Linked Open Data
Overview

Theoretical Part
• Linked Data
• Status Quo
• RDF Data Cubes

Practical Part
Open Data → Linked Open Data
(using LODRefine / OpenRefine)
Linked Data
Sir Tim Berners-Lee
What is Linked Data?

“The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.”

— Tim Berners-Lee

Source: w3.org/DesignIssues/LinkedData.html
The 4 Rules

1. Use **URIs** as names for things.

2. Use **HTTP URIs** so that people can look up those names.

3. When someone looks up a URI, **provide useful information**, using the standards (RDF, SPARQL).

4. Include **links to other URIs**, so that they can discover more things.

Source: w3.org/DesignIssues/LinkedData.html
The 5 Star Scheme

* Available **on the web** (whatever format) with an **open license**, to be Open Data

**** As above, plus: Available as **machine-readable structured data** (e.g. excel instead of image scan of a table)

*** All the above, plus: Use **non-proprietary format** (e.g. CSV instead of Excel)

**** All the above, plus: Use **open standards** from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

***** All the above, plus: **Link your data** to other people’s data to provide context

Source: w3.org/DesignIssues/LinkedData.html
The Linking Open Data Project

“The goal of the W3C SWEO Linking Open Data community project is to extend the Web with a data commons by publishing various open data sets as RDF on the Web and by setting RDF links between data items from different data sources.”

Source: w3.org/.../CommunityProjects/LinkingOpenData
The Linked Open Data Cloud

Source: http://richard.cyganiak.de/2007/10/lod/
Finding Linked Data

The Data Hub / LOD

Sindice
Linked Data Vocabularies

schema.org

Linked Open Vocabularies
Serving Linked Data

**Intended Use?**

- HTML files with embedded data
- RDF files / dump
- SPARQL Endpoints

**Data Storage?**

- Wrapped Relational DBs
- Triple Stores

Source: w3.org/.../CommunityProjects/LinkingOpenData
Status Quo
The Linked Open Data Cloud (in Theory)

Source: http://richard.cyganiak.de/2007/10/lod/
The Linked Open Data Cloud (in Praxis)
SPARQL Endpoints Status

### Availability

#### Performance
- 10.000 is the most common result-size threshold
- Cold Tests: 1.13
- Warm Tests: 0.44
- Average ASK: 1.13
- Average JIN: 2.45
- Average JIN: 3.23

### Interoperability
- Solution Modifiers: 84.11%
- Common Operators and Filters: 76.16%
- Graph and other: 82.12%

### Discoverability
- 18.83% of the endpoints provide a Vcard description
- 12.33% of the endpoints provide a SD description
- 72.87% of the endpoints have no description

Server name:
RDF Data Cubes
Data Cube

Source: en.wikipedia.org/wiki/File:OLAP_slicing.png
RDF Data Cube

Vocabulary (and design pattern) for describing multi-dimensional data (such as statistics)

Based on …
- RDF
- SDMX (Statistical Data and Metadata eXchange)

Source: en.wikipedia.org/wiki/File:OLAP_slicing.png
Source: w3.org/TR/vocab-data-cube/
The Cube Model

Observations
contain the actual data (e.g. statistical values)

Dimensions
identify the observation (e.g. year, geographic region)

Measures
describe the phenomenon being observed (e.g. number of sales, average age)

Attributes
for any additional metadata (e.g. unit of measure)
Example RDF Data Cube

<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>2001</td>
<td>0.0833333333</td>
</tr>
<tr>
<td>Austria</td>
<td>2002</td>
<td>0.0833333333</td>
</tr>
<tr>
<td>Austria</td>
<td>2003</td>
<td>0.5681816162</td>
</tr>
<tr>
<td>Austria</td>
<td>2004</td>
<td>0.6666666667</td>
</tr>
<tr>
<td>Austria</td>
<td>2006</td>
<td>0.7</td>
</tr>
<tr>
<td>Austria</td>
<td>2007</td>
<td>1</td>
</tr>
<tr>
<td>Austria</td>
<td>2009</td>
<td>1</td>
</tr>
<tr>
<td>Austria</td>
<td>2010</td>
<td>1</td>
</tr>
<tr>
<td>Belgium</td>
<td>2001</td>
<td>0</td>
</tr>
<tr>
<td>Belgium</td>
<td>2002</td>
<td>0.0833333333</td>
</tr>
</tbody>
</table>

Displaying 10 of 199 results
Open Data → Linked Open Data
Today’s Goals

1. Import CSV file into LODRefine (OpenRefine)

2. Turn entity strings into URIs

3. Export as Linked Data

4. Upload to SPARQL Endpoint, explore in CODE Query Wizard, visualize in CODE Vis Wizard
1) Import CSV file into LODRefine

LODRefine

code.zemanta.com/sparkica

Touristen in Graz

http://data.graz.gv.at/daten/package/touristen-in-graz
2) Turn entity strings into URIs

**German DBpedia:** [http://de.dbpedia.org/](http://de.dbpedia.org/)
(e.g. [http://de.dbpedia.org/resource/Österreich](http://de.dbpedia.org/resource/Österreich))

(e.g. [http://reference.data.gov.uk/id/month/2012-01](http://reference.data.gov.uk/id/month/2012-01))

**OpenRefine – Pad with leading zeros:** [https://github.com/OpenRefine/OpenRefine/wiki/Recipes#pad-with-leading-zeroes](https://github.com/OpenRefine/OpenRefine/wiki/Recipes#pad-with-leading-zeroes)

**OpenRefine – Replace:** [https://github.com/OpenRefine/OpenRefine/wiki/Recipes#replacing-chars-punctuation-etc-using-regular-expressions](https://github.com/OpenRefine/OpenRefine/wiki/Recipes#replacing-chars-punctuation-etc-using-regular-expressions)
3) Export as Linked Data

**URIs for Observations**

http://kti.tugraz.at/semtech2013/observation/{YEAR}/{ID}

**Property “Label”**

http://www.w3.org/2000/01/rdf-schema#label

“Gäste in Graz aus {Herkunft} im Zeitraum {Jahr}-{Monat}”

**Property “Timespan”**

http://kti.tugraz.at/semtech2013/vocab/timespan

**Property “Region of Origin”**

http://kti.tugraz.at/semtech2013/vocab/regionOfOrigin

**Property “Arrivals”**

http://kti.tugraz.at/semtech2013/vocab/arrivals

**Property “Overnight Stays”**

http://kti.tugraz.at/semtech2013/vocab/overnightStays
4) Upload, Explore & Visualize

**Upload**

http://code.know-center.tugraz.at:8890/conductor

**SPARQL Endpoint**

http://code.know-center.tugraz.at:8890/sparql

**Explore & Visualize**

http://code.know-center.tugraz.at

Select “Semantic Technologies 2013”
Recommended Literature

**Linked Data:**
*Evolving the Web into a Global Data Space*

Read it for free at [linkeddatabook.com](http://linkeddatabook.com)
Thanks :)