Offre de stage – Conception et mise en place d'un thésaurus de variables pour l'expérimentation agronomique

Contexte

Terres Inovia est l'institut technique agricole dédié aux grandes cultures oléoprotéagineuses et au chanvre. Notre mission est d'améliorer la compétitivité de nos cultures par l'innovation et une expertise indépendante, en adaptant la production agricole et la valorisation des produits aux différents contextes climatiques, économiques et aux demandes sociétales.

À l'ère du numérique, Terres Inovia est amené à gérer toujours plus de données, en particulier des données issues d'essais agronomiques conduits par ses stations expérimentales. L'institut a structuré ses actions de gestion et de valorisation des données au sein de son programme Data & Numérique. Dans ce cadre, la direction de la Data et de l'Innovation Numérique recrute un stagiaire pour concevoir et déployer un référentiel de variables essentiel pour l'homogénéisation et la centralisation des données expérimentales dans un esprit FAIR [1].

Descriptif du stage

Dans le cadre de la mise en œuvre de sa gouvernance des données et d'une politique de gestion et de diffusion de ses données de référence, Terres Inovia souhaite déployer une gestion centralisée de ses variables, en particulier celles utilisées en expérimentation. Ce travail s'inscrit dans le cadre du projet CASDAR Star [2] et d'un partenariat qui comprend l'INRAe (UMR MISTEA de Montpellier) et l'Institut Français de la vigne et du vin (IFV).

Créer et diffuser un standard afin d'homogénéiser le nom et le contenu des variables en interne est un défi important dont pourraient bénéficier des sujets récurrents tels que l'évaluation de variétés, de produits phytosanitaires et de leviers agronomiques. Notre démarche vise à s'inspirer d'outils et de ressources existants (Crop Ontology [3]) pour n'en garder et déployer que les éléments utiles et pertinents pour une application opérationnelle au sein de l'Institut. En effet, les approches et outils du web-sémantique sont très intéressants pour aller vers une gestion optimale des connaissances et en même temps présentent des limites qui peuvent constituer des freins à leur adoption.

L'objectif du stage est de mettre en application ces démarches pour les variables d'observation utilisées par les équipes de Terres Inovia. Le stagiaire, dans un premier temps, devra confronter les modèles de variables existants aux variables actuellement mobilisées par nos outils, puis dans un second temps, traduira nos variables dans le modèle choisi pour en faire à terme un thésaurus, tel que défini dans la publication de présentation du projet CASDAR STAR [4] :

« Un thésaurus est un ensemble organisé de termes contrôlés et normalisés représentant les concepts d'un domaine de connaissance. Les termes sont reliés entre eux par des relations de synonymie (terme équivalent), de hiérarchie (terme générique et terme spécifique) et d'association (terme associé) ; chaque terme appartient à une catégorie ou domaine. »

Les livrables du stage pourront faire l'objet d'une publication via le portail HAL du projet [5].

Profil

Vous êtes actuellement en formation de niveau Bac +4 ou +5 de type école d'ingénieur ou université en agronomie. Vous avez une appétence pour le domaine des sciences des données et l'informatique. Vous aimez prendre des initiatives et être force de proposition. Vous avez un bon relationnel et savez travailler aussi bien de manière autonome qu'en équipe. Les compétences recherchées sont :

- Bases en agronomie
- Bases du langage de programmation R
- Organisation, rigueur, précision et capacités rédactionnelles
- Compréhension écrite de l'anglais
- Facilités pour l'abstraction

Vous serez à monter en compétences dans les domaines suivants :

- Outil de gestion et de versionning de codes informatiques : Git
- Technologies de manipulation et de visualisation de données : R ou Python

Conditions

- Durée du stage : 4 à 6 mois au cours de l'année 2026
- Localisation : Baziège (31)
- Indemnité de stage : 1 000 € / mois
- Tickets restaurant et indemnités réglementaires pour les transports en commun

Modalités de candidature

- CV
- LM où vous chercherez à répondre succinctement à la question suivante :
  « à partir du guide de la Crop ontology [3], comment décomposeriez-vous une variable d'observation nommée « matière fraîche des parties aériennes du colza en entrée hiver /m² » en trait (=entité+attribut) x méthode x unité ? Argumentez vos choix. »

Contact

Pierre-Etienne ALARY
TERRES INOVIA
110 Chemin de la côte Vieille
31450 BAZIEGE

pe.alary [at] terresinovia.fr

Références

[1] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[2] STAR : Standardiser les données expérimentales et techniques pour faciliter leur réutilisation et accélérer l'innovation et le développement agricole : application aux travaux sur les biosolutions. rd-agri.fr/casdar_projet_24acn16050119

[3] Pietragalla, J.; Valette, L.; Shrestha, R.; Laporte, M.-A.; Hazekamp, T.; Arnaud, E. (2024) Guidelines for creating crop-specific Ontology to annotate phenotypic data: version 2.2. Alliance Bioversity International and CIAT. 47 p. https://hdl.handle.net/10568/110906

[4] Florence Amardeilh, Arnaud Charleroy, Baptiste Darnala, Xavier Delpuech, Catherine Roussey, et al.. Standardiser les données expérimentales pour faciliter l'innovation dans le domaine des bio solutions : le projet CASDAR STAR. *8. Atelier INtégration de sources/masses de données hétérogènes et Ontologies, dans le domaine des sciences du VIVant et de l'Environnement (IN-OVIVE), adossé à la conférence Ingénierie des Connaissances @PFIA*, AFIA-Association Française d'Intelligence Artificielle, Jul 2025, Dijon, France. 5p. https://hal.inrae.fr/hal-05074927v1

[5] https://hal.inrae.fr/CASDAR_STAR/