

The essentials for Semantic Web application development

Linked Data

Josef Petrák, petrakj@vse.cz

KIZI, VŠE Praha



Outline

- Introduction into RDF
- Diving into Linked Data
- Linking Open Data
 - Overview of the initiative, and Basics
 - Best Practices detailed

Introduction into RDF

About RDF

- RDF is a data model
- It is syntax independent
- A graph of triples
- Triple in the form ?subject ?predicate ?object
- All resources are URIs (vocabulary terms as well)
- Blank node is a local resource without global ID.
- Literal is a string value (may be typed or have a language)

- For more complete intro, see rdf:about website <http://www.rdfabout.com/> or W3C RDF Primer <http://www.w3.org/TR/rdf-primer/>.

Diving into Linked Data

Purpose of Linked Data

- Linked Data are „Best Practices“ for:
 - Ontology design and usage
 - Data and metadata management
 - Application design issues

The very first idea

- *„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.“*
- The four rules
 1. Use URIs as names of things.
 2. Use HTTP URIs so that people can look up those names.
 3. When someone looks up a URI, provide useful information.
 4. Include links to other URIs. so that they can discover more things.
- See *Tim Berners-Lee, Linked Data; in Design Issues, <http://www.w3.org/DesignIssues/LinkedData.html>*

Linking resources

1. In one source, using rdf:IDs

```
<rdf:Description about="#denise"  
  <fam:child rdf:Resource="#edwin">  
  <fam:child  
    rdf:Resource="http://smiths.name#carol">  
</rdf:Description>
```

2. Using HTTP 303 See Also header

3. Using RDF property „rdfs:seeAlso“

Linking Open Data

W₃C SWEO Community Project

Overview

- The initiative is active since June 2007.
- News, list of committers and their contributions, FAQ, information resources; see <http://esw.w3.org/topic/SweolG/TaskForces/CommunityProjects/LinkingOpenData>
- Tutorial about *How to publish Linked Data on the Web*, see <http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>

Basics of Linked Data

- Best Practices: avoid blank nodes, avoid reification, use RDF collections and containers carefully
- Content Negotiation, URIs dereferencing, „Cool URIs“
- RDF Links, URI Aliases
- Ontology / Schema reuse, best practices on term definitions
- „Must include“ information in RDF description of a resource
- Wrapping non-RDF databases or APIs, and opening the close-world silos

Best Practices Detailed

Dereferencing HTTP URIs

- Information resource: a resource with representation (page, document, image). The server should return HTTP 200 OK.
- Non-information resource: a resource without representation (vocabulary terms, human beings, etc.). The server should return HTTP 303 See Other, and Location of related information resource headers.
- Demo
 - Non-information resource <http://keg.vse.cz/resource/person/josef-petrak>
 - Information resource <http://keg.vse.cz/person/josef-petrak>

Content Negotiation

- Serving the best acceptable representation of information resource to the agent while dereferencing the HTTP URIs.
- Based on HTTP `Accept` header.
- *HTML for browsers, RDF for Semantic Web agents!*

URI Aliases

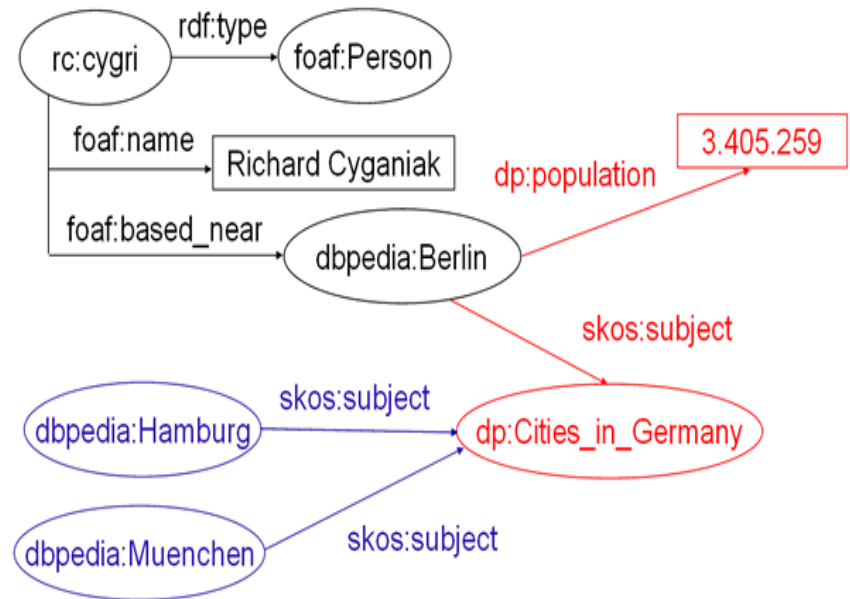
- The Semantic Web is open world, thus **one resource** may have **multiple identifiers**.
- To map identical resources, create aliases.
- Using OWL property `owl:sameAs`.
- <http://keg.vse.cz/resource/person/josef-petrak>
`owl:sameAs` <http://jspetrak.zapisky.info/>

„Cool URIs“ for the Semantic Web

- Guidelines on effective creating resource URIs.
- Applied on non-information resources.
- Two strategies – 303 URIs vs. hash URIs.
- For further information see <http://www.dfki.uni-kl.de/~sauermann/2006/11/cooluris/>

RDF Links

- *Foundation of Web of Data*
- RDF Links point to other resources with additional information.
- Used for navigation cross data, and crawling by search engines.
- Created using various vocabulary terms with resource as object.



„Bad features“ of RDF

- *Blank nodes*: make impossible to link external resource. **Prefer to use URI references!**
- *Reification*: unclear semantics, difficult to query. **Attach metadata instead!**
- *RDF collections and containers*: difficult to query. **Use multiple triples with the same predicate instead!**

Reusing vocabulary terms

- Prefer reuse instead of creating ontologies from scratch.
- Good for data understanding among agents.
- De-facto standard vocabularies: **FOAF** (for contacts and social networks), **SIOC** (on-line communities and discussions), **DOAP** (software project descriptions), **Dublin Core** (library metadata), **SKOS** (taxonomy description), **Creative Commons** (licencing)
- Of course, you can mix vocabularies together!
- For more complete list, see <http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/CommonVocabularies>

How to define vocabulary terms

1. **Do not define new vocabulary from scratch**, but combine existing ones, and add new terms.
 2. **Provide metadata for both machines and humans.** Use `rdfs:label` and `rdfs:comment` properties.
 3. **Make term URIs dereferencable.**
 4. **Make use of other people's terms.** Map terms relations using `rdfs:subClassOf` and `rdfs:subPropertyOf`.
 5. **State all important information explicitly.** Humans can do guesswork, machines can't.
 6. **Do not create over-constrained models.**
- See *Best Practice Recipes for Publishing RDF Vocabularies*, <http://www.w3.org/TR/swbp-vocab-pub/>

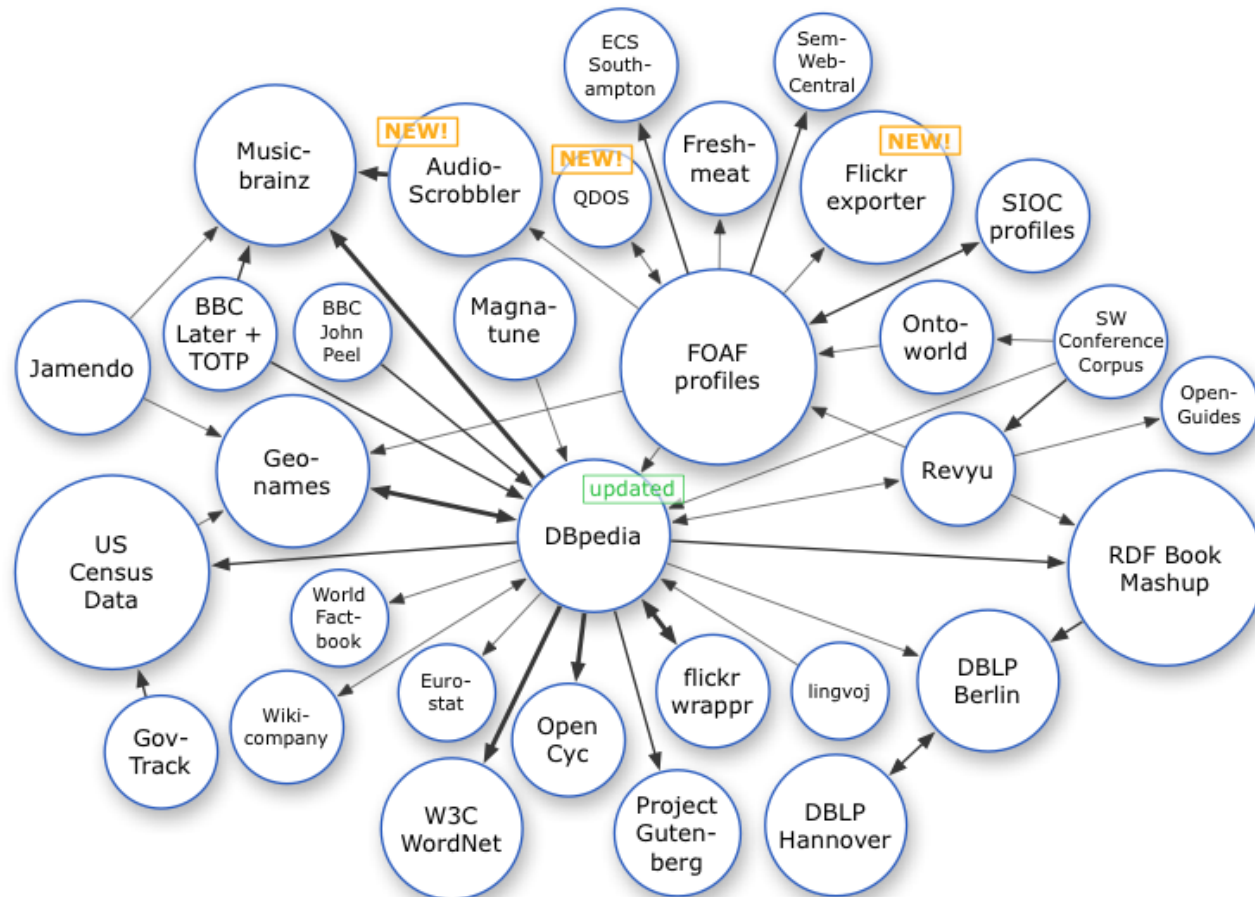
Don't forget in your RDF data

- *Description*: all triples about the resource (resource URI is in the subject).
- *Backlinks*: all triples about other resources which are in relation to described resource (resource URI is in the object).
- *Related description*: information about related resources. **Consider ideal data size!**
- *Metadata*: information about author, origin or licence of the data.
- *Syntax*: RDF is syntax-independent. Prefer commonly used syntaxes such as RDF/XML (official standard), and Turtle or N3 which are well human-readable.

Database content in RDF

- Various well-known information databases on the Web now publish (or are published by 3rd party projects) data as Linked Data in RDF.
 - Wikipedia as DBpedia, www.dbpedia.org
 - Geonames
 - CIA Fact Book, Eurostat Countries and Regions
 - MusicBrainz
- You can reuse and link data from these databases.

Map of Linked Data sources



Wrapping APIs to produce RDF

- Several useful information is served by APIs.
- The information can not crawled by search engines.
- APIs can not be accesses by generic data browser.

- Wrapper may
 - Assign URIs to non-information resources
 - Do the URI dereference and server RDF.

- Example applications: Amazon and Googe Base APIs wrapped by RDF Book Mashup, WordPress with SIOC Exporter plug-in

Small demo

- Let's browse the interlinked Semantic Web!

„Using Tabulator <http://www.w3.org/2005/ajar/tab> browse Timbl's profile
<http://www.w3.org/People/Berners-Lee/card>“

Questions?

Thank you.
