Building Linked Data For Both Humans and Machines

Wolfgang Halb, Yves Raimond and Michael Hausenblas
LDOW2008, 2008-04-22, Beijing, China
Agenda

- **riese**
  - RDFizing and Interlinking the Eurostat data Set Effort
- LinkingOpenData
- Eurostat
- Architecture
- Schema & Data
- Querying
- Interlinking
- Demo
- Inside
RDFizing and Interlinking the Eurostat data Set Effort

http://riese.joanneum.at

- Contributing to Linking Open Data project
- Offer Semantic Web version of Eurostat data
- ... for both humans and machines
LinkingOpenData: Current State
Eurostat

- Eurostat (http://ec.europa.eu/eurostat)
- Publishes statistics in these themes:
  - General and regional statistics
  - Economy and finance
  - Population and social conditions
  - Industry, trade and services
  - Agriculture and fisheries
  - External trade
  - Transport
  - Environment and energy
  - Science and technology
- About the European Union in detail and additional statistics for some major non-European countries
Eurostat

- data dump provided as download (tables)
- updated twice a day
- additionally needed:
  - dictionary files to translate the data codes used
  - table of contents for structure

- Size of Eurostat data
  - 5 GB data dump in approx. 4,000 files
  - 350 million data values
  - 80,000 different data codes
riese: architecture

Eurostat
http://europa.eu/estatref/download/everybody/

TOC
[toc]
dictionaries
[tav]
datasets
[tav]

RDFising & Interlinking
[SWI Prolog, SesSQL server]

Templates
[XML/PHP, SPARQL]

Data & Schema
[RDF/XML]

Rendering & Serving
[Apache 2.2, PHP 5]

open linked datasets

human users
[XHTML]

Semantic Web agents
[RDF]

mixed port
indexer port
global query port
riece: schema & data
Survey Modelling Statistics in RDF, Lee Feigenbaum
- “The riese approach seems the best combination of flexibility and usability.” [1]

Generalised framework for statistics modelling to be released soon

Example data:

```xml
<riese:Dataset
    rdf:about="http://riese.joanneum.at/data/eb040"
    dc:title="Inflation rate"
    riese:data_end="2006"
    riese:data_start="1980"
    riese:last_update="08/01/2008"/>
```

<riese:Item dc:title="Inflation rate Austria 2006"
    rdf:value="1.7"
    <riese:dimension rdf:resource="http://riese.joanneum.at/dimension/geo/at"/>
    <riese:dataset rdf:resource="http://riese.joanneum.at/data/eb040"/>
</riese:Item>
riese: schema & data

- XHTML + RDFa example:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN" "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en"
    xmlns:riese="http://riese.joanneum.at/schema/core#">
    ...
    <head>
        ...
    </head>
    <body about="http://riese.joanneum.at/data/economy/"
        instanceof="riese:Dataset">
        <span class="toc-entry"><a href="http://riese.joanneum.at/data/bop/" rel="skos:narrower"
                class="dim">Balance of payments - International transactions</a></span>
        Last update: <span property="dc:date" datatype="xsd:date">2008-01-09</span>
    </body>
</html>
```
riese: querying

```
1 SELECT *
2 WHERE
3 { ?item riese:dimension dim:geo_at.
5   ?dataset dc:title ?ds_title
6   FILTER regex(?ds_title, "food", i))
```
riese: interlinking

- **Generic approach**
  - Restrict source dataset to potential interlinking candidates
  - Look up identifying feature from source in target dataset
  - Restrict results by appropriate classifications or identifiers
  - Create the interlink

- **User Contributed Interlinking**
riese: demo
riese: inside

- Server:
  - Apache 2.2
  - SWI-Prolog
  - p2r
  - PHP 5
  - RDF/XML documents in the file system
    (to be replaced by p2r data-pump)

- Client
  - XHTML+RDFa
  - Javascript/Yahoo! User Interface Library [YUI]
Let’s discuss!