

## Semantic application engineer for data modeling and interoperability (2 positions)

- Employer: INRAE
- Title: Semantic Application Engineer
- Topics: semantics, ontology-based modeling, ecological data, relational and graph-oriented database, interoperability, information system, pipelines, data repository
- Submission deadline: Open until filled
- Location: Avignon, France
- Duration: 12 to 16 months
- Date Posted: 5<sup>th</sup> June 2022
- Contact : [christian.pichot@inrae.fr](mailto:christian.pichot@inrae.fr), object = « AnaEE semantics»

INRAE is looking for two application engineers for the implementation of semantic pipelines dedicated to the annotation and publication of data/metadata in ecology.

### General context

---

Scientific communities have created national and international Research Infrastructures (RIs) to share their tools and results. Thus, the Analysis and Experimentation on Ecosystems (AnaEE) RI provides services for the study of continental ecosystems and participates in the European environmental infrastructures (ENVRI) cluster.

In order to ensure interoperability and proper reuse of the data produced, AnaEE is developing a distributed data management strategy based on semantic interoperability. A semantic workflow, prototyped by AnaEE-France RI and ENVRI-plus, enables the automated generation and publication of standardised datasets (NetCDF) and metadata records (geoDCAT and ISO19115/19139, <http://ceur-ws.org/Vol-2969/paper11-s4biodiv.pdf>).

The proposed jobs aim at implementing these services for the different AnaEE experimental platforms in France or in other European countries. The job is supervised by INRAE and carried out in collaboration with CNRS (France) and CREA (Italy).

### Description of the job

---

The engineer will be responsible for the implementation of the data semantic processing workflow. Data are initially managed in the local information systems of the experimental platforms. The first step is to model them as graphs using a domain ontology (the Extensible Observation Ontology [OBOE] extended for experimentation). The second step is to process this information through pipelines developed in shell scripts and Java programs. The initial data are thus converted into semantic metadata graphs which are used to define the scopes of the data/metadata products to be generated and published. The process involves several softwares such as yEd for graph edition, Java *ad hoc* developments for the generation of RDF triples, Blazegraph for their management and exposure (SPARQL endpoint), Dataverse as a file repository and for generation of identifiers (DOI). Pipelines are dockerised for an easier deployment.

The engineer will support the AnaEE experimental platforms in the semantic modeling of their experimentations and the connection of their information systems to the pipeline. He/she will also contribute to the evolution of the functionalities offered by the system.

The engineer will be part of the AnaEE semantics team and will be based at INRAE-PACA, Avignon. He/she will interact with the other engineers of the team as well as with scientists and IT persons from the experimental platforms. One position is devoted to implementation for French platforms and the other for European platforms. Positions will require occasional travels to visit platforms.

**Key qualifications**

- knowledge in core semantic technologies (RDF, OWL, SPARQL) and semantic web
- Experience in the field of ontology based modeling
- Knowledge in SQL, mysql or postgresql RDB environments
- Good knowledge of Linux environment and Shell language
- Experience in data treatment pipelines
- Appreciated knowledge in
  - blazegraph and (web)protege
  - software development (Java)
  - data/metadata formats and standards
  - biological sciences, ecology
- Good communication skills
- Ability to work in a team environment and interact with scientists and technicians
- Fluent technical English for one of the two jobs

**Education and Experience**

- MS or Ph D in Computer Science or relevant field  
OR initial background in ecology and additional education in computer science/ semantic data management
- 2+ years work experience preferred