométrie de masse couplés à des systèmes de chromatographie gazeuse et de chromatographie liquide) pour l'analyse, la caractérisation et la quantification de composés biochimiques complexes dans des matrices végétales.

Le stage se déroulera en interaction étroite avec les ingénieurs de la plateforme « Exploration du Métabolisme » PFEM⁴ et la DipSO⁵, Direction pour la Science Ouverte en charge de l'accompagnement à l'ingénierie des ressources sémantiques à INRAE (sites d'Angers et de Nantes).

Réunion régulières avec les partenaires du projet en visio-conférence et en présentiel.

³ <https://www6.inrae.fr/p2m2/>

⁴ <https://pfem.isc.inrae.fr/>

⁵ <https://www6.inrae.fr/dipso/>

Responsables scientifiques du stage / contact :

Co-encadrants : Sophie Aubin, Olivier Filangi, Magalie Weber

Contact : Olivier Filangi olivier.filangi@inrae.fr