

Jesualdo Tomás Fernández-Breis¹, José Antonio Miñarro-Giménez¹, María del Carmen Legaz-García¹, Mikel Egana Aranguen²

¹ Facultad de Informática, Universidad de Murcia, CP 30100, Murcia, Spain

² Centre for Plant Biotechnology and Genomics UPM- INIA, CP 28223, Madrid, Spain

emails:{jfernand, jose.minya, mdclg3}@um.es, mikel.egana.aranguen@upm.es

Towards Orthology Linked Datasets

Abstract: OGOLOD[1] is a Linked Open dataset that integrates information about orthology and genetic disorders. This dataset combines information from orthology sources such as Inparanoid, OrthoMCL, COG and Homologene, with genetic disorders information from OMIM. The development of OGOLOD required the creation of the OGO ontology, providing the formal model for our domain of interest. This ontology reuses semantic content from well-known ontologies like the Gene Ontology, the Human Phenotype Ontology, the Relations Ontology and the NCBI Taxonomy.

The integration methodology [2] applied for the development of OGO has two main steps: mapping and generation. The first step is manual and requires the definition of mappings between the resources and the domain ontology, and the definition of identity conditions for the information entities. The second step is automatic: processing and applying the mappings generate the integrated semantic content.

In the last years, we have developed a software tool, SWIT, that allows us to define the mappings between relational or XML data and a domain ontology to generate semantic repositories, in both RDF and OWL. Hence, this tool can be used to apply the previously developed semantic integration methodology. Given that most of currently available orthology resources publish their data in XML or structured files, we believe that our tool would be helpful in order to achieve a common semantic representation of orthology data upon the agreement of the domain ontology. This would be the first step towards the publication of this orthology dataset in the Linked Open Data cloud, which would require the definition of links with external relevant datasets as we did in OGOLOD with LinkedLifeData and Bio2RDF.

References:

[1] Miñarro-Giménez JA, Egaña Aranguen M, Villazón Terrazas B, Fernández-Breis JT, Translational research combining orthologous genes and human diseases with the OGOLOD dataset, Semantic Web journal: Special Call for Linked Dataset descriptions

[2] Miñarro-Gimenez JA, Madrid M, Fernández-Breis JT. OGO: an ontological approach for integrating knowledge about orthology. BMC Bioinformatics vol 10:S10:S13