

New Paper: “The ChEMBL database as linked open data”

Egon Willighagen 

Published May 9, 2013

Citation

Willighagen, E. (2013). New Paper: “The ChEMBL database as linked open data”. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/h24n0-r8e92>

Keywords

Chembl, Rdf, Cito, Cheminf, Ontology

Copyright

Copyright © Egon Willighagen 2013. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

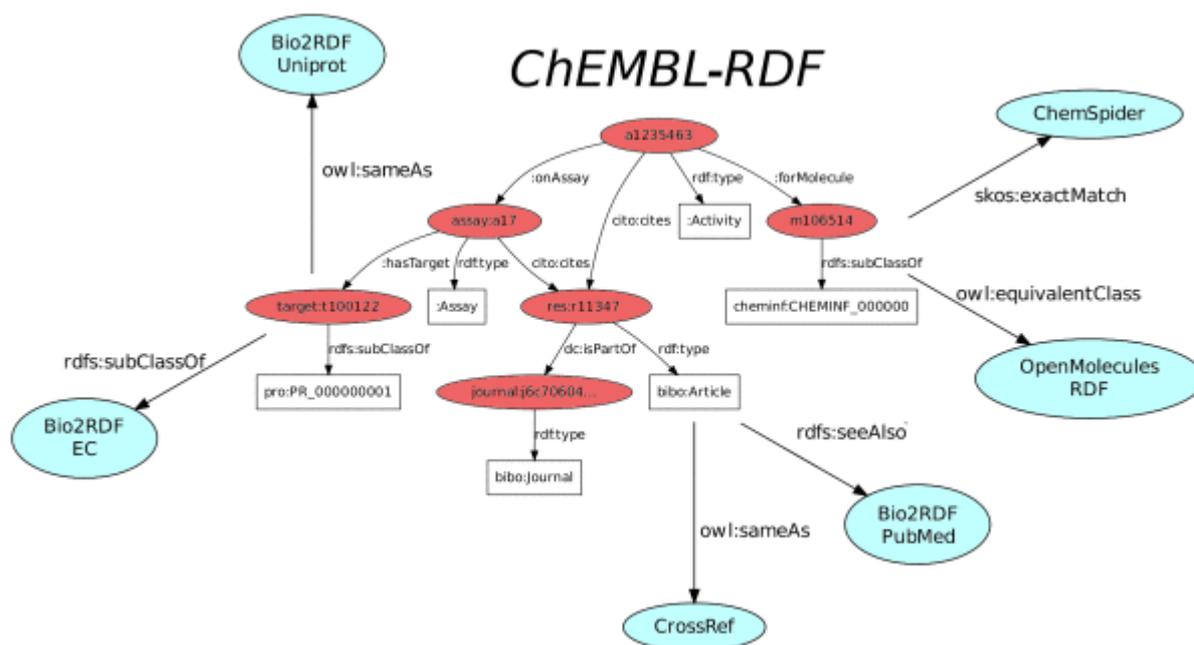
chem-bla-ics

Update: Mark wrote up a [blog post](#) on the RDF that the ChEMBL team itself.

Yesterday, the paper “The ChEMBL database as linked open data” (doi:10.1186/1758-2946-5-23) by Andra Waagmeester (@andrawaag), Ola Spjuth (@ola_spjuth), Peter Ansell (@p_ansell), Antony Williams (@chemconnector), Valery Tkachenko, Janna Hastings, Bin Chen (@binchenindiana), David J Wild (@davidjohnwild), and me appeared in the OA [JChemInf](#) journal.

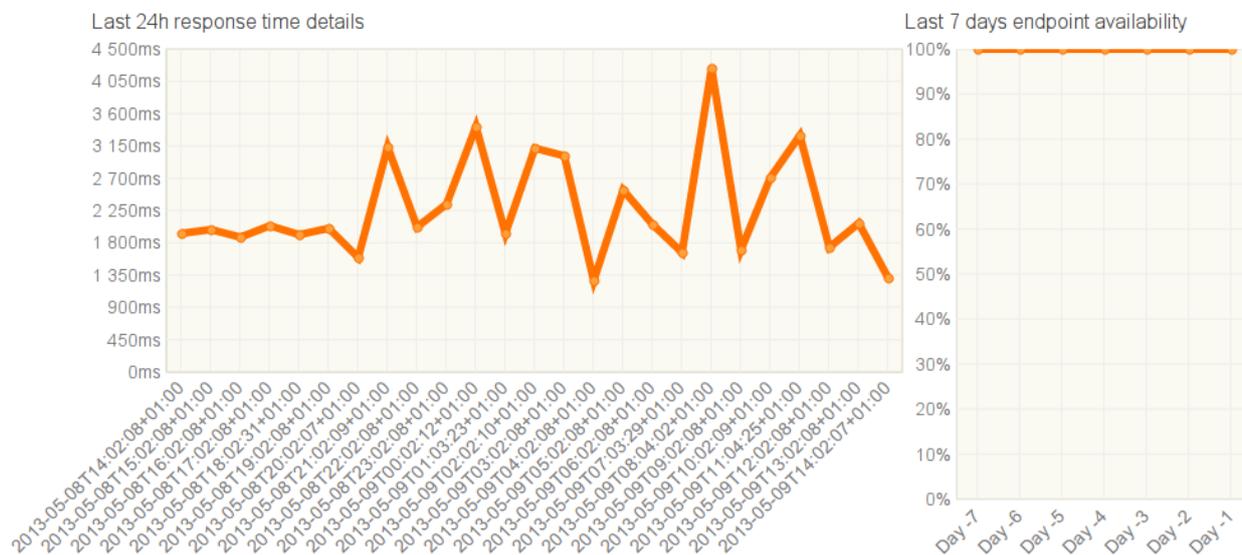
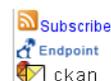
I am also indebted to the [ChEMBL](#) team (@chembl) for both providing such valuable data under a liberal Open Access license and their critical reading of the manuscript! **Additionally, I would like to stress that the ChEMBL team will create their own RDF version of ChEMBL and that this paper is not describing the version they will release.**

BTW, the [source of the paper](#) is available from GitHub. And the (original) [scripts to create RDF from the MySQL dump of ChEMBL](#) are also on GitHub.



This paper outlines the [RDF](#) as it has evolved from various earlier projects. The above diagram visualizes the basic structure (red), various Linked Data resources linked too (blue) and illustrates how various ontologies are used, such as the [CHEMINF](#), [BIBO](#), and [CITO](#) ontologies.

Additionally, various applications and links are described developed by various co-authors. For example, Peter worked on the use in [Bio2RDF](#) and Bin and David on [Chem2Bio2RDF](#). Andra developed an extension for his (#altmetric) [CitedIn](#) resource, giving credit to a paper when data in it is extracted into ChEMBL. Ola, Valery, and Anthony developed a [Bioclipse Decision Support](#) extension, which supports a nearest neighbor search in ChEMBL using [ChemSpider](#). Of course, Ola also hosts the [SPARQL end point](#) of which you can monitor the uptime at the also cool [mondeca.com](#) service:

**ChEMBL-RDF (@ Uppsala University) SPARQL Endpoint Details**[Back to Endpoint Status List](#)

(Yes, I think I have all the cool buzzwords covered in this paper. Sadly, marketing is needed nowadays as a scientist. Where is the time that you could rant on page after page in all your domain specific jargon, not having to worry if your reader would understand it immediately, or without a university degree...)

What this paper does not describe, is all the things I did with ChEMBL-RDF in the [Open PHACTS](#) project ([@Open_PHACTS](#)), which includes the use of [QUDT](#) and the [jQUDT](#) library for unit normalization outlined in [this document](#) and the use of VOID for link sets as described in [this document](#).