

MEETING ABSTRACT

Open Access

Speeding up research with the Semantic Web

Marco Roos*, Erik A Schultes, Barend Mons

From 6th European Conference on Rare Diseases and Orphan Products
Brussels, Belgium. 23-25 May 2012

Data for Rare Diseases are often distributed. Ideally, we can combine relevant data and biological insights from any place in the world and use it directly as input for computational analysis. However, too often data is poorly described making it hard to find, hard to assess its quality, and hard to integrate with other data. A valid question is: 'Why can't we analyse data as if it came from one global database?'. Here we introduce the Semantic Web as an enabling technology for making data interoperable and thereby expediting biological insight.

The Semantic Web 'language' is RDF: the Resource Description Framework. It uses the 'hyperlink' mechanism known from the internet to refer to data instead of web pages. Meaningful relations are specified as triples: subject, predicate, object. For example, 'CAPN3', 'interacts with', 'ParvB'. Written in RDF:

<http://www.uniprot.org/uniprot/P20807>

<http://www.conceptwiki.org/index.php/Concept:adf6044e-5c2b-11df-b0cb-001517ac506c>

<http://bio2rdf.org/geneid:29780>

While RDF is meant for computers, we see that: (i) RDF triples convey meaning; (ii) hyperlinks specify the location of data, which might be different databases (even within a triple); (iii) data items are also references to other RDF documents with more triples (e.g. try <http://www.uniprot.org/uniprot/Q13547> in a browser). A hyperlink can be in any number of triples, effectively creating the world wide database of meaningfully linked data that is needed in the study of Rare Diseases. Ontologies can also be encoded in RDF, thereby extending the functionality to a global knowledge base. New experiments and discoveries can continually add information to this knowledge base.

For example, the Semantic Web can help us to find drug targets for Rare Diseases. For this purpose, OpenPhacts [1] is integrating compounds from Chemspider [<http://chemspider.com>], proteins from UniProt [[\[uniprot.org\]\(http://www.uniprot.org\)\], pathways from WikiPathways \[<http://wiki-pathways.org>\], and documents from PubMed \[<http://www.ncbi.nlm.nih.gov/pubmed/>\]. We also make DNA sequence variations from the Leiden Open Variation Database \(LOVD \[<http://www.lovd.nl>\]\) available in RDF, and visualised via the UCSC genome browser.](http://</p></div><div data-bbox=)

However, a number of barriers must be overcome. First, databases pre-dating the Semantic Web are used abundantly and must be integrated. This is usually an expensive and tedious task. Secondly, building a scientific reputation often conflicts with data sharing. Therefore, we have developed a data publishing framework called Nanopublication: an application of RDF that links authorship to individual datum (attribution). This creates a transparent and equitable incentive for data sharing. Nano publications also provide incentives for the exposure of legacy data.

In conclusion, Nano publications and Semantic Web technology makes data easier to find and directly applicable to integrative analyses.

Acknowledgements

We thank Frank van Harmelen, Paul Groth and Andrew Gibson for sharing their expertise on the Semantic Web. This work was supported by Open PHACTS [<http://openphacts.org>], funded by the Innovative Medicines Initiative of the EU and EFPIA [<http://www.imi.europa.eu>], and 'Workflow Forever' [<http://wf4ever-project.org>], funded by the Seventh Framework Programme of the European Commission (Digital Libraries and Digital Preservation area ICT-2009.4.1 project reference 270192).

Published: 22 November 2012

Reference

1. Williams AJ, Harland L, Groth P, Pettifer S, Chichester C, Willighagen EL, Evelo CT, Blomberg N, Ecker G, Goble C, Mons B: **Open PHACTS: semantic interoperability for drug discovery**. *Drug Discovery Today* 2012, Available from: <http://linkinghub.elsevier.com/retrieve/pii/S1359644612001936>.

doi:10.1186/1750-1172-7-S2-A11

Cite this article as: Roos et al.: Speeding up research with the Semantic Web. *Orphanet Journal of Rare Diseases* 2012 **7**(Suppl 2):A11.

* Correspondence: M.Roos1@uva.nl

Human Genetics Department, Leiden University Medical Center, Leiden, 2333 ZA, The Netherlands