Maastricht University Leading in Learning!

The Web – What is the Issue?

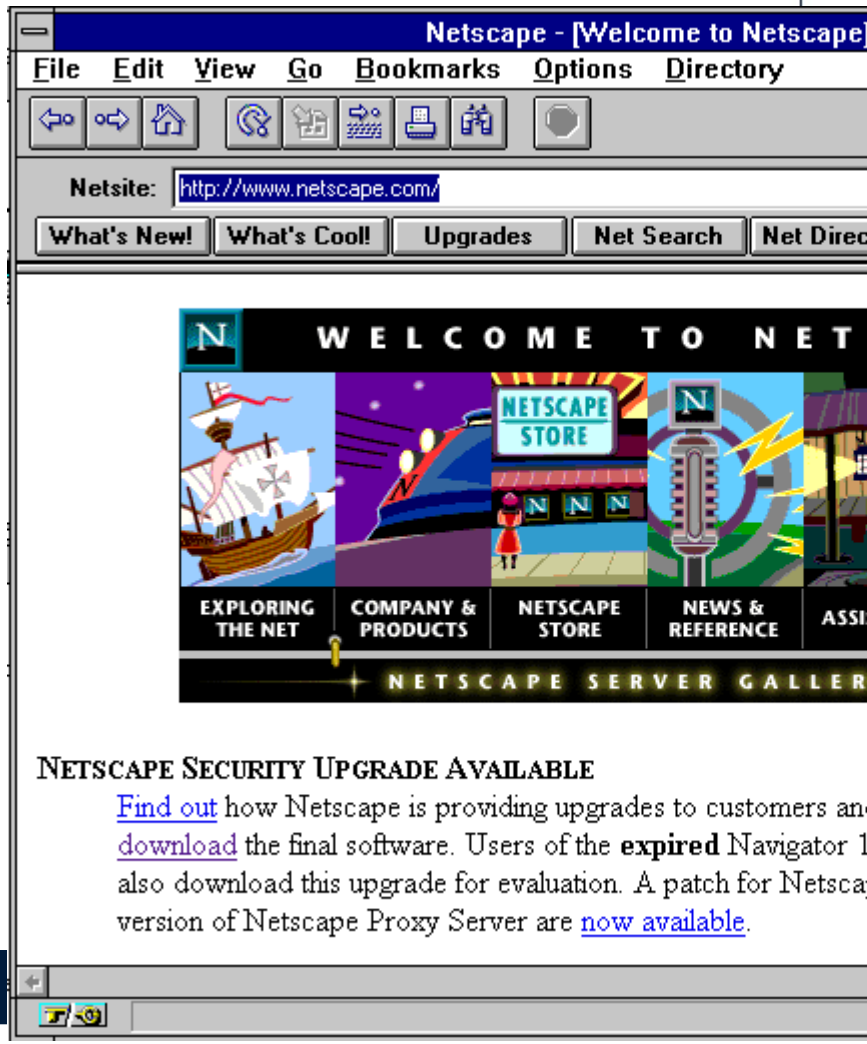
Egon Willighagen (@egonwillighagen)

21 Oct 2014, What's in a Name, Burlington House, London, UK

Web → <html></html>

- World Wide Web
  - *hyperlinked* web pages
- Web Browser
- Nowadays:
  - dynamically created
  - Interactive

- Is there an issue?



Data in HTML

doi:10.1021/ja01193a005.

Table I: Normal Paraffins

Cpd.	t_0	w_0	p_0	SMILES
n-Butane	-0.5	10	1	CCCC
n-Pentane	36.1	20	2	CCCCC
n-Hexane	68.7	35	3	CCCCCC
n-Heptane	98.4	56	4	CCCCCCC
n-Octane	125.7	84	5	CCCCCCCC
n-Nonane	150.8	120	6	CCCCCCCCC
n-Decane	174.0	165	7	CCCCCCCCCC
n-Undecane	195.8	220	8	CCCCCCCCCCC
n-Dodecane	216.2	286	9	CCCCCCCCCCCC

What is HTML?

- Aimed at human
  - "Hypertext"
  - later applets, flash, javascript, HTML5's canvas
- Terminology
  - title, paragraph, link, list, ...
- Visualization
  - bold, italic, ...
- CSS: separate content from looks

HTML draft version timeline [edit]

October 1991

HTML Tags,[5] an informal CERN document listing 18 HTML tags, was first mentioned in pu

June 1992

First informal draft of the HTML DTD,[22] with seven[23][24][25] subsequent revisions (July 1
August 18, November 17, November 19, November 20, November 22)

November 1992

HTML DTD 1.1 (the first with a version number, based on RCS revisions, which start with 1
than 1.0), an informal draft[25]

June 1993

Hypertext Markup Language[26] was published by the IETF IIIR Working Group as an Inter
rough proposal for a standard). It was replaced by a second version[27] one month later, fol
further drafts published by IETF itself[28] that finally led to HTML 2.0 in RFC1866

November 1993

HTML+ was published by the IETF as an Internet-Draft and was a competing proposal to th
May 1994.

April 1995 (authored March 1995)

[20]

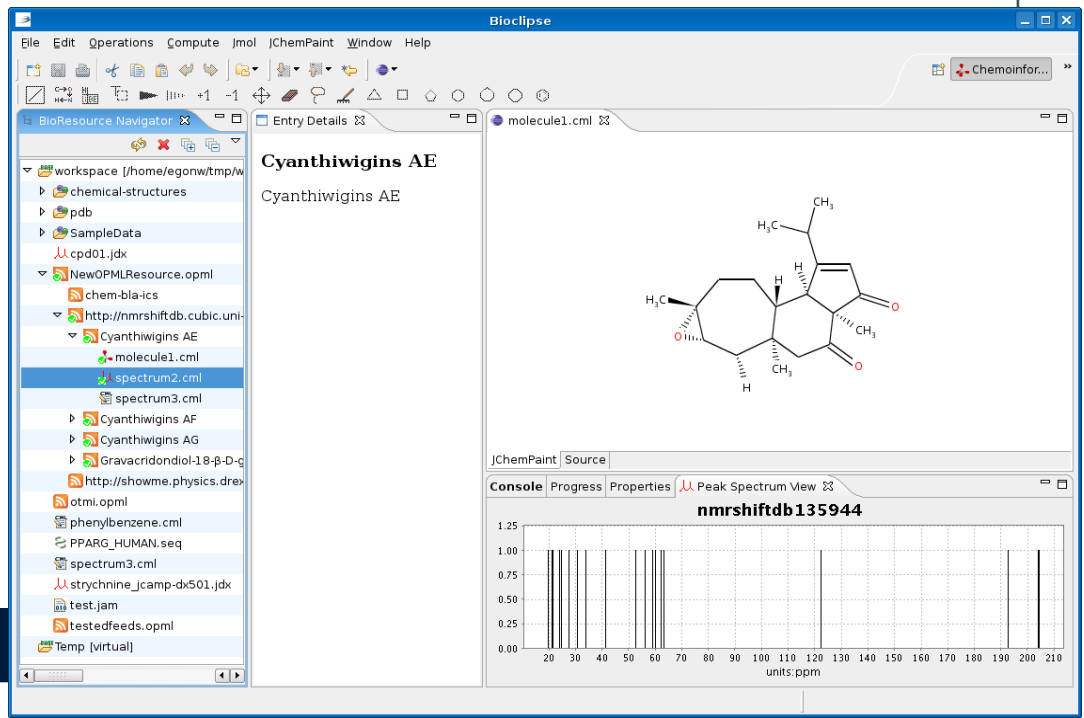
Where are the semantics?

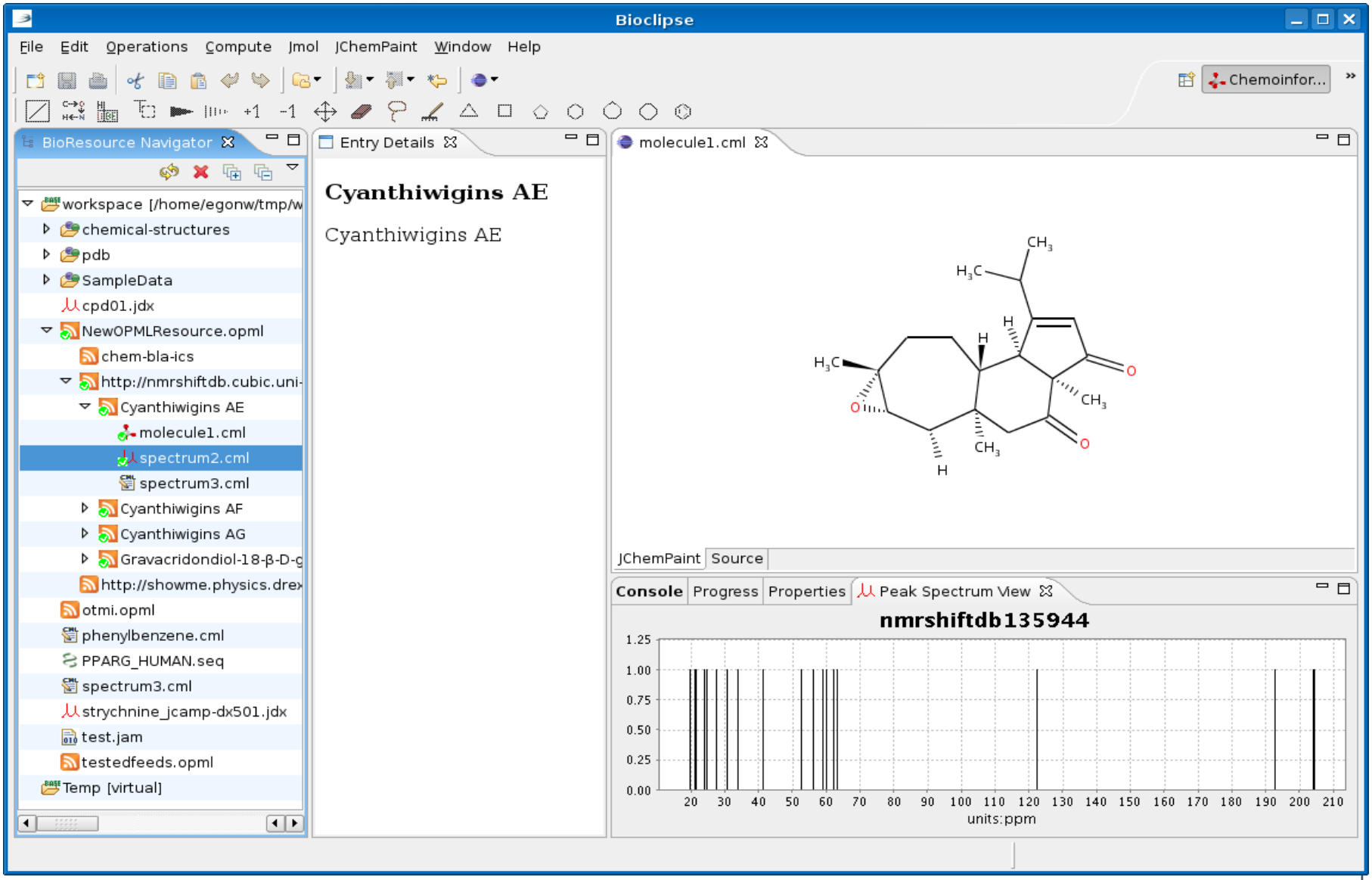
Where are the semantics?

Mama, die, die, die...

Solution #1: <head> + mime types

```
<head>
    <link rel="alternate"
    type="application/rss+xml" title="RSS"
    href="http://some.website.org/rss/" />
</head>
```





Murray-Rust, P., Rzepa, H.S., Williamson, M.J. & Willighagen, E.L. Chemical markup, XML, and the world wide web. 5. applications of chemical metadata in RSS aggregators. J Chem Inf Comput Sci 44, 462-469 (2004)

Solution #1: <head> + mime types

- <head>

  <link rel="meta" type="application/rdf+xml" href="meta.rdf"/>

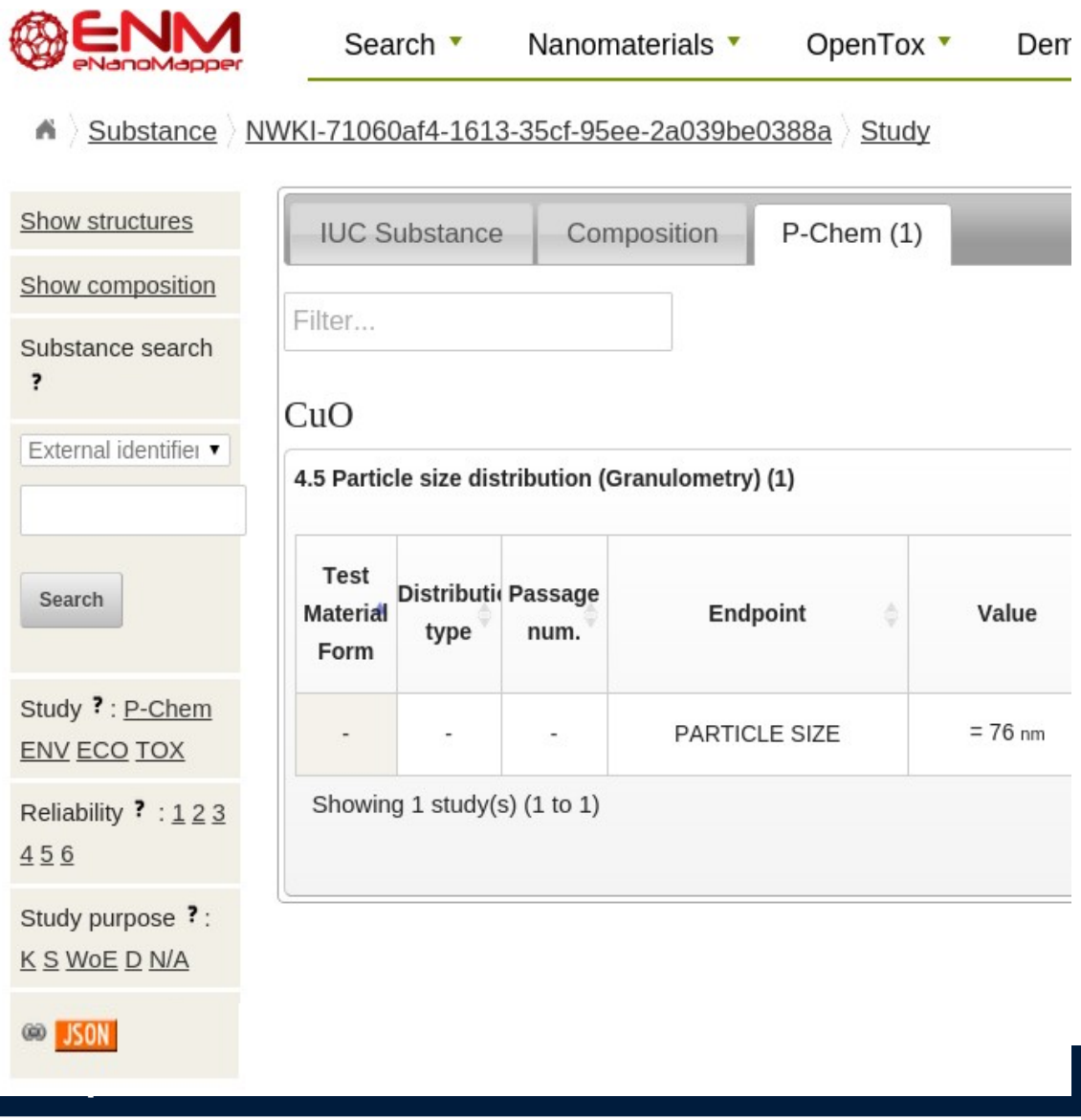
  </head>

- Any format. Yes, also JSON.
- Any Chemical MIME type!

  H. S. Rzepa, P. Murray-Rust and B. J. Whitaker, J. Chem. Inf. Comp. Sci., 1998, 38, 976-982.

Maastricht University Leading in Learning!

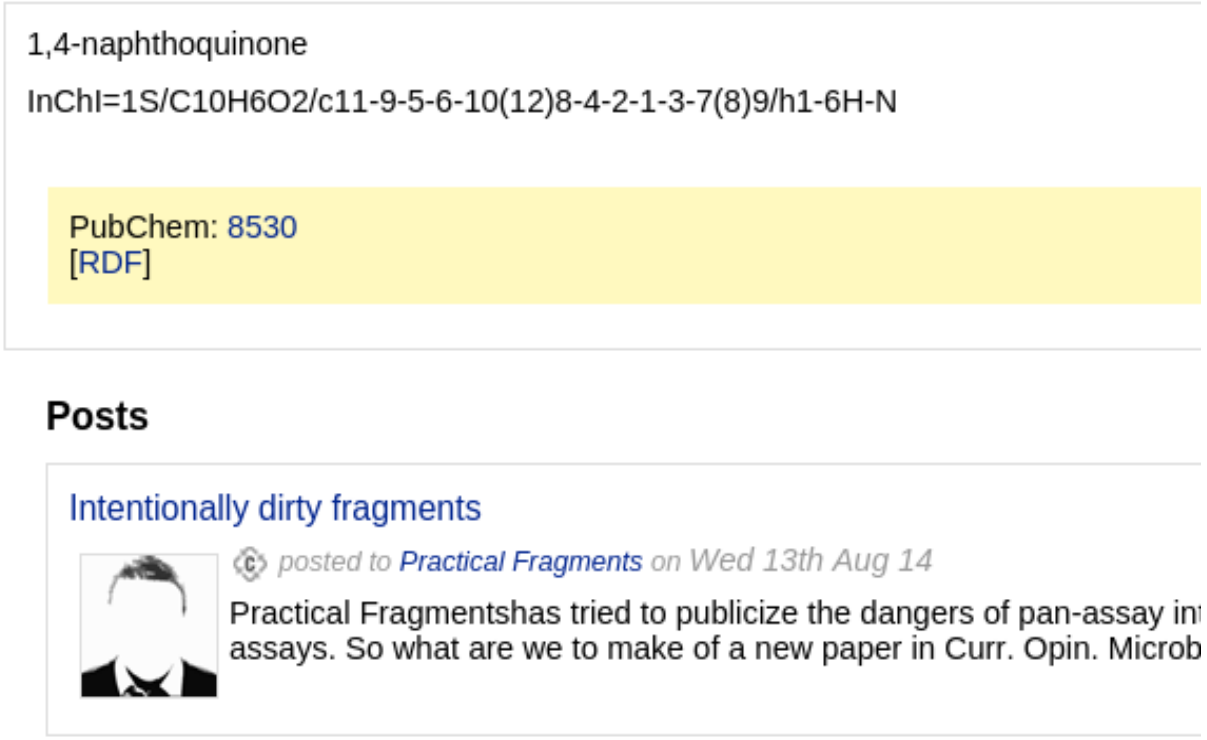
Also JavaScript Object Notation (JSON)




```
{
- study: [
  - {
      uuid: "NWKI-580091eb-ba1e-4500-
    - owner: {
      - substance: {
            uuid: "NWKI-71060af4-161
        },
      - company: {
            uuid: "NWKI-9f4e86d0-c85
            name: "NanoWiki"
        }
      },
    - citation: {
        title: null,
        year: "0",
        owner: null
      },
    - protocol: {
        topcategory: "P-CHEM",
      - category: {
            code: "PC_GRANULOMETRY_S
            title: "4.5 Particle siz
        },
        endpoint: "PARTICLE SIZE",
      - guideline: [
            ""
        ]
      },
    - parameters: {
        DISTRIBUTION_TYPE: null,
        TESTMAT_FORM: null
      },
    - reliability: {
```

Solution #2: microformats

- for SMILES: CCO
- for CAS registry numbers: 50-00-0
- for InChI: InChI=1/CH4/h1H4



Solution #2: microformats

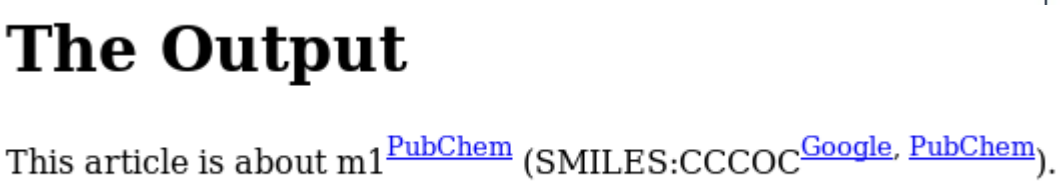
```
<html xmlns="http://www.w3.org/1999/xhtml"
     xmlns:chem="http://www.blueobelisk.org/chemistryblogs/">

<head>
 <title>m1</title>
 <script type="text/javascript" src="sechemtic.js" />
</head>

<body onload="addGoogleAndPubChemLinks(1,1)">
  <h1>The Output</h1>
  <p>This article is about <span class="chem:compound">m1</span>
  (SMILES:<span class="chem:smiles">CCCOC</span>).</p>

</body>

</html>
```



Willighagen, E. et al. Userscripts for the life sciences. BMC Bioinformatics 8, 487+ (2007)

Solution #3: schema.org

Documentation for health/medical types

Without Markup | **Microdata** | RDFa | JSON-LD

```
<div itemscope itemtype="http://schema.org/Drug">
  <h1>Drug Info: <span itemprop="name">PainAwayMax</span></h1>
  PainAwayMax is a brand of
  <span itemprop="alternateName">ibuprofen</span> manufactured by
  <span itemprop="manufacturer" itemscope itemtype="http://schema.org/Organization">
    <span itemprop="name">FooCorp</span>
  </span>.
  It is an
  <span itemprop="drugClass" itemscope itemtype="http://schema.org/DrugClass">
   <span itemprop="name">analgesic</span>
  </span> in the class of
```

Solution #4: HTML+RDFa

doi:10.1021/ja01193a005.

Table I: Normal Paraffins

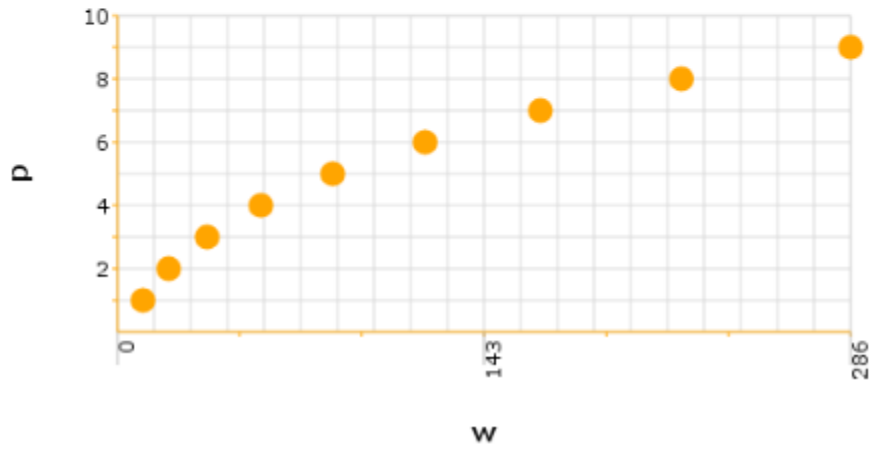
Cpd.	t_0	w_0	p_0	SMILES
n-Butane	-0.5	10	1	CCCC
n-Pentane	36.1	20	2	CCCCC
n-Hexane	68.7	35	3	CCCCCC
n-Heptane	98.4	56	4	CCCCCCC
n-Octane	125.7	84	5	CCCCCCCC
n-Nonane	150.8	120	6	CCCCCCCCC
n-Decane	174.0	165	7	CCCCCCCCCC
n-Undecane	195.8	220	8	CCCCCCCCCCC
n-Dodecane	216.2	286	9	CCCCCCCCCCCC

Solution #4: HTML+RDFa

```
<h2>Table I: Normal Paraffins</h2>
<table>
  <tr>
    <td>Cpd.</td>
    <td>t<sub>0</sub></td>
    <td>w<sub>0</sub></td>
    <td>p<sub>0</sub></td>
    <td>SMILES</td>
  </tr>
  <tr typeof="cc:molecule" about="http://rdf.openmolecules.net/?InChI=1/C4H10/c1-3-4-2/h3-4H2,1-2H3">
    <td><span property="dc:title">n-Butane</span></td>
    <td><span property="cc:t0" datatype="xs:float">-0.5</span></td>
    <td><span property="cc:w0" datatype="xs:integer">10</span></td>
    <td><span property="cc:p0" datatype="xs:integer">1</span></td>
    <td><span property="chem:smiles">CCCC</span></td>
  </tr>
  <tr typeof="cc:molecule" about="http://rdf.openmolecules.net/?InChI=1/C5H12/c1-3-5-4-2/h3-5H2,1-2H3">
    <td><span property="dc:title">n-Pentane</span></td>
    <td><span property="cc:t0" datatype="xs:float">36.1</span></td>
    <td><span property="cc:w0" datatype="xs:integer">20</span></td>
```

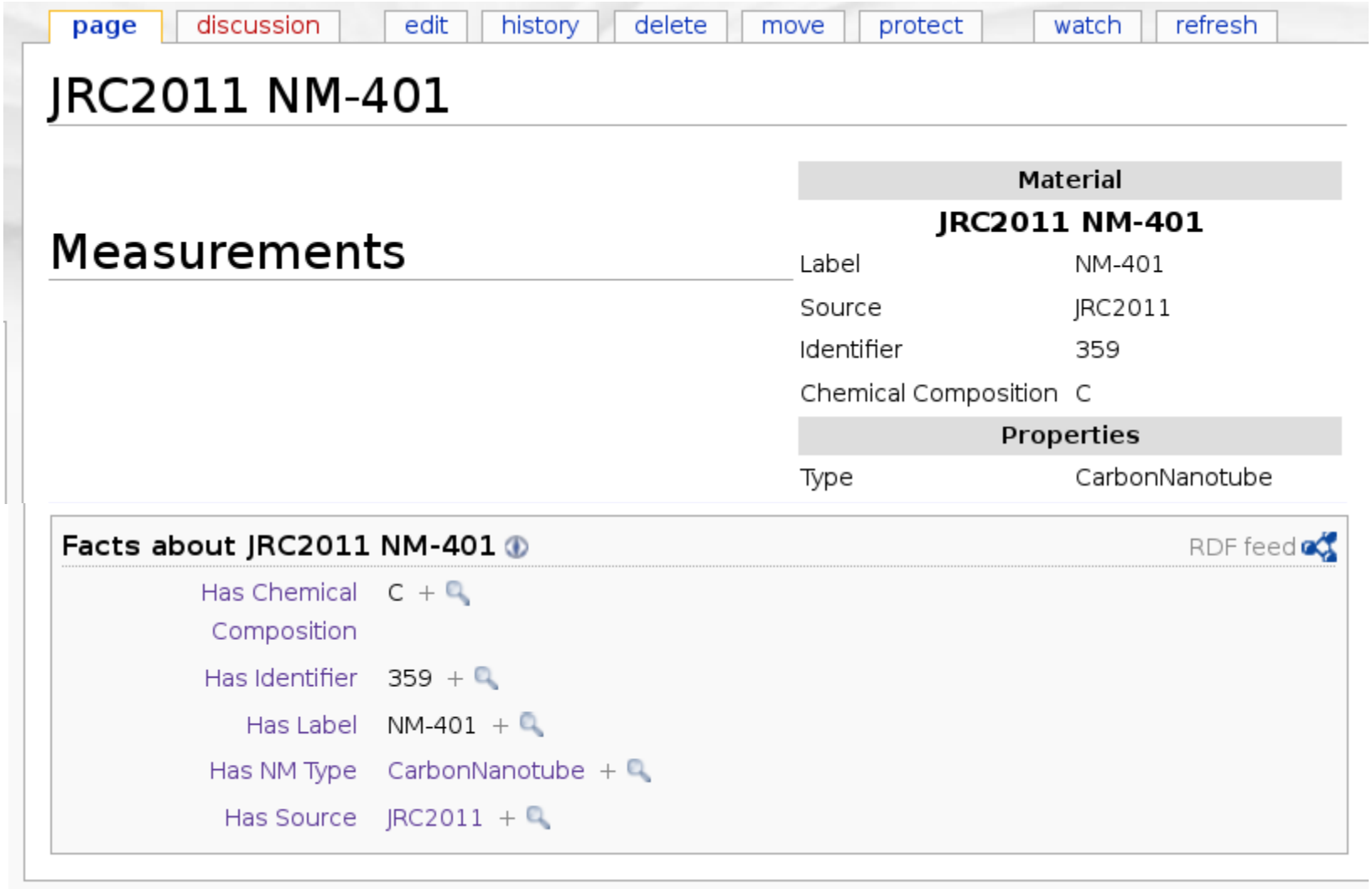
```
PREFIX cc: <http://github.com/egonw/cheminformatics.classics/1/#>
SELECT *
FROM
<http://www.w3.org/2007/08/pyRdfa/extract?
uri=http%3A%2F%2Fegonw.github.com%2Fcheminformatics.classics%2Fclassic1.html&for
mat=pretty-xml&warnings=false&parser=lax&space-preserve=true>
{
    ?mol cc:w0 ?w ;
         cc:p0 ?p .
}
```

Plot Results



Orion Jankowski, SPARQL to Chart, 2010
http://www.pharmash.com/posts/2010-09-27-sparql-to-chart.html

Solution #5: Semantic MediaWiki



Solution #6: RDF/XML + XSLT

OpenMolecules RDF

About	http://rdf.openmolecules.net/?InChI=1/CH4/h1H4
Identifier	info:inchi/InChI=1/CH4/h1H4
InChI	InChI=1/CH4/h1H4
Source	Chemical blogspace
Source	ChEBI
Source	DBPedia
owl:sameAs	http://dbpedia.org/resource/Methane
Source	ChemSpider
owl:sameAs	http://www.chemspider.com/Chemical-Structure.291.rdf#Compound
foaf:homepage	http://www.chemspider.com/Chemical-Structure.291.html
Source	OpenTox
foaf:homepage	http://apps.ideaconsult.net:8080/ambit2/compound/738

More about the Semantic Web

RDF technologies in chemistry

Edited by: Dr Martin Paul Braendle, Dr Egon Willighagen
Collection published: 13 May 2011
Last updated: 19 May 2011

The American Chemical Society (ACS) Division of Chemical Information (CINF) invited scientists from around the world to present their use of RDF technologies in chemistry on 22nd-23rd August 2010 at the 240th ACS National Meeting in Boston, USA. During three half-day sessions, the speakers demonstrated a mix of smaller and larger initiatives where Resource Description Framework (RDF) technologies are used in cheminformatics and bioinformatics as Open Standards for data exchange, common languages (ontologies), and problem solving. This Thematic Series introduces work presented at that meeting, showing the current advances in cheminformatics using these RDF technologies.

Article processing charges for this Thematic Series have been partially funded by Pfizer, Inc. Pfizer, Inc. has had no input into the content of the publication or the articles themselves. All articles in the series have been independently prepared by the authors and have been subject to the journal's standard peer review process.

Methodology Open Access Highly accessed
Chemical Entity Semantic Specification: Knowledge representation for efficient semantic cheminformatics and facile data integration
Leonid L Chepelev, Michel Dumontier
Journal of Cheminformatics 2011, **3**:20 (19 May 2011)
Abstract | Full text | PDF | PubMed | Cited on BioMed Central | Graphical abstract

The Web – What is the issue?

We're not using these innovative solutions enough.

Acknowledgments

- People whom I worked with on solutions
  - H. Rzepa, P. Murray-Rust, P. Maas, N. Jeliazkova, N. O'Boyle, R. Guha, C. Steinbeck, D. Wild, ...
- Projects that allow(ed) me to continue adoption of these methods
  - Open PHACTS (IMI-funded)
  - eNanoMapper (EC-funded)

- WikiPedia contributors (for images)