chem-bla-ics

Dereferencable InChIs: OpenMolecules RDF

Egon Willighagen

Published April 15, 2012

Citation

Willighagen, E. (n.d.). In *chem-bla-ics*. chem-bla-ics. https://doi.org/10.59350/eg04s-efd96

Keywords

Chemistry, Rdf, Inchi, Opendata

Copyright

Copyright © Egon Willighagen 2012. 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.

About four and a half years ago, I started OpenMolecules RDF, a spin off from Chemical blogspace (Cb, which is still up and running thanks to Peter Maas!) where I started using InChIs in URIs . My interest came from the dereferencability, the ability to take an InChI and find information about the chemical structure representated by it. Because information about anything is scattered around the internet, and we need something decentralized. Moreover, at the time searching of InChIs with search engines like Google did not work well at all: InChIs were tokenized in inconvenient ways.

Originally, these URIs for InChIs were provided (and still are) by Cb, this July five years ago:

`http://cb.openmolecules.net/rdf/?InChI=1/CH4/h1H4`

for which soon after a separate domain was instantiated (thanx to Geoff!):

`http://rdf.openmolecules.net/?InChI=1/CH4/h1H4`

Mind you, **OpenMolecules RDF** is a decent citizen of the Linked Open Data network, though not much linked to. The ChEMBL-RDF data is, and love to hear if there are other link sets pointing there. On the outlinking side, it points to ChEBI (via Bio2RDF), DBPedia, ChemSpider (for 10k structures), the NMRShiftDB, and Cb itself. This post describes the adding of the link to DBPedia .

In the past few years, I have written up bits on OpenMolecules RDF. The main reference is our chapter in *Beautiful Data* (Willighagen, 2010), where I used the URIs for the solubility data . It was later also described in the *Linking the Resource Description Framework to cheminformatics and proteochemometrics paper* (Willighagen, 2011), and another book chapter (Guha, 2011).

This blog features a few more use cases, such as the ability to use these URIs to bookmark molecules or to annotate them with tags with Connotea (which resulted in a nice lunch with the Nature people at the time). The link to Connotea is disabled at the moment, though.

At this moment the system still holds, though there is problem in that browsers can put practical limits on URIs length, which limits the maximum size of the InChI. Virtuoso does this too.