BioGateway: an integrated RDF store for life sciences

Erick Antezana[1,2], Ward Blondé[1,2], Mikel Egaña[3], Alistair Rutherford[4], Robert Stevens[3], Bernard De Baets[5], Vladimir Mironov[6], and Martin Kuiper[6]

[1] Dept. of Plant Systems Biology, VIB, Gent, Belgium
[2] Dept. of Molecular Genetics, Ghent University, Belgium
[3] School of Computer Science, The University of Manchester, UK
[4] http://www.netthreads.co.uk, Edinburgh, Scotland, UK
[5] Dept. of Applied Mathematics, Biometrics and Process Control, Ghent University, Belgium
[6] Dept. of Biology, Norwegian University of Science and Technology, Norway
http://www.semantic-systems-biology.org/biogateway

Abstract

Life science researchers have been in need for integrated knowledge systems for decades and even more so in recent years. BioGateway is a knowledge base that demonstrates the application of Semantic Web technologies in the domain of Life Sciences: it provides URIs for the data, connected by RDF triples. Unlike other systems it has RDF optimised for querying and it provides a single click access to the data through a library of preconstructed queries.

Two ontologies were created in order to provide a scaffold to integrate all the RDF graphs: MetaOnto and BioMetarel. MetaOnto describes the ontologies that are stored in the system, and BioMetarel provides a set of relationships that are used for linking different resources in RDF triples. A light-weight derivative of BioMetarel (Biorel) was added to each of the graphs in the store. The integration of data in BioGateway has been achieved on the basis of the use of BioMetarel, the use of the same URIs for equivalent resources in the data sources (SwissProt, GOA, NCBI taxonomy) and the orthogonality of OBO ontologies with respect to the classes.

BioGateway provides a library of optimised, easily customisable SPARQL queries that make the resources more accessible to layman users and experts. The library was split into a section with *biological queries* and a section with *ontological queries*. The *biological queries* were designed for the life science users, and they show the most relevant part of the knowledge base. On the other hand, the set of *ontological queries* show how SPARQL can be used to explore the structure of BioGateway.

The visualisation of triple-based resources poses a special challenge. It is necessary to develop and deploy new interfaces to manipulate, query and visualize this knowledge in an intuitive way. The results of queries can be viewed both in a SPARQL browser with colored graphics as in a tabular format.