Proposition de contribution au séminaire INRAE "Semantic Linked Data"

Christian Pichot, INRAE, christian.pichot@inrae.fr
Ph. Clastre, B. Jaillet, D. Maurice, G. Monet, R. Yahiaoui

Titre :
Developing semantic interoperability in ecology and ecosystem studies : semantic modeling and annotation for FAIR data production.

Résumé :
The study of ecosystem functioning requires multidisciplinary approaches and mobilises numerous research teams. The data produced are very abundant but most often poorly standardised. In this context the development of semantic interoperability is a major challenge for the sharing and reuse of data. This objective is implemented within the framework of the AnaEE (Analysis and Experimentation on Ecosystems, ESFRI) Research Infrastructure dedicated to experimentation on ecosystems and biodiversity. The modelling of the experimental system is based on the OBOE ontology extended for disciplinary needs. It covers the measured variables, the different components of the experimental context, from sensor and plot to network, by the atomic decomposition of the observed entities, their characteristics and their qualification, the units and naming standards. This modelling allows the semantic annotation of relational databases and flat files for the production of graph databases.
A distributed Information System (IS) is developed, based on semantic interoperability of its components and the use of common vocabularies (AnaeeThes thesaurus and OBOE-based ontology). Discovery and access portals are fed by information (rdf triples) produced by the semantic annotation of the AnaEE distributed resources: relational databases and modeling platforms.
A first pipeline is developed for the automation of the annotation process and the production of the semantic data, annotation that may represent a huge conceptual and practical work. A second pipeline is devoted to the exploitation of these semantic data through the generation i) of standardized GeoDCAT and ISO metadata records and ii) of data files (NetCDF format) from selected perimeters (experimental sites, years, experimental factors, measured variables...).
Carried out on all the data generated by the experimental platforms, this practice produces semantically interoperable data that meets the linked opendata standards. The work carried out contributes to the development and use of semantic vocabularies within the ecology research community. The genericity of the tools make them usable in different contexts of ontologies and databases.