Practical experiences with developing Linked Data applications in the MELODIES project

Jon Blower
University of Reading

on behalf of MELODIES consortium
“If ... the Web made all the online documents look like one huge book, [Linked Data] will make all the data in the world look like one huge database.”
Linked Data: high-level viewpoint

• Linked Data builds bridges between data silos
• Based on wide agreement across communities
• Uses Web standards, not those of a particular community
Linked Data: technical viewpoint

http://5stardata.info

1. make your stuff available on the Web (whatever format) under an open license
2. make it available as structured data (e.g., Excel instead of image scan of a table)
3. use non-proprietary formats (e.g., CSV instead of Excel)
4. use URIs to denote things, so that people can point at your stuff
5. link your data to other data to provide context
What does Linked Data mean for users?
MELODIES Project concept
# The MELODIES Services

<table>
<thead>
<tr>
<th>Improving Emission Inventories</th>
<th>Site-specific information for land management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban Ecosystem Accounting</td>
<td>Assessment of Good Environmental Status for oceans</td>
</tr>
<tr>
<td>Desertification indicators</td>
<td>Marine transport</td>
</tr>
<tr>
<td>Customised crisis, disaster and risk mapping</td>
<td>Groundwater modelling</td>
</tr>
</tbody>
</table>
Common themes

• Integration of data from multiple sources
• Cross-querying to generate new information
Example MELODIES service: Urban Explorer application

Urban Explorer Land Use / Land Cover theme
See http://www.melodiesproject.eu/node/33
Open Data inputs to Urban Explorer
(Sources: OpenStreetMap, Prague authorities, LinkedGeoData.org, more)
European Historical Flood Database

**Input:**
- EFAS, EMS, HRL Layers, Local Component, EU-DEM, CLC
- Sentinel-1, EarthExplorer, Archives (Envisat, ERS, ...)

**Disaster events:**
- EM-DAT, MunichRe Nathan, Darmouth Flood Observatory, National databases

**Processing:**
- Public Linked Data
- Private Linked Data
- Public Linked Open Data

**Links:**
- www.terranea.de
- www.eoxplore.com
Common themes

• Integration of data from multiple sources
• Cross-querying to generate new information
• Handling geospatial Linked Data
Geospatial Linked Data tools

- **Strabon**
  - spatiotemporal RDF store
  - GeoSPARQL / stSPARQL support

- **GeoTriples**
  - Convert geospatial data to RDF

- **Ontop-spatial**
  - “Wrap” existing geo-databases

- **SILK**
  - Discover links in datasets
Sextant: visualizing and integrating time-evolving Linked Data
Common themes

• Integration of data from multiple sources
• Cross-querying to generate new information
• Handling geospatial Linked Data
• Handling raster/gridded data
  – Satellite images, weather forecasts, ...
The CoverageJSON data format
https://covjson.org

• Rich and efficient JSON encoding of geospatial scientific data
  – n-D Gridded and non-gridded data
  – continuous and categorical data
• Designed for web and mobile applications

• Bridge between science and Linked Data worlds
  – Metadata in JSON-LD (i.e. RDF in disguise)
  – Data in JSON arrays for efficiency
  – Can be used with no knowledge of RDF
Land cover remapping application
youtu.be/dxfmTkBdn90

- Data processed in-browser
- Metadata is RDF “in disguise” using JSON-LD
Common themes

• Integration of data from multiple sources
• Cross-querying to generate new information
• Handling geospatial Linked Data
• Handling raster/gridded data
  – Satellite images, weather forecasts, ...
• Selecting appropriate vocabularies
• Publishing data to ensure maximum re-use
MELODIES CKAN portal

Data discovery / publishing using GeoDCAT-AP vocabulary
Science Datasets

The web contains specialized repositories for datasets in many scientific domains: life sciences, earth sciences, material sciences, and more. Similarly, many governments maintain repositories of civic and government data. However, much of that structured data is not readily available to search engines, which must extract the data from HTML pages in order to provide search services to users. When webmasters provide structured markup, they enable search engines to “understand” this metadata, which in turn improves data discovery, leading scientists to the information they need for their work.

For example, consider this dataset that describes historical snow levels in the Northern Hemisphere. This page contains basic information about the data, like spatial coverage and units. Other pages on the site contain additional metadata: who produces the dataset, how to download it, and the license for using the data. With structured data markup, these pages can be more easily discovered by other scientists searching for climate data in that subject area.

Dataset markup is available for you to experiment with before it’s released to general availability. When you implement the markup, you’ll see previews in the Structured Data Testing Tool. You won’t, however, see your datasets appear in Search.

Let Google know that you added structured markup to your dataset by providing your details in our partner interest form.

What qualifies as a dataset?

For purposes of inclusion, we take a broad view of what qualifies as a dataset:
Summary

• Exploring Open Linked Data through eight real applications, with emphasis on user’s point of view

• We are developing tools and standards for producing and using *geospatial* linked data

• Application developers see benefits of Linked Data flexibility and ease of data integration
  – but data conversion is time-consuming
  – RDF is very unfamiliar to most
  – selecting vocabularies is a major barrier

• Need to build bridges to user communities to encourage adoption
  – Flexibility is good for data providers, very difficult for users
  – Consider other ways to represent links and semantics (e.g. JSON and JSON-LD)
  – Take opportunities to integrate with mass-market search
Thank you!

j.d.blower@reading.ac.uk
@Jon_Blower

http://melodiesproject.eu
@MelodiesProject