

EGU22-10213

<https://doi.org/10.5194/egusphere-egu22-10213>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Developing semantic interoperability in ecosystem studies: semantic modelling and annotation for FAIR data production

Christian Pichot¹, Nicolas Beudez², Cécile Callou³, André Chanzy², Alyssa Clavreul², Philippe Clastre¹, Benjamin Jaillet¹, François Lafolie², Jean-François Le Galliard⁴, Chloé Martin³, Florent Massol⁴, Damien Maurice⁵, Nicolas Moitrier², Ghislaine Monet⁶, Hélène Raynal⁷, Antoine Schellenberger⁸, and Rachid Yahiaoui⁸

¹INRAE, UR 0629 URFM, AVIGNON, France (christian.pichot@inrae.fr)

²INRAE, UMR 1114 EMMAH, , AVIGNON, France

³CNRS, UMS 3468 BBEEs, Paris, France

⁴CNRS, UMS 3194 CEREEP, St-Pierre-lès-Nemours, France

⁵INRAE, UMR 1434 SILVA, Champenoux, France

⁶INRAE, UMR 0042 CARTEl, Thonon les Bains, France

⁷INRAE, UMR 1248 AGIR, AUZEVILLE TOLOSANE, France

⁸INRAE, US 1106 InfoSol, Ardon ORLEANS, France

The study of ecosystem characteristics and functioning requires multidisciplinary approaches and mobilises multiple research teams. Data are collected or computed in large quantity but are most often poorly standardised and therefore heterogeneous. In this context the development of semantic interoperability is a major challenge for the sharing and reuse of these data. This objective is implemented within the framework of the AnaEE (Analysis and Experimentation on Ecosystems) Research Infrastructure dedicated to experimentation on ecosystems and biodiversity. A distributed Information System (IS) is developed, based on the semantic interoperability of its components using common vocabularies (AnaeeThes thesaurus and OBOE-based ontology extended for disciplinary needs) for modelling observations and their experimental context. The modelling covers the measured variables, the different components of the experimental context, from sensor and plot to network. It consists in the atomic decomposition of the observations, identifying the observed entities, their characteristics and qualification, naming standards and measurement units. This modelling allows the semantic annotation of relational databases and flat files for the production of graph databases. 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 without such automation. 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 will produce 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.