chem-bla-ics

RDF-ing molecular space

Egon Willighagen

Published July 31, 2007

Citation
Willighagen, E. (n.d.). In *chem-bla-ics*. chem-bla-ics. https://doi.org/10.59350/9r29x-8y455

Keywords
Chemistry, Rdf, Inchi, Rdfa

Copyright
Copyright © Egon Willighagen 2007. 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.

RDF might be the solution we are looking for to get a grip on the huge amount of information we are facing. microformats , and RDFa , are just solutions along the way, and Gleaning Resource Descriptions from Dialects of Languages (GRDDL) might be an important tool to get the web RDF-ied.

One important aspect of RDF is that any resource has a unique URI. These make look like a URL or even like `urn:doi:10.1186/1471-2105-8-59`. The recent blogs by Pierre (*URL +1, LSID -1*) and Roderic (*Rethinking LSIDs versus HTTP URI*) illustrate the pro and cons of the different alternatives.

bioGUID

As usual, the bioinformaticians are less conservative and ahead of chemists in trying new options, and several interesting website have emerged. For example, bioGUID makes the bridge between a simple URI and a resolvable URL. And, importantly, it spit RDF. This is the output for http://bioguid.info/doi:10.1109/MIS.2006.62:

```
<?xml version="1.0" encoding="utf-8"?>
<?xml-stylesheet type="text/xsl" href="http://bioguid.info/xsl/html.xsl"?>
<rdf:RDF xmlns:bioguid="http://bioguid.info/schema/0.1/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rss="http://purl.org/rss/1.0/"
  xmlns:prism="http://prismstandard.org/namespaces/1.2/basic/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description rdf:about="http://bioguid.info/doi:10.1109/MIS.2006.62">
    <rdf:type rdf:resource="http://bioguid.info/schema/0.1/Publication"/>
    <rdfs:comment>Generated by transforming XML returned by CrossRef's
      OpenURL service.</rdfs:comment>
    <dc:creator>Shadbolt</dc:creator>
    <dc:title>The Semantic Web Revisited</dc:title>
    <dcterms:issued>2006</dcterms:issued>

    <prism:publicationDate>2006</prism:publicationDate>
    <dc:identifier rdf:resource="doi:10.1109/MIS.2006.62"/>
    <rdfs:comment>info URI scheme</rdfs:comment>
    <dc:identifier rdf:resource="info:doi/10.1109/MIS.2006.62"/>
    <rdfs:comment>CrossRef resolver</rdfs:comment>
    <rss:link>http://dx.doi.org/10.1109/MIS.2006.62</rss:link>
    <prism:publicationName>IEEE Intelligent Systems</prism:publicationName>

    <prism:volume>21</prism:volume>
    <prism:number>3</prism:number>
```

```
    <prism:startingPage>96</prism:startingPage>
    <prism:issn>10947167</prism:issn>
  </rdf:Description>
</rdf:RDF>
```

(BTW, interesting is the use of XSLT to create HTML; it's doing the opposite of GRDDL! And this is probably the right way. Cheers Roderic!)

InChI

I wanted something similar for molecules. The unique identifier is the InChI, of course. The InChI itself is not a proper URI, so I set up a webpage to work around that (if only I had realized this some time ago, I would have urged IUPAC to use the prefix 'inchi:' instead of 'InChI='). The result is, currently, looking like http://cb.openmolecules.net/rdf/rdf.php?InChI=1/CH4/h1H4. I do not use a XSLT yet, but will do so shortly. The RDF looks like:

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:iupac="http://www.iupac.org/">

<rdf:Description
 rdf:about="http://cb.openmolecules.net/rdf/?InChI=1/CH4/h1H4">

 <iupac:inchi>InChI=1/CH4/h1H4</iupac:inchi>

 <pubchem:cid xmlns:pubchem="http://pubchem.ncbi.nlm.nih.gov/#">297</pubchem:cid>
 <pubchem:name xmlns:pubchem="http://pubchem.ncbi.nlm.nih.gov/#">methane</pubchem:name>
 <cb:discussedBy xmlns:cb="http://cb.openmolecules.net/#">http://chemistrylabnotebook.blo
 <cb:discussedBy xmlns:cb="http://cb.openmolecules.net/#">http://wwmm.ch.cam.ac.uk/blogs/
 <cb:discussedBy xmlns:cb="http://cb.openmolecules.net/#">http://chem-bla-ics.blogspot.co
 <cb:discussedBy xmlns:cb="http://cb.openmolecules.net/#">http://chem-bla-ics.blogspot.co

</rdf:Description>

</rdf:RDF>
```

The system uses PHP to create the output, and has a basis pluggable system: a plugin basically spits a RDF fragment for the given InChI, and at this moment it only has a plugin for Cb, but I plan a few more. It needs some tuning and any and all feedback is most welcome. Note that the actual URI might change a bit.