Linked Data

Making things more accessible

Hugh Glaser

Data-Intensive Research: how should we improve our ability to use data?
e-Science Institute, Edinburgh
17th March 2010
Overview

- Technology -> Use
- Linked Data tutorial
- Doing the linking
- Using the linking
Motivation

• **Publishing data becomes an imperative**
  – Freedom of Information
  – Public confidence
  – Reputation

• **Linking Data becomes an imperative**
  – Research benefits from the use of multiple datasets
  – Research benefits from datasets from different researchers
  – Connecting publications with the subject data

• **Data Fusion**
  – Breaking silos
  – Socio-political problem
  – data.gov.uk
How to Join?

- Hypertext Style/ Cool URIs don't change.  
  - [http://www.w3.org/Provider/Style/URI](http://www.w3.org/Provider/Style/URI)

- Cool URIs for the Semantic Web  
  - [http://www.w3.org/TR/cooluris/](http://www.w3.org/TR/cooluris/)

- Linked Data - Design Issues  
  - [http://www.w3.org/DesignIssues/LinkedData.html](http://www.w3.org/DesignIssues/LinkedData.html)

- How to publish Linked Data on the Web  
  - [http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/](http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/)

- Linked Data  
  - [http://linkeddata.org/](http://linkeddata.org/)
1. Use URIs as names for things

- **Duh!**

- **Everything**
  - If you don't name something you can't talk about it
  - Things of course
  - Workflows
  - Ideas
  - ...

- **Cool URIs**
  - Think of the consumer/customer
    - [https://secure.ecs.soton.ac.uk/gizmos/person_by_username.php?username=hg](https://secure.ecs.soton.ac.uk/gizmos/person_by_username.php?username=hg)
    - [https://secure.ecs.soton.ac.uk/person/username/hg](https://secure.ecs.soton.ac.uk/person/username/hg)
  - Library of Congress - *Chronicling America*: Historic American Newspapers
    - [http://chroniclingamerica.loc.gov/](http://chroniclingamerica.loc.gov/)
    - Cool URIs – 71233 page views per week to 447992 - factor of 6.3

- **Ambiguity**
  - URIs help to avoid it, especially if you...
2. Use HTTP URIs so that people can look up those names

- HTTP URI names come with an excellent look up mechanism
- And ownership, etc.
- Make sure they are Cool
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)

- Otherwise I don’t know what you mean
- Deliver some human readable data
  - html
- Deliver some machine processable data
  - RDF
  - JSON
  - CSV
  - text
- Query is good
  - SPARQL
RDF - Triples

- “Ian Millard works for the University of Southampton”

  person-04860 name Ian Millard
  person-04860 works for Southampton University
  Southampton University label The University of Southampton
RDF - Triples

• “Ian Millard works for the University of Southampton”

  soton-rkb:person-04860 akt:works_for dbpedia:Southampton_University .
  dbpedia:Southampton_University rdfs:label “The University of Southampton” .

• This is actually RDF!
  – Given the namespaces

• I could put this in an RDF store now
**SPARQL**

```sparql
PREFIX soton-rkb: <http://southampton.rkbexplorer.com/id/>
PREFIX akt:  <http://www.aktors.org/ontology/portal#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT * WHERE
```

<table>
<thead>
<tr>
<th>org</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://southampton.rkbexplorer.com/id/UoS">http://southampton.rkbexplorer.com/id/UoS</a></td>
<td>University of Southampton</td>
</tr>
<tr>
<td><a href="http://southampton.rkbexplorer.com/id/org-ecs">http://southampton.rkbexplorer.com/id/org-ecs</a></td>
<td>School of Electronics and Computer Science, University of Southampton</td>
</tr>
</tbody>
</table>
4. Include links to other URIs. so that they can discover more things

- "Foreign" URIs
  - dbpedia:Southampton_University

- Equivalence
  - owl:sameAs
  - skos:exactMatch
  - ...

© 2010 Seme4
Seme4.com
Where to keep the equivalence?

- **Equivalence is different to the substantive data**
  - Accuracy
  - Provenance
  - License
  - Trust
  - Maintenance
  - Process

- **Separate knowledge base**

- **Too many identifiers?**
  - The answer is not to create another one

- **A little middleware**
  - sameAs aware
Discovery v. Description

- **Equivalence**

- **http://sameas.org/ is discovery**

- **For description you need your own**

- **Many advantages to distributed architecture**
  - Scalable
  - Multiple
  - Rollback
  - Deprecation
  - Canonisation (URI reduction)
<sameAs>
interlinking the Web of Data

The Web of Data has many equivalent URLs. This service helps you to find co-references between different data sets. Enter a known URI, or use Sindice to search first.

Why not try searching for the string "London", or finding other equivalent identifiers for http://dbpedia.org/resource/London?

Currently serving 30618791 URLs in 10796012 bundles!

about • contact
The Web of Data has many equivalent URLs. This service helps you to find co-references between different data sets. Enter a known URI, or use Sindice to search first.

Equivalent URLs for http://dbpedia.org/resource/London –
9. http://www.okkam.org/ens/d1d7d3fb0-4418-41a1-8866-89c1458a1c66
10. http://www.okkam.org/entity/oke0ab6619-e28a-438d-b53b-cdd5e8157b30
12. http://data.ordnancesurvey.co.uk/id/7000000000041428
15. http://rdf.freebase.com/ns/guid.9202a8c04000641f80000000000242b2
18. http://sw.opencyc.org/concept/Mx4rvVjWPJwpEbGdrcN5Y29ycA

rdf+xml  n3  json  text  show fewer items
<sameAs>
interlinking the Web of Data

The Web of Data has many equivalent URLs. This service helps you to find co-references between different data sets. Enter a known URI, or use Sindice to search first.

<sameAs> http://webscience.org/person/23

Sindice
Enter a literal (text) search...

Equivalent URIs for http://webscience.org/person/23 –

1. http://southampton.rkbexplorer.com/id/person-1ec5be2388d15446d69e61...d7ff327d1e
2. http://southampton.rkbexplorer.com/id/person-2a8f4c2138ec4ce328c712...fd42c1a2tb
3. http://southampton.rkbexplorer.com/id/person-2a8f4c2138ec4ce328c712...34d38c7878


rdf+xml · n3 · json · text

Currently serving 30918791 URLs in 10796012 bundles!

about · contact.
<sameAs>
interlinking the Web of Data

The Web of Data has many equivalent URLs.
This service helps you to find co-references
between different data sets.
Enter a known URI, or use Sindice to search first.

Equivalent URLs for http://dbpedia.org/resource/14-Methoxymetopon –
<sameAs>
3. http://rdf.freebase.com/ns/guid.9202a8c04000841f800000000812d3e8

Currently serving 30918791 URLs in 10796012 bundles!

about · contact
(C) Sindice
Data Fusion

- Break the silos
- At low risk
- And low cost
- With little or no process/workflow change
Using Linked Data and sameAs: RKBExplorer.com

- Unified view of disparate resources
- Open Web
- Distributed query
- Distributed sameAs
- Services delivering RDF, JSON, csv, ...
  - http://www.rkbexplorer.com/service/
- Different interfaces possible
Brian Randell is a British computer scientist, specializing in research in software fault tolerance and dependability.
Brian Randell is connected to Andrea Bondavalli

They are linked by 4 relations.

**Publications**

They have co-authored 3 papers:

- Dependability Analysis of Iterative Fault Tolerant Software Considering Correlation
- Rational Design of Multiple-Redundant Systems: Adjudication and Fault Treatment
- Dynamic Adjustment of Dependability and Efficiency in Fault-Tolerant Software

**Projects**

They are both members of the ReSIST Resilience for Survivability in IST project.
MyExperiment: dotAC Explorer

Location of Project Members

Details

Names:
MyExperiment

Starts:
01/03/2007

Ends:
31/03/2009

Web Address:
http://www.myexperiment.org/id/project-279

Description:
The MyExperiment Virtual Research Environment enables you to do your research.

Participating Organisations

University of Southampton

Project Members

Cagio Gomba
Darius Michaelides
David De Floure
Frederique VanTil

Related JISC Projects

Meeting Memory Technology: Informing Collaboration
MyExperiment Repository Enhancement
Collaborative Orthopaedic Research Environment
Internet-Based Behavioural Research
Virtual Research Integration Collaboration
Building Research and Innovation Networks

© 2010 Seme4

Seme4.com
Concluding Remarks

- Linked Data is not hard
  - URIs are easy
  - RDF is just another format

- Coreference is important

- Ian Millard did the work

- http://eprints.ecs.soton.ac.uk/18691/