

Updates from the EBI RDF platform

Thomas Liener¹, Tony Burdett¹, and Simon Jupp¹

EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK
tliener@ebi.ac.uk, tburdett@ebi.ac.uk, jupp@ebi.ac.uk

The EBI RDF platform[1] was released in 2013 and provides access to several EBI databases as RDF Linked Data. The RDF platform provides a SPARQL endpoints and a Linked Data browser to the data in addition to making data dumps available over FTP. At release the RDF platform¹ contained data from the Reactome, BioModels, BioSamples, Gene Expression Atlas and ChEMBL databases in addition to the collaboration with Uniprot to deliver Uniprot RDF. Since the release, additional EBI databases have begun generating RDF exports that will be made available via the RDF platform. These include the Ensembl core data² and the text-mined data from Europe PubMedCentral³. Additional datasets in preparation are the Genome Wide Association Studies (GWAS) dataset⁴, a new version of the Expression Atlas data and the complete set of ontologies from the Ontology Lookup Service (OLS)⁵.

This poster will present the recent updates to the platform and our plans for a consolidated infrastructure to deal with the increasing size and demand on the platform. In a bid to improve the performance of queries that span multiple datasets and provide a consolidated view over the EBI data, we are moving to provide a single SPARQL endpoint to access all the EBI RDF data. We have developed a series of scripts for the automated build and deployment of RDF datasets based on metadata described using the HCLS dataset description best practice guidelines⁶. Over the coming year we will expose analytics of common query patterns (e.g. percentage of queries that combine data from multiple datasets, most heavily accessed concepts and common predicates) that will bring new insight into the usage of the data. We will use these analytics to deliver an enhanced user interface with improved search and browsing capabilities along with smart visualisations that reflect common usage patterns.

References

- [1] Jupp et al. The EBI RDF platform: linked open data for the life sciences, *Bioinformatics* (2014) 30 (9): 1338 - 1339

¹ <http://www.ebi.ac.uk/rdf/>

² <https://www.ebi.ac.uk/rdf/services/ensembl/>

³ <https://europepmc.org/>

⁴ <https://www.ebi.ac.uk/gwas>

⁵ <https://www.ebi.ac.uk/ols>

⁶ <https://www.w3.org/TR/hcls-dataset>