Semantic Web, Linked Data, and Semantic 3D Media

Ivan Herman, W3C

FOCUS 3D Conference, Sophia Antipolis, France
2010-02-11
Significant buzz…

- There is quite a buzz around “Semantics”, “Semantic Technologies”, “Semantic Web”, “Web 3.0”, “Data Web”, etc, these days
- New applications, companies, tools, etc, come to the fore frequently
Significant buzz...

- It is, of course, not always clear what these terms all mean:
  - “Semantic Web” is a way to specify data and data relationships; it is also a collection of specific technologies (RDF, OWL, GRDDL, SPARQL, …)
  - “Semantic Technologies”, “Web 3.0” often mean more, including intelligent agents, usage of complex logical procedures, etc
Lots of Tools (not an exhaustive list!)

- **Categories:**
  - Triple Stores
  - Inference engines
  - Converters
  - Search engines
  - Middleware
  - CMS
  - Semantic Web browsers
  - Development environments
  - Semantic Wikis
  - …

- **Some names:**
  - Jena, AllegroGraph, Mulgara, Sesame, flickurl, …
  - TopBraid Suite, Virtuoso environment, Falcon, Drupal 7, Redland, Pellet, …
  - Disco, Oracle 11g, RacerPro, IODT, Ontobroker, OWLIM, Talis Platform, …
  - RDF Gateway, RDFLib, Open Anzo, DartGrid, Zitgist, Ontotext, Protégé, …
  - Thetus publisher, SemanticWorks, SWI-Prolog, RDFStore…
  - …
The main lesson:

- *Anybody can start developing Semantic Web applications*
There is a great community

- There are *lots* of tutorials, overviews, and books around
  - some of them good, some of them bad, just as with any other areas…
- Active developers’ communities
  - blogs, IRC channels, mailing lists, various fora: more than what one person can oversee…
- Some measures claim that there are over $10^7$ Semantic Web documents on the Web
Some deployment communities

- Major communities pick the technology up: digital libraries, defence, eGovernment, energy sector, financial services, health care, oil and gas industry, life sciences, social web applications …
So what *is* the Semantic Web?
There is a growing number of application patterns referring to the Semantic Web:

- data integration using RDF, SKOS, OWL, …
- knowledge engineering with complex ontologies
  - using, e.g., OWL and/or rule based reasoning
- better data management, archiving, cataloguing, etc
  - e.g., digital library applications
- managing, coordinating, combining Web services
- intelligent software agents
- improving search (usually using domain specific vocabularies…)
- etc
Is this where we are?
Maybe, but being an elephant is not necessary bad! 😊

- it shows that the Semantic Web is a mature technology
- that there is lots of interest, applications
- various application areas pick what they need...
  - e.g., some need sophisticated knowledge management, so they go for complex ontology’s…
  - some concentrate on semantically simpler vocabularies but large volume of data
- …and that is fine, there is room for many!
But it is good to (re-)emphasize some principles

The Semantic Web:

- extends the principles of the Web from documents to data; create a Web of data
• It is the Semantic **Web**, and not only Semantics!
  • data, ontologies, vocabularies, etc, can (and should!) be shared, reused, potentially on Web scale
    • *the “network effect” on data*

• The major importance of RDF is that it provides an abstract integration layer for data on the Web
Applications are not always very complex…

- Eg: simple semantic annotations of data provides easy integration (eg, with MusicBrainz, Wikipedia, geographic data sets, etc)

- What is needed: some simple vocabularies, simple annotation
  - annotation can be generated by a server automatically, or
  - added by the user via some user interface

- This extra data can be in some microformats, in RDFa, …
A relatively simple application

- Goal: reuse of older experimental data
- Keep data in databases or XML, just export key “fact” as RDF
- Use a faceted browser to visualize and interact with the result

Courtesy of Nigel Wilkinson, Lee Harland, Pfizer Ltd, Melliyal Annamalai, Oracle (SWEO Case Study)
But is there already a Web of Data out there?
Linking Open Data Project

- Goal: “expose” open datasets in RDF
- Set RDF links among the data items from different datasets
- Set up query endpoints
- Billions triples, millions of “links”
Example data source: DBpedia

- **DBpedia** is a community effort to:
  - extract structured ("infobox") information from Wikipedia
  - provide a SPARQL endpoint to the dataset
  - interlink the DBpedia dataset with other datasets on the Web
Extracting structured data from Wikipedia

@prefix dbpedia <http://dbpedia.org/resource/>.
@prefix dbterm <http://dbpedia.org/property/>.

```
dbpedia:Amsterdam
  dbterm:officialName "Amsterdam" ;
  dbterm:longd "4" ;
  dbterm:longm "53" ;
  dbterm:longs "32" ;
  ...;
  dbterm:leaderName dbpedia:Job_Cohen ;
  ...;
  dbterm:areaTotalKm "219" ;
  ...;
dbpedia:ABN_AMRO
  dbterm:location dbpedia:Amsterdam ;
  ...
```
Automatic links among open datasets

Processors can switch automatically from one to the other…
The real value is in the links!
Applications can access the data directly (via the URI-s)

There are several “instantiations” of part of the cloud that user can access

- these store copies of several “blobs”
  - possibly with some inferred triples based on, eg, OWL
- often offering a SPARQL endpoint to query to cloud
Example for cloud exploration
**Example for cloud exploration**

<table>
<thead>
<tr>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosts of the Summer Olympic Games</td>
<td>populated place</td>
</tr>
<tr>
<td>Port cities and towns in the Netherlands</td>
<td>second-order administrative division</td>
</tr>
<tr>
<td>Port cities and towns of the North Sea</td>
<td>Settlements established in the 13th century</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>alternate name</th>
<th>Amsterdam@en</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>Amsterdam, Gemeente Amsterdam</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>population</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000000</td>
</tr>
<tr>
<td>738434</td>
</tr>
<tr>
<td>741636</td>
</tr>
<tr>
<td>743000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>children features</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://sws.geonames.org/2759793/contains.rdf">http://sws.geonames.org/2759793/contains.rdf</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>feature class</th>
</tr>
</thead>
<tbody>
<tr>
<td>geo-ont:A</td>
</tr>
<tr>
<td>geo-ont:P</td>
</tr>
<tr>
<td>geo-ont:P</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>feature code</th>
</tr>
</thead>
<tbody>
<tr>
<td>capital of a political entity</td>
</tr>
<tr>
<td>populated place</td>
</tr>
<tr>
<td>second-order administrative division</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>country</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.geonames.org/countries/NL">http://www.geonames.org/countries/NL</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>map</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.geonames.org/2759793/gemeente-amsterdam.html">http://www.geonames.org/2759793/gemeente-amsterdam.html</a></td>
</tr>
<tr>
<td><a href="http://www.geonames.org/2759794/amsterdam.html">http://www.geonames.org/2759794/amsterdam.html</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>nearby features</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://sws.geonames.org/2759794/nearby.rdf">http://sws.geonames.org/2759794/nearby.rdf</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>parent feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kingdom of the Netherlands, Provincie Noord-Holland</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>wikipedia article</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://en.wikipedia.org/wiki/Amsterdam">http://en.wikipedia.org/wiki/Amsterdam</a></td>
</tr>
</tbody>
</table>
Example for cloud exploration
Example for cloud exploration
Example for cloud exploration

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kingdom of the Netherlands</td>
<td>capital</td>
</tr>
<tr>
<td>Netherlands</td>
<td>location</td>
</tr>
<tr>
<td>A-Film Distribution B.V.</td>
<td></td>
</tr>
<tr>
<td>ABN AMRO Holding N.V.</td>
<td></td>
</tr>
<tr>
<td>Akzo Nobel N.V.</td>
<td></td>
</tr>
<tr>
<td>Amsterdamse Trade Bank N.V.</td>
<td></td>
</tr>
<tr>
<td>APX B.V.</td>
<td></td>
</tr>
<tr>
<td>Armada Music</td>
<td></td>
</tr>
<tr>
<td>Barlaeus Gymnasium</td>
<td></td>
</tr>
<tr>
<td>Bugaboo International B.V.</td>
<td></td>
</tr>
<tr>
<td>Celtel</td>
<td></td>
</tr>
<tr>
<td>Central European Media Enterprises Ltd.</td>
<td></td>
</tr>
<tr>
<td>CNH Global N.V.</td>
<td></td>
</tr>
<tr>
<td>Commodore Gaming BV</td>
<td></td>
</tr>
<tr>
<td>Corporate Express N.V.</td>
<td></td>
</tr>
<tr>
<td>dance4life</td>
<td></td>
</tr>
<tr>
<td>De Bijenkorf</td>
<td></td>
</tr>
<tr>
<td>De Meer Stadion</td>
<td></td>
</tr>
<tr>
<td>Delta Lloyd Groep N.V.</td>
<td></td>
</tr>
</tbody>
</table>
This is a major community project

- anybody can participate; to subscribe to the list:
  - http://lists.w3.org/Archives/public/public-lod/

- or look at the project site:
  - http://esw.w3.org/topic/SweolG/TaskForces/CommunityProjects/LinkingOpenData

- if you know of open data sets: contact the project to incorporate it with the rest!
Using the LOD to build Web site: BBC
Using the LOD to build Web site: BBC
Using the LOD to build Web site: BBC
Using the LOD cloud on an iPhone

Courtesy of Chris Bizer and Christian Becker, Freie Universität, Berlin
Using the LOD cloud on an iPhone

Courtesy of Chris Bizer and Christian Becker, Freie Universität, Berlin
Using the LOD cloud on an iPhone

Search Engines
- Sindice
- FalconS

Marbles Engine

Amazon EC2

Shared Cache

HTTP GET

Linked Data on the Web

Courtesy of Chris Bizer and Christian Becker, Freie Universität, Berlin
Publication of data: Library of Congress
Subject Headings
Publication of data: Library of Congress
Subject Headings
And what about ontologies?

- I.e.: where does, e.g., OWL come into the LOD picture?
- Ontologies are necessary to properly integrate data
  - “term used in this dataset relates to the term used there this and this way…”
- OWL, Rules, RDF vocabularies are vital
And what about ontologies?

- But…
  - Ontologies/vocabularies can be very simple (few terms)
  - Expressivity vs. complexity of management and usage has always be balanced
  - *You are perfectly decent Semantic Web citizen even if you do not use complex OWL (or not use OWL at all…)*
How does this apply to Semantic 3D?
Caveat: I am an outsider, sorry if I bang on open doors…
Think of the data out there from the start!

- Modeling 3D objects with Semantic technologies (OWL, SKOS, etc) is important
  - (and looks fairly complex from where I stand…)
- But… think of the data out there!
  - applications may use this in many different ways…
- Also: contribute if you can, make your data widely available!
An artificial example

- There is, of course, the 3D modeling aspect
- But there may be, also, additional data on the artifact
- This can be connected to the outside world
An artificial example

### About:

**http://dbpedia.org/resource/Ramesses_I**

An Entity in Data Space: dbpedia.org

Merneptah Ramesses I was the founding Pharaoh of Ancient Egypt's 19th dynasty. The dates for his short reign are not completely known but the time-line of late 1292-1290 BC is frequently cited as well as 1295-1294 BC.

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbprop:wikimedia:entity</td>
<td><a href="http://upload.wikimedia.org/wikipedia/commons/thumb/1/16/Statue_ram-c">http://upload.wikimedia.org/wikipedia/commons/thumb/1/16/Statue_ram-c</a></td>
</tr>
<tr>
<td>dbprop:abstract</td>
<td>Merneptah Ramesses I was the founding Pharaoh of Ancient Egypt’s 19th Dynasty. In reality his brief reign marked the end of the Middle Kingdom, but he is sometimes included in the 18th Dynasty.</td>
</tr>
<tr>
<td>dbprop:birth</td>
<td>1457 BC</td>
</tr>
<tr>
<td>dbprop:death</td>
<td>1290 BC</td>
</tr>
<tr>
<td>dbprop:father</td>
<td>Amenemhat III, Nineteenth Dynasty of Egypt</td>
</tr>
<tr>
<td>dbprop:gender</td>
<td>Male</td>
</tr>
<tr>
<td>dbprop:height</td>
<td>5'02”</td>
</tr>
<tr>
<td>dbprop:hair</td>
<td>Black</td>
</tr>
<tr>
<td>dbprop:hasPhotoCollection</td>
<td><a href="http://www4.wiwiss.fu-berlin.de/flickrapp/photos/Ramesses_I">http://www4.wiwiss.fu-berlin.de/flickrapp/photos/Ramesses_I</a></td>
</tr>
</tbody>
</table>
An artificial example
You can of course make it nicer...
Obviously, there are other datasets

- Use Geodata for precise information on Egypt
- Use the LOC data to give precise subject descriptions
- …
Make the data available!

- Make the data available to others!
- Wouldn’t it be cool to see your data appear on an iPhone? 😊
What does it mean in practice?

- Add (meta)data to your artifacts
  - use http URI-s
  - use public vocabularies if possible
- *Add links to other public datasets*
  - that is how others will find you!
- Make your data and vocabularies public
- If you can: set up a SPARQL endpoint
Practice in other areas

- I used a very “webby” example with Wikipedia
- Of course other areas have their own datasets that can be used
Colored LOD cloud...
Eli Lilly’s Target Assessment Tool

Courtesy of Susie Stephens, Eli Lilly (SWEO Case Study)
Linked Open eGov Data (US example)
You publish the raw data, others use it...

Examples from RPI’s Data-gov Wiki, Jim Hendler & al.
An example with UK government data

Data.gov.uk Newspaper

Friday, October 17, 2008

Over the last three days we’ve been working on a slide project. A design exercise if you like.

We’ve been thinking about the beta Data.gov.uk repository, and wanted to explore putting some of the information contained within into people’s hands in a form that is accessible, timely, and relevant.

And perhaps unsurprisingly, we thought a good way to do that was with a newspaper. So here it is, the Postcode Paper:

It’s a prototype of a service for people moving into a new area. In our exercise we imagined you might receive it after paying your council tax for the first time.

It gathers information about your area, such as local services, environmental information and crime statistics.
Conclusions

- The Semantic Web is, primarily a *Web of Data*
- Think of the out there when looking at 3D Media
  - use data out there
  - link your data to the rest of the Web of Data
- Making these available opens up nice application facilities!
Thank you for your attention!

- These slides are publicly available on:

  http://www.w3.org/2010/Talks/0211-Sophia-IH/