CORE
Provided by Nature Precedings
Metadata, citation and similar papers at core.ac.uk

Maastricht University Leading in Learning!

CC BY SA

Faculty of Health, Medicine and Life Sciences

Exposing WikiPathways as Linked Open Data

Please scan me
or visit qrate.me/23819

Andra Waagmeester[A], Helena Deus[B], Chris Evelo[A]

[A] Department of Bioinformatics - BiGCaT, Maastricht University, the Netherlands
[B] Digital Enterprise Research Institute, National University of Ireland, Galway

Open PHACTS
Open Pharmacological Space

Introduction

Biology has become a data intensive science. Discovery of new biological facts increasingly relies on the ability to find and match appropriate biological data elements.

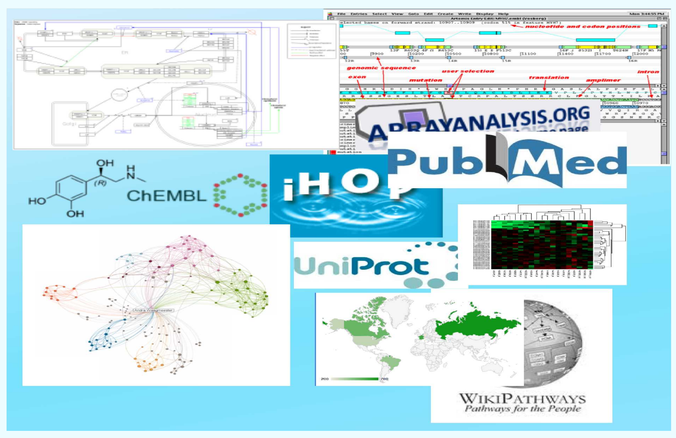


Figure 1. Modern Biologists are required to process and integrate a wide variety of knowledge formats.

Linked Open Data Cloud

Semantic web and Linked Data technologies eliminate the barriers between database silos by relying on a set of standards and best practices for representing and describing data. The architecture of the semantic web relies on the architecture of the web itself for integrating and mapping universal resource identifiers (URI), coupled with basic inference mechanisms to enable matching concepts and properties across data sources. Semantic Web and Linked Data technologies are increasingly being successfully applied as integration engines for linking biological elements.

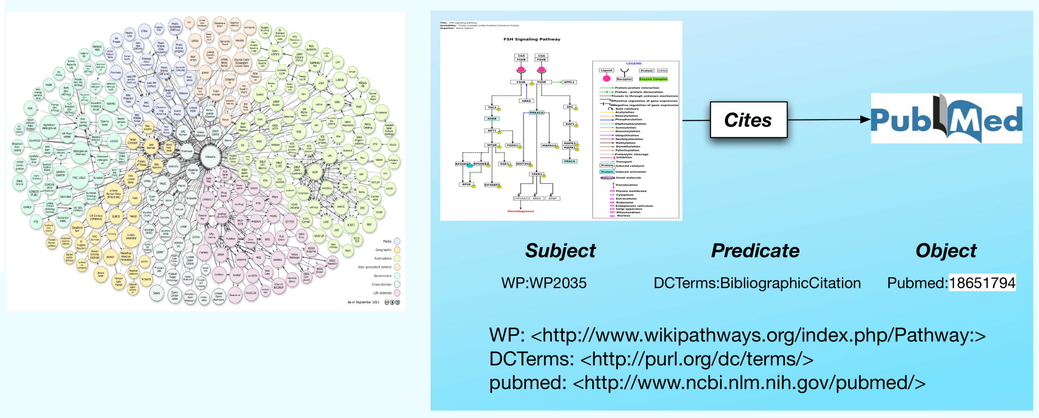
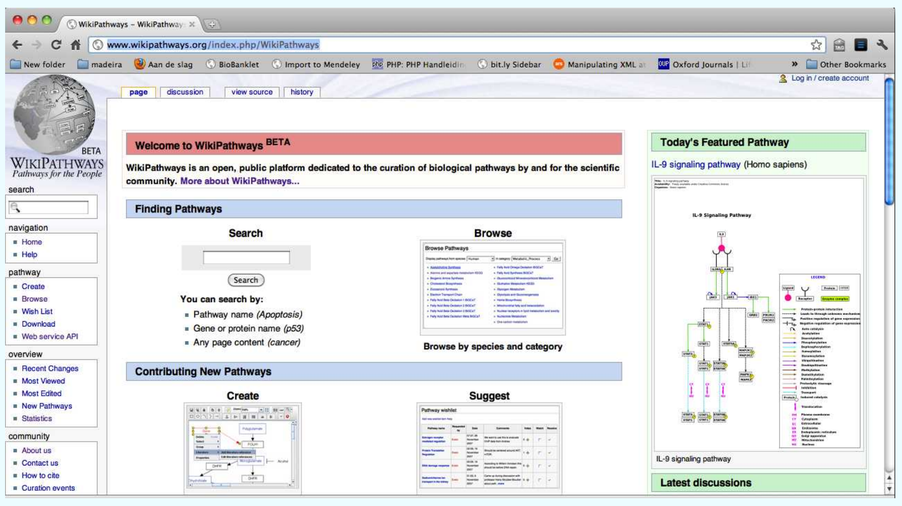


Figure 2: (A) "Linking Open Data cloud diagram (LOD), by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/", (B) Linking to the LOD requires exposing your data in so called triples.

WikiPathways (http://www.wikipathways.org)

WikiPathways is an online repository of Pathways, where the content is curated by the online community. WikiPathways enables automatic querying with 24 predefined functions (ie. Listorganisms, listpathways, findinteractions, etc) (see: http://wikipathways.org/index.php/Help:WikiPathways_Webservice/API)



Exposing WikiPathways as Linked Open Data.

Exposing data to the Linked Open Data Cloud is a 4-step process:

1. Understand your data
2. Select ontologies and controlled vocabularies
3. Map the data and the ontologies/vocabularies
   (e.g. Jena: http://jena.sourceforge.net/ and (tripliser: http://daverog.github.com/tripliser/).
4. Host the data on a triplestore (http://www.w3.org/wiki/LargeTripleStores)

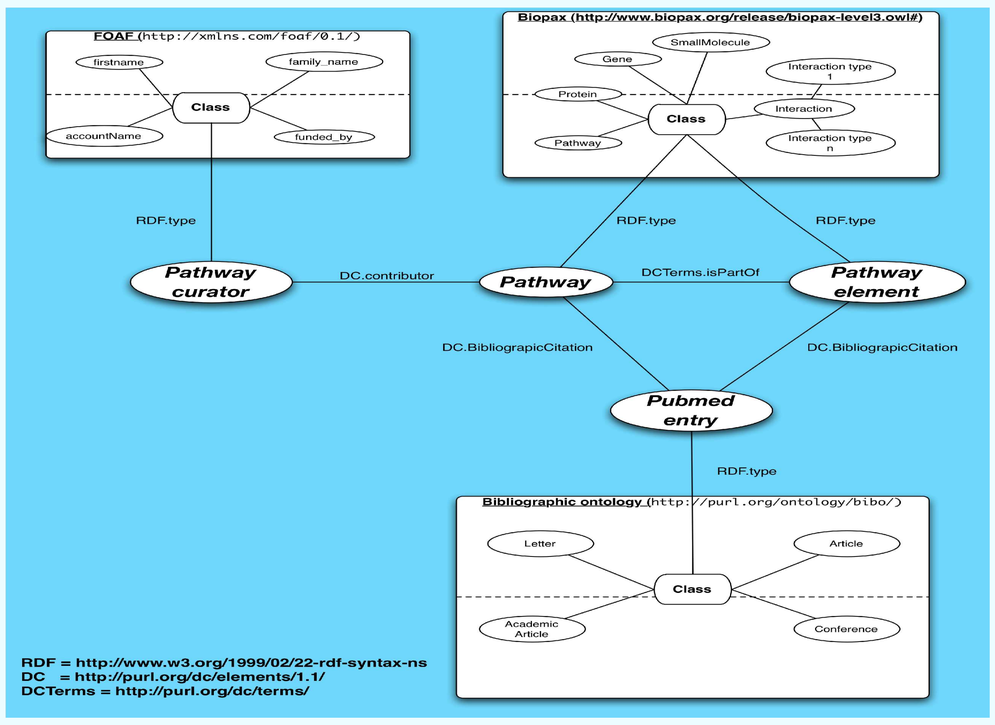


Figure 4: This illustrates a selected mapping of WikiPathways classes to three ontologies/ Controlled vocabularies

Tangible Results

With WikiPathways contact available as Linked Open Data we can:

1. Do SPARQL queries on WikiPathways content at http://semantics.bigcat.unimaas.nl:8000/sparql/
2. Integrate our content with other resources automatically (see Figure 5)
3. provide much more complex API calls through so calles SPARQL queries (see: semantics.bigcat.unimaas.nl/sparqlBookmarker.html)

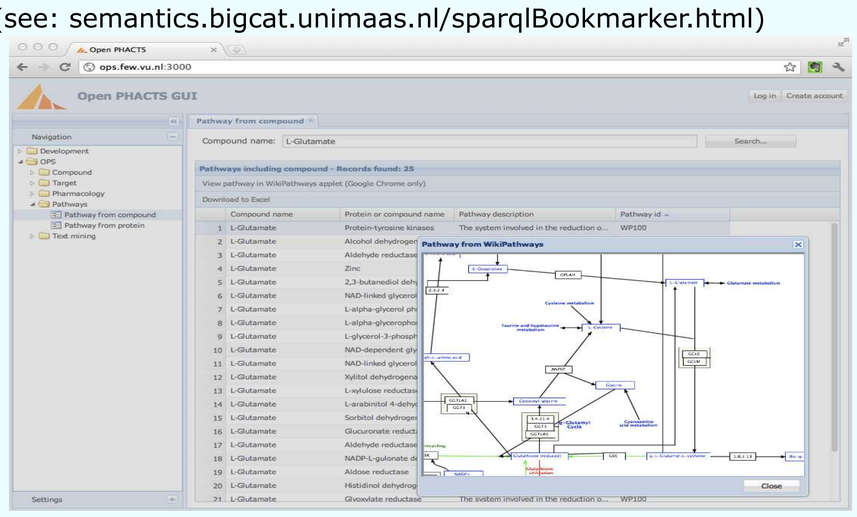


Figure 5: Here WikiPathways LOD is integrated with chemical dataresources which used the same vocabularies and ontologies in a prototype of chemical compound browser. The pathways where a selected set of compounds are active were selected using LOD.

Acknowledgements

We acknowledge financial support provided by the IMI-JU, grant agreement No 115191.

Innovative Medicines Initiative

SEVENTH FRAMEWORK PROGRAMME

Correspondence to:
Andra Waagmeester
Andra.Waagmeester@maastrichtuniversity.nl
www.bigcat.unimaas.nl

Dept of Bioinformatics
T +3143 388 2913
F +3143 388 4149

Maastricht University
P.O. Box 616
6200 MD Maastricht, The Netherlands

Nature Precedings : doi:10.1038/npre.2011.6679.1 : Posted 9 Dec 2011