

Stage M2 PRO (R&D)

Analyse, conception, et déploiement d'un système d'intégration de données hétérogènes au sein de l'institut INRAE

contacts : Patrice Buche (patrice.buche@inrae.fr) Pierre Bisquet (pierre.bisquet@inrae.fr) Federico Ulliana (ulliana@lirmm.fr)

Mots clés : *intégration de données, données hétérogènes, ontologies, règles, raisonnements, web sémantique*

Contexte

Cette proposition de stage se place dans le cadre du développement d'un prototype de **système d'intégration de données** hétérogènes au sein d'**INRAE** (*Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement*, www.inrae.fr). Plus précisément, cela sera destiné au soutien de l'infrastructure nationale INRAE **CALIS** (www6.inrae.fr/calis).

Mission

La mission du stage concerne **l'étude et la mise en œuvre d'une solution logicielle** pour l'intégration de données au format **relationnel** et **RDF**. Les développements s'appuieront sur les bibliothèques **OnTop** (<https://ontop-vkg.org/>) et **Graal** (<https://graphik-team.github.io/graal/>) qui permettent des approches pour l'intégration de données à base d'ontologies et de règles.

Objectifs du stage

Les tâches prévues dans le cadre du stage sont les suivantes.

- *Analyse des besoins liés à l'intégration des données*
- *Comparaison des fonctionnalités offertes par OnTop et Graal*
- *Déploiement de l'outil d'intégration*
- *Rédaction de la documentation utilisateur*

Profil

Le candidat idéal présente un **fort intérêt pour** la thématique de **l'intégration de données**. Des compétences en **bases de données relationnelles** et **RDF/SPARQL** sont requises. Des bases solides en **Java** sont nécessaires pour toutes les activités de développement logiciel.

Stage PRO R&D. Cette proposition de stage professionnel porte sur une thématique R&D. Concrètement, cela signifie que le stage concerne une problématique pratique issue d'activités de recherche, mais aucune activité de recherche n'est prévue. Le stage se déroulera dans un laboratoire de recherche, et sera co-encadré par l'équipe ICO de l'UMR IATE d'INRAE (<https://umr-iate.cirad.fr/equipes/ico/presentation>) et l'équipe GraphIK du laboratoire LIRMM (www.lirmm.fr).

Plus de détails. Afin de doter les communautés scientifiques, académiques, comme privées, de capacités technologiques renouvelées et performantes, l'infrastructure nationale de recherche Consommateur/Alimentation/Santé (**CALIS**) a été labellisée par INRAE en 2021 à partir de trois pôles, construits sur des dispositifs collectifs reconnus (eux-même labellisés) de production, de traitement et gestion de données relatives (1) aux comportements des consommateurs, (2) à la construction de la qualité des aliments et à leur caractérisation, et (3) à la caractérisation de leurs propriétés nutritionnelles et de leurs impacts sur la santé chez l'Homme. L'objectif global est de permettre une meilleure exploitation des données produites en interne par l'institut.

L'objectif du stage est de concevoir et prototyper un schéma global d'intégration des sources de données produites par des infrastructures représentatives de la diversité dans CALIS (plateformes procédé alimentaires, plateformes de caractérisation de produits alimentaires, ...). Du point de vue technique, ces plateformes sont toutes pourvues à minima d'un système d'information permettant la réservation des équipements (bases de données relationnelles accessibles par un web service ou fichiers CSV). D'autres équipements capitalisent une sélection de leurs données dans des bases de données RDF [1, 2, 3, 4, <https://www6.inrae.fr/cati-icat-atweb>]. Les données RDF sont accessibles par l'intermédiaire de SPARQL endpoints.

Références

1. L. Ibanescu, J. Dibie, S. Dervaux, E. Guichard, J. Raad, PO2- A Process and Observation Ontology in Food Science. Application to Dairy Gels, in: E. Garoufallou, I. S. Coll, A. Stellato, J. Greenberg (Eds.), MTSR 2016, 2016, Proceedings, Vol. 672 of Communications in Computer and Information Science, 2016, pp. 155–165.
2. A. Muljarto, J. Salmon, P. Neveu, B. Charnomordic, P. Buche, Ontology-based model for food transformation processes - application to winemaking, in: S. Closs, R. Studer, E. Garoufallou, M. Sicilia (Eds.), MTSR2014, 2014. Proceedings, Vol.478 of Communications in Computer and Information Science, Springer, 2014
3. C. Lousteau-Cazalet, A. Barakat, J. P. Belaud, P. Buche, G. Busset, B. Charnomordic, S. Dervaux, S. Destercke, J. Dibie, C. Sablayrolles, C. Vialle, A decision support system for eco-efficient biorefinery process comparison using a semantic approach, Comput. Electron. Agric. 127(2016).
4. M. Munch, P. Willemin, C. E. Manfredotti, J. Dibie, S. Dervaux, Learning probabilistic relational models using an ontology of transformation processes, in: On the Move to Meaningful Internet Systems. OTM 2017 Conferences - Confederated International Conferences: CoopIS, C&TC, and ODBASE 2017, Springer, 2017, pp. 198–215.