The Web of Data
The Web of Data
The Web of Data
The Web of Data
The Web of Data

Enabling networked knowledge.
LOD Hubs = datasets that attract many inlinks

Music metadata community uses BBC Music identifiers

UK government data community uses Ordnance Survey identifiers

Library data community uses Library of Congress Subject Headings
Standard identifiers
Standard identifiers

LOD hubs are bar codes for a specific community.

0 123456 789012
For example, government data
For example, government data

- 5000 datasets about:
  - Politicians
  - Companies
  - Schools
  - Administrative areas
  - Motorways
  - Government departments
How to Attract Links
How to Attract Links
# Reconciliation

<table>
<thead>
<tr>
<th>City</th>
<th>State</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambridge</td>
<td>Massachusetts</td>
<td>United States</td>
</tr>
</tbody>
</table>
Reconciliation

DBpedia

UNIVERSITY OF CAMBRIDGE

Cambridge Bay in Canada

City | State | Country
--- | --- | ---
Cambridge | Massachusetts | United States
Reconciliation

Cambridge city in Maryland

Cambridge city in Canada

<table>
<thead>
<tr>
<th>City</th>
<th>State</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambridge</td>
<td>Massachusetts</td>
<td>United States</td>
</tr>
</tbody>
</table>
Reconciliation

DBpedia

label=Cambridge
type = City

In the state of Massachusetts

Cambridge city in Massachusetts

<table>
<thead>
<tr>
<th>City</th>
<th>State</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambridge</td>
<td>Massachusetts</td>
<td>United States</td>
</tr>
</tbody>
</table>
Approaches

- SPARQL
- SPARQL + full-text search
- Silk Server
- Semantic Web search engines
- Based on regular expressions

**Pros**
- Standardised
- Zero-effort approach

**Cons**
- Slow
- Not good at text search
- No ranked results
SPARQL + full-text search

- Based on full-text extension for SPARQL

**Pros**
- More forgiving string matching
- Ranking
- Zero-effort (depending on your SPARQL store)

**Cons**
- Proprietary syntax
Silk Server

- **Pros**
  - Powerful declarative link specification
  - Variety of similarity functions

- **Cons**
  - Configuration needs to be prepared
  - Silk Server needs to be deployed
  - Silk Server tightly couples its input and reference data
Semantic Web Search Engine

- Based on Sindice API
- Pros
  - Zero-effort approach (if your dataset is indexed in Sindice)
  - Search distributed RDF datasets (e.g. FOAF profiles)
- Cons
  - Noisy
- Data Interlinking benchmark (part of IM@OAEI2010)
- We reconciled DailyMed against:
  - DBpedia SPARQL endpoint (http://dbpedia.org/sparql)
  - Sider dump file (part of the benchmark)
Results

- SPARQL with REGEX is unsuitable (performance)
- Except if labels are very consistent
- Type restrictions are very effective
- Silk has best recall (but requires custom link spec)
Google Refine + RDF
List of people from DERI

- Fadi Maali
- Gofran Shukair
- Souleiman Hasan
- Richard Cyganiak
- Michael Hausenblas
- Manfred Hauswirth
- Stefan Decker
- Lukasz Porwol
- Alexandre Passant
- Owen
- Maciej Dabrowski
RDF Schema Alignment

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

**Base URI:** http://localhost:3333/edit

**Available Prefixes:**
- rdfs
- dcat
- foaf
- spat
- dct
- wns
- rdf
- skos

```
http://deri.ie/deri
<table>
<thead>
<tr>
<th>foaf:Organization</th>
<th>add rdf:type</th>
</tr>
</thead>
<tbody>
<tr>
<td>foaf:member</td>
<td></td>
</tr>
<tr>
<td>full name</td>
<td></td>
</tr>
<tr>
<td>URI</td>
<td></td>
</tr>
<tr>
<td>foaf:Person</td>
<td></td>
</tr>
<tr>
<td>add rdf:type</td>
<td></td>
</tr>
<tr>
<td>foaf:name</td>
<td></td>
</tr>
<tr>
<td>full name cell</td>
<td></td>
</tr>
<tr>
<td>foaf:depiction</td>
<td></td>
</tr>
<tr>
<td>image URI</td>
<td></td>
</tr>
<tr>
<td>add rdf:type</td>
<td></td>
</tr>
<tr>
<td>add property</td>
<td></td>
</tr>
</tbody>
</table>
```
## Preview URI values

### Expression

```plaintext
value.urlify()
```

### Table:

<table>
<thead>
<tr>
<th>row</th>
<th>value</th>
<th>value.urlify()</th>
<th>resolved against the base URI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fadi Maali</td>
<td>fadi-maali</td>
<td><a href="http://localhost:3333/fadi-maali">http://localhost:3333/fadi-maali</a></td>
</tr>
<tr>
<td>2</td>
<td>Gofran Shukair</td>
<td>gofran-shukair</td>
<td><a href="http://localhost:3333/gofran-shukair">http://localhost:3333/gofran-shukair</a></td>
</tr>
<tr>
<td>3</td>
<td>Souleiman Hasan</td>
<td>souleiman-hasan</td>
<td><a href="http://localhost:3333/souleiman-hasan">http://localhost:3333/souleiman-hasan</a></td>
</tr>
<tr>
<td>4</td>
<td>Richard Cyganiak</td>
<td>richard-cyganiak</td>
<td><a href="http://localhost:3333/richard-cyganiak">http://localhost:3333/richard-cyganiak</a></td>
</tr>
<tr>
<td>5</td>
<td>Michael Hausenblas</td>
<td>michael-hausenblas</td>
<td><a href="http://localhost:3333/michael-hausenblas">http://localhost:3333/michael-hausenblas</a></td>
</tr>
<tr>
<td>6</td>
<td>Manfred Hauswirth</td>
<td>manfred-hauswirth</td>
<td><a href="http://localhost:3333/manfred-hauswirth">http://localhost:3333/manfred-hauswirth</a></td>
</tr>
</tbody>
</table>
Example

List of people from DERI

Fadi Maali
Gofran Shukair
Souleiman Hasan
Richard Cyganiak
Michael Hausenblas
Manfred Hauswirth
Stefan Decker
Lukasz Porwol
Alexandre Passant
Owen
Maciej Dabrowski

Find related RDF datasets

Related RDF datasets

- List of domains:
  - dblp.l3s.de
  - www.mendeley.com
  - www.deri.ie
  - data.semanticweb.org
  - twitter.com
  - demo.semantic-web.at
  - www.semanlink.net
  - www.aifb.kit.edu
  - semanticweb.org
  - www.linkedin.com
  - richard.cyganiak.de
  - www.slideshare.net

Add domain-specific Sindice service

Enabling networked knowledge.
Example
Example

Reconciliation result facets

Resource Preview
Reconcile against a SPARQL endpoint

Add SPARQL-based reconciliation service

Name: DBpedia
A human readable name

Endpoint details
Endpoint URL: http://dbpedia.org/sparql

Graph URI:
Leave empty to use the default graph

Type: Virtuoso
This determines the syntax that will be used for search

Label properties
Select properties that are used to label resources in the endpoint. These properties will be used to match resources:

- rdfs:label
- skos:prefLabel
- dcterms:title
- dc:title
- foaf:name
- Other...

OK Cancel
Reconcile against an RDF dump

Add file-based reconciliation service

This will set up a new reconciliation service based on an RDF file that provides entity URIs and entity labels.

Name: NYTimes Places
A human readable name

File details
- Load file from URL:
- Upload file: Choose File No file chosen
- File format: Auto-detect

Label properties
Select properties that are used to label resources in the endpoint. These properties will be used to match resources:
- rdfs:label
- skos:prefLabel
- dcterms:title
- dc:title
- foaf:name
- Other...

OK Cancel
5-star plan for open data

★ Make your stuff available on the Web

★★ Make it available as structured data
(e.g., an Excel sheet instead of image scan of a table)

★★★ Use a non-proprietary format
(e.g., a CSV file instead of an Excel sheet)

★★★★ Use linked data format
(i.e., URIs to identify things, and RDF to represent data)

★★★★★ Link your data to other people’s data to provide context
Making this easier!

New Dataset → Hub
- RDF Extension for Google Refine
- Reconciliation will be in the upcoming next version

Thanks!