An Introduction to the Potential of Linked Data -
RKBExplorer, ResearchSpace and sameAs.org

Kultivate Project, Linked Data Workshop
HEFCE, Centre Point, 12th December 2011

Hugh Glaser
Seme4 Limited and University of Southampton, UK
Linked Data – some motivation

Linked data is data in which real-world things are given addresses on the web (URIs), and data is published about them in machine-readable formats at those locations. Other datasets can then point to those things using their URIs, which means that people using the data can find out more about something without that information being copied into the original dataset.

This page lists the sectors for which we currently publish linked data and some additional resources that will help you to use it. Most sectors have one or more SPARQL endpoints, which enable you to perform searches across the data; you can access these interactively on this site.

Reference data covers the central workings of government, including organisational structures where these have been made available as RDF.

Browse

- Visualisation
- Government Departments
- Other Public Bodies
- Ministers
- Members of Parliament
- Members of the House of Lords

Search

- SPARQL Endpoint provided by Talis
- Search provided by Talis
- SPARQL Endpoint provided by TSO

Ordnance Survey

Ordnance Survey have released a number of their products as linked data, including postcode units and administrative areas.

Browse

- Ordnance Survey Linked Data
data.nytimes.com
For the last 150 years, The New York Times has maintained one of the most authoritative news vocabularies ever developed. In 2009, we began to publish this vocabulary as linked open data.

The Data
As of 13 January 2010, The New York Times has published approximately 10,000 subject headings as linked open data under a CC BY license. We provide both RDF documents and a human-friendly HTML versions. The table below gives a breakdown of the various tag types and mapping strategies on data.nytimes.com.

<table>
<thead>
<tr>
<th>Type</th>
<th>Manually Mapped Tags</th>
<th>Automatically Mapped Tags</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>People</td>
<td>4,978</td>
<td>0</td>
<td>4,978</td>
</tr>
<tr>
<td>Organizations</td>
<td>1,489</td>
<td>1,592</td>
<td>3,081</td>
</tr>
<tr>
<td>Locations</td>
<td>1,910</td>
<td>0</td>
<td>1,910</td>
</tr>
</tbody>
</table>

9,960

Browse individual data records:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

SKOS Files
Download all of the data records as SKOS Files.

People
Organizations
Locations

Using Our Linked Data
Want to learn more about the nuts and bolts of our RDF documents? This page provides technical documentation. This blog post provides step-by-step instructions for building your own NYT Linked Data Application.
### Linked Data – a little motivation

Ordinance Survey is Great Britain's national mapping agency, providing the most accurate and up-to-date geographic data, relied on by government, business and individuals. OS OpenData is the opening up of Ordinance Survey data as part of the drive to increase innovation and support the "Making Public Data Public" initiative. As part of this initiative Ordinance Survey has published a number of its products as Linked Data. Linked Data is a growing part of the Web where data is published on the Web and then linked to other published data in much the same way that web pages are interlinked using hypertext. The term Linked Data is used to describe a method of exposing, sharing, and connecting data via URLs on the Web. To find more Linked Data published as part of this initiative please go to data.gov.uk. If you are not familiar with Linked Data, OS OpenData products are also available in alternative formats from the OS OpenData website. Ordinance Survey can provide support for the Ordinance Survey OpenData products, but cannot give advice or support on using RDF, SPARQL or SPARQL Endpoints.

Ordinance Survey has published three separate linked data resources: the 1:50 000 Scale Gazetteer, Code-Point Open and the administrative geography gazetteer for Great Britain.

<table>
<thead>
<tr>
<th>Title</th>
<th>Ordnance Survey Linked Data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Ordnance Survey is Great Britain's national mapping agency, providing the most accurate and up-to-date geographic data, relied on by government, business and individuals. OS OpenData is the opening up of Ordinance Survey data as part of the drive to increase innovation and support the &quot;Making Public Data Public&quot; initiative. As part of this initiative Ordinance Survey has published a number of its products as Linked Data. Linked Data is a growing part of the Web where data is published on the Web and then linked to other published data in much the same way that web pages are interlinked using hypertext. The term Linked Data is used to describe a method of exposing, sharing, and connecting data via URLs on the Web. To find more Linked Data published as part of this initiative please go to data.gov.uk. If you are not familiar with Linked Data, OS OpenData products are also available in alternative formats from the OS OpenData website. Ordinance Survey can provide support for the Ordinance Survey OpenData products, but cannot give advice or support on using RDF, SPARQL or SPARQL Endpoints.</td>
</tr>
<tr>
<td>Creator</td>
<td>Ordinance Survey</td>
</tr>
<tr>
<td>Publisher</td>
<td>Ordinance Survey</td>
</tr>
<tr>
<td>Date issued</td>
<td>2010-10-25</td>
</tr>
<tr>
<td>Example resource</td>
<td>SO16 4GU Portwood Southampton Abstract Geometry Southampton Itchen The County of Hampshire The City of Southampton Southampton Common</td>
</tr>
<tr>
<td>SPARQL endpoint</td>
<td>Sparql</td>
</tr>
<tr>
<td>UIR lookup point</td>
<td>Data?about=</td>
</tr>
<tr>
<td>Vocabulary used</td>
<td>The administrative geography and civil voting area ontology Geometry Ontology FOAF OWL Gazetteer Ontology Postcode Ontology</td>
</tr>
<tr>
<td>UIR regex pattern</td>
<td><a href="http://data.ordnancesurvey.co.uk/fd/+">http://data.ordnancesurvey.co.uk/fd/+</a></td>
</tr>
<tr>
<td>See also</td>
<td>Semantic.html</td>
</tr>
<tr>
<td>Coverage</td>
<td>England Scotland Wales</td>
</tr>
</tbody>
</table>
Iron Maiden

Formed 1975.
English heavy metal band

Biography

Iron Maiden are an English heavy metal band from Leyton in East London, formed in 1975. The band is directed by founder, bassist and songwriter Steve Harris. Since their inception, the group has released a collective total of thirty albums: fourteen studio albums, seven live albums, four EPs and five compilations.

Pioneers of the New Wave of British Heavy Metal, with such an original sound they achieved success during the early 1980s and after several lineup changes, went on to release a series of platinum and gold albums. These include the US platinum-selling landmarks The Number of the

Now On The BBC

Bruce Dickinson's Friday Rock Show
Bruce plays brand new and classic metal with an array of top flight guests on 6 Music

Latest Tracks Played On The BBC

<table>
<thead>
<tr>
<th>Song</th>
<th>Radio</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run To The Hills</td>
<td>6 Music</td>
<td>06/04/2010</td>
</tr>
<tr>
<td>Run To The Hills (Live at Maida Vale)</td>
<td>BBC Radio 1</td>
<td>02/04/2010</td>
</tr>
<tr>
<td>Aces High</td>
<td>6 Music</td>
<td>27/02/2010</td>
</tr>
</tbody>
</table>
BBC World Cup 2010 dynamic semantic publishing

Post categories: World Cup, linked data, metadata, semantic, semantic web, web publishing
Jem Rayfield | 10:00 UK time, Monday, 12 July 2010

The World Cup 2010 website is a significant step change in the way that content is published. From first using the site, the most striking changes are the horizontal navigation and the larger, format high-quality video. As you navigate through the site it becomes apparent that this is a far deeper and richer use of content than can be achieved through traditional CMS-driven publishing solutions.

The site features 700-plus team, group and player pages, which are powered by a high-performance dynamic semantic publishing framework. This framework facilitates the publication of automated metadata-driven web pages that are light-touch, requiring minimal journalistic management, as they automatically aggregate and render links to relevant stories.
Information Management: A Proposal

Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control

http://www.w3.org/History/1989/proposal.html
“The ability of the semantic web to cheaply but effectively integrate data and breakdown data silos provides museums with a long awaited opportunity to present a richer, more informative and interesting picture.”

» Dominic Oldman, British Museum
The Rosetta Stone

Object type:
stone (all objects)

Title (object):
The Rosetta Stone

Materials:
granodiorite (all objects)

Place (findspot):
Excavated/Findspot: Fort Saint Julien (all objects)

Date:
196BC

Period/Culture:
Provenance (scope note | all objects)

Description:
Part of grey and pink granodiorite stele bearing priestly decree concerning Ptolemy V in three blocks of text: Hieroglyphic (14 lines), Demotic (32 lines) and Greek (53 lines).

Inscriptions:
Inscription Type: inscription
Inscription Language: Greek

Dimensions:
Length: 112.3 centimetres (max)
Width: 76.7 centimetres
Thickness: 28.4 centimetres

Condition:
Fair (incomplete)

Curator’s comments:
Compass text:
The Rosetta Stone

Object types:
- stele (all objects)

Title (object):
The Rosetta Stone

Materials:
- granite (all objects)
- Place (findspot):
  Excavated/Findspot: Port Saint Julien (all objects)

Date:
- 196BC

Period/Culture:
- Ptolemaic

Description:
Part of grey and pink granite stele bearing priestly decree concerning Ptolemy V in three blocks of text: Hieroglyphic (14 lines), Demotic (32 lines) and Greek (53 lines).

Inscriptions:
- Inscription Type: Inscription
- Inscription Language: Greek
- Inscription Type: Inscription
- Inscription Script: Hieroglyphic
- Inscription Language:

Inscription Comment: The inscription is a decree passed by a council of priests, one of a series of edicts the royal cult of the 15-year-old Ptolemy V on the first anniversary of his coronation.

Inscription Type: Inscription
- Inscription Script:demotic
- Inscription Language:

Dimensions:
- Length: 112.3 centimetres (max)
- Width: 78.7 centimetres
- Thickness: 26.4 centimetres

Department: Ancient Egypt & Sudan
- Registration number: 24
- BM/EP number: 24
- Additional IDs
- B5.24
- Bibliographic reference
Description
Part of grey and pink geminite stela bearing priestly decree concerning Pharaoh V in three blocks of text: Hieroglyphic (14 lines), Demotic (12 lines) and Greek (53 lines).

Inscriptions
Inscription Type: Inscription
Inscription Language: Greek

Dimensions
Length: 112.3 centimetres (max)
Width: 75.7 centimetres
Thickness: 28.4 centimetres

June 1999 – Temporary Exhibition

Reason for treatment
Temporary Exhibition

Treatment proposal
1. Test clean return edgels.
2. Clean and prepare for "Rosetta Stone, Cracking Codes and Decipherment" exhibition.

Condition
Object coated with protective treatment - identified as camphor wax, hand-grease and dirt.
Inscriptions bear traces of hard pink material, principally near edgels and at corners. It closely matches the pink-granite vein for colour and is covered in places by black penins ink - see annotated report. Inscription was painted with water-based white Plaka gouache, by Carol Andrews in 1982 to replace white powder. (White powder was probably modelling clay applied as a slip-over earlier surface; the excise would have been removed with a cloth). While powders (clay) continued to absorb areas of damage. Returns on either side have white deposit with preserved horizontal directionality, but areas of damage do not have this deposit. (This is possibly gypsum plaster, spilled during cleaning of the inscription in Egypt). Also found: traces of reddish compacted clay on back and returns and a looser white material under cross-member of old mount, but elsewhere consolidated and turned brown by camphor wax. Text painted in white on left return: "CAPTURED IN EGYPT BY THE BRITISH ARMY IN 1801" on right return: "PRESENTED BY KING GEORGE III."
British Museum

• That was a while ago
  – Science data
  – Conservation data
  – Linked to the Catalogue
  – Live on the web site

• Then
  – Catalogue (CIDOC/CRM)

• ResearchSpace.org
  – More institutions
  – Research environment
Round-topped limestone stela (222BC-204BC)

Round-topped limestone stela divided into five registers (400BC)

Fragment from the lower part of a painted sandstone stela (4th BC BC(late))

Upper part of a painted round-topped sandstone stela (4thC BC(late))

Object ID: http://bm2.rkbexplorer.com/id/object/YCA56669

**Details**

**Owner:** The British Museum

**Physical description:** Upper part of a painted round-topped sandstone stela: beneath a winged sun-disc with pendant uraei and heaven sign, Alexander the Great is depicted as pharaoh, offering two vessels of wine to the sacred Buchis bull, supported on a plinth. Parts of six lines of hieroglyphic text are preserved, recording the death of the sacred bull in the fourth year of Alexander, and noting that it was alive during the reign of Darius (III), who controlled Egypt from 336 BC until the conquest of Alexander in 332 BC. Part of the lower part of the stela bearing six rows of Hieroglyphic text has been re-affixed; this was originally registered separately as 1929.1016.202 (EA 1719).

**Image**

**Annotation list:**

- note: The lower part is actually referenced as another object in the British Museum collection. 27/01/2011

**Science**

**Title:** The Identification of Egyptian Limestone Sculpture likely to Deteriorate

**Reason:** Conservation Research Lab Report

**Conservation**

Remove and investigate the makeup between the two fragments (1929 10-16 167 & 1927 10-16 202). The fill between the two fragments should be recessed and retouched to match the surrounding stone. Investigate the consolidant that has been used on the lower fragment, this has caused the stone to darken comparative to the upper fragment. Fill and cap the laminations to the surface of the stone, take care not to mask any of the text when filling. Remove the white material on the surface (modern paint/plaster) from the upper fragment. Remove the number from the front of...
Our Example: RKB and RKBExplorer

• RKB (ReSIST Knowledge Base) and RKBExplorer
• Knowledge-enabled infrastructure for cooperation in research into resilient systems
• Reasonably mature system
  – Actually essentially unchanged in 5 years
  – Still running (even though) unfunded
RKBExplorer.com – a person

Rotated people

Srinandan Dasmehaputra
Yannis Kalloglu
Terry Elliott
Nicholas Gibbins
Harith Alani
Paul Lewis
David De Roure
Bo Hu
Kieron O'Hara
Nigel Shadbolt
Kieron O'Hara

People

Kieron O'Hara
Harith Alani
Nicholas Gibbins
David De Roure
Paul Lewis
Paul Smart
Terry Elliott
Bo Hu
Srinandan Dasmehaputra
Yannis Kalloglu
Nick Jennings

Organisations

University of Southampton
Intelligens, Agents and Multimedia Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science
UNIVERSITY OF NOTTINGHAM

Publications

Facilitating Knowledge Management in Perasive Health Care Systems
A Controlled Natural Language Interface for Semantic Media Wiki Using the Rabbit Language
Contextualizing Tags in Collaborative Tagging Systems
Lifelogging: Privacy and Empowerment with Memories for Life
On Measuring Expertise in Collaborative Tagging Systems
On Measuring Expertise in Collaboration

Research Areas

I.2.6. Learning
I.2.11. Distributed Artificial Intelligence
I.3.4. Knowledge Representation Formalisms and Methods
I.3.6. Systems and Software
I.3.5. Online Information Services
I.2.0. General
I.2.7. Natural Language Processing

Full Name:
Nigel Shadbolt
Works For:
University of Southampton
Telephone:
+44 (0)23 8095 7682
Fax:
+44-802-090339
Email Address:
rns@ecs.soton.ac.uk
Web Address:
http://www.ecs.soton.ac.uk/people/nrs
Web Address:
http://en.wikipedia.org/wiki/Nigel_Shadbolt
Description:
Nigel Richard Shadbolt is Professor of Computer Science, University of Southampton in southern
Or a paper
Or a couple of people

And how they are linked
And why are they linked?

Carl Lagoze is connected to Dean Krafft

They are linked by 9 relations.

**Publications**

They have co-authored 6 papers:

- Core services in the architecture of the national science digital library (NSDL)
- An information network overlay architecture for the NSDL
- Metadata aggregation and "automated digital libraries": A retrospective on the NSDL experience

(3 more)

**Affiliations**

They are both affiliated to Cornell University.

**Projects**

They are both members of 2 projects:

- NSDL Technical Network Services: A Cyberinfrastructure Platform for STEM Education
- Collaborative Project: Core Integration - Leading NSDL toward Long-Term Success
Before and after inserting a paper in the Southampton ePrints repository and RKB has noticed
Note the position of Les Carr in Hugh’s related People
One of Many KBs, including other ePrints

<table>
<thead>
<tr>
<th>acm.rkbexplorer.com</th>
<th>italy.rkbexplorer.com</th>
</tr>
</thead>
<tbody>
<tr>
<td>budapest.rkbexplorer.com</td>
<td>kaunas.rkbexplorer.com</td>
</tr>
<tr>
<td>citeseer.rkbexplorer.com</td>
<td>kisti.rkbexplorer.com</td>
</tr>
<tr>
<td>cordis.rkbexplorer.com</td>
<td>laas.rkbexplorer.com</td>
</tr>
<tr>
<td>courseware.rkbexplorer.com</td>
<td>lisbon.rkbexplorer.com</td>
</tr>
<tr>
<td>darmstadt.rkbexplorer.com</td>
<td>newcastle.rkbexplorer.com</td>
</tr>
<tr>
<td>dblp.rkbexplorer.com</td>
<td>nsf.rkbexplorer.com</td>
</tr>
<tr>
<td>dbpedia.org</td>
<td>pisa.rkbexplorer.com</td>
</tr>
<tr>
<td>deepblue.rkbexplorer.com</td>
<td>rae2001.rkbexplorer.com</td>
</tr>
<tr>
<td>deploy.rkbexplorer.com</td>
<td>resex.rkbexplorer.com</td>
</tr>
<tr>
<td>epsrc.rkbexplorer.com</td>
<td>roma.rkbexplorer.com</td>
</tr>
<tr>
<td>eurecom.rkbexplorer.com</td>
<td>southampton.rkbexplorer.com</td>
</tr>
<tr>
<td>ft.rkbexplorer.com</td>
<td>ulm.rkbexplorer.com</td>
</tr>
<tr>
<td>ibm.rkbexplorer.com</td>
<td>unlocode.rkbexplorer.com</td>
</tr>
<tr>
<td>ieee.rkbexplorer.com</td>
<td>wiki.rkbexplorer.com</td>
</tr>
<tr>
<td>irit.rkbexplorer.com</td>
<td>xxx.yyy.zzz</td>
</tr>
</tbody>
</table>

Range from a few 100 to more than 10,000,000 “facts”
iPhone App

Hugh Glaser

Telephone  +44-1703-593670

Works For School of Electronics and Computer Science, University of Southampton

Full Name Hugh Glaser

Fax  +44-1703-593045

Web  http://www.ecs.soton.ac.uk/

Address  people/hg

Projects

- ReSIST Resilience for Survivability in IST
- HELIOS

Research Areas

- Static Analysis
Semantic Web is not “All or Nothing”

Hugh Glaser
School of Electronics and Computer Science
University of Southampton
Southampton
SO17 1BJ
United Kingdom

Position: Visiting academic staff in Intelligence, Agents, Multimedia Group
Fax: +44 (0)23 8059 3045
Email: hg@ecs.soton.ac.uk
URL: http://id.ecs.soton.ac.uk/person/21 [browse]

Interests: genetic algorithms, ontology maintenance, semantic grid, semantic squirrels, semantic web, semantic wiki, service composition, ubiquitous computing, web 2.0

The group secretary for the Intelligence, Agents, Multimedia Group is Jane Morgan.

Additional Contact Details

I may be here

Duties

ECS - Staff Training Budget
ECS - Travel Budget
Staff Teaching allocation (with Mark Zwolinski)
Professor David C De Roure

School of Electronics and Computer Science
University of Southampton
Southampton
SO17 1BJ
United Kingdom

Positions: Head of Group in Grid and Pervasive Computing Group, Academic staff in Intelligence, Agents, Multimedia Group

Extension: 22418
Telephone: +44 (0)23 8059 2418
Email: dder@ecs.soton.ac.uk
URI: http://id.ecs.soton.ac.uk/person/47 [browse]

Interests: agent-based computing, autonomic computing, arts and humanities applications, audio, autonomic computing, chemoinformatics, collaborative tools, computers and music, decentralised systems, distributed programming languages, distributed systems, e-who, e-learning, e-research, e-science, e-social science, electronic lab notebooks, electronic systems design, environmental monitoring, folksonomies, grid, grid computing, grid standards, hardware-software co-design, hypertext, information systems, intelligent sensor networks, interaction, ipv6, linux, lisp, middleware, multi user dungeons, multicast, multimedia, music, music information retrieval, nature inspired computing, open access, open hypermedia, open journals, open source, pervasive computing, provenance, rdf, recommender systems, scheme, scholarly communications, scripting languages, self-organising systems, semantic annotation, semantic grid, semantic web, sensor networks, service-oriented architectures, simulation, social networks, social tagging, surge, technology-enhanced learning, ubiquitous computing, virtual research environments, visual programming, web, web 3.0, web science, web services, web standards, workflow

The group secretary for the Intelligence, Agents, Multimedia Group is Jane Morgan.

Biography

David De Roure is a Professor of Computer Science in the School of Electronics and Computer Science at the University of Southampton, UK. A founding member of the School's Intelligence, Agents, Multimedia Group, he leads the e-Research activities and is a Director of the Pervasive Systems Centre. Closely involved in the UK e-Science programme, David is National Strategic Director of e-Social Science, Chair of OMIII-UK and a Co-Director of e-Research South. His work focuses on creating new research methods in and between multiple disciplines and his projects draw on Web 2.0, Semantic Web and workflow technologies.

David works with:
- Professor Dame Wendy Hall
- Dr Danijus T Michaelides
- Dr Les A Carr
- Dr Mark J Weal
- Dr Don Cruickshank
- Professor Nigel R Shadbolt
- Kevin R Page
- Dr Gary B Wills
- Dr David E Miller
- Professor Luc Moreau
- David R Newman
- Hugh Gaiser
- Dr Stuart E Middleton
- Dr Nicholas Gibeaux
- Professor Nicholas Jenkins
- Professor Hugh C. Davis
- Dr Monica Mc Schraefel
- Ken E Meacham
- Professor AIG "Tony" Hey
- Dr Robert J Walters
- Christopher Gutteridge
- Dr Harith Alani
- Yang Yang
- Saad A Alhammad
- Dr Simon Green MD MB (explain)
The LOD “cloud”

Linked Data Principles
1. Use URIs as names for things

• **Everything**
  – If you don’t name something you can’t talk about it
  – Things of course
  – Year of publication
  – 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)

• **RESTful Interfaces**

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

- So they know what you mean
- Deliver some human readable data
  - html
- Deliver some machine processable data
  - RDF
  - JSON
  - CSV
  - text
“Hugh Glaser works for the University of Southampton”

person-00021
  name
    "Hugh Glaser"

person-00021
  works-for
    Southampton_University

Southampton_University
  label
    "The University of Southampton"

How to Publish Linked Data on the Web:- http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/
RDF (Resource Description Framework)

“Hugh Glaser works for the University of Southampton”

<http://southampton.rkbexplorer.com/id/person-00021>
  <http://xmlns.com/foaf/0.1/name>
  "Hugh Glaser" .

<http://southampton.rkbexplorer.com/id/person-00021>
  <http://aktors.org/ontology/portal#works-for>
  <http://dbpedia.org/resource/Southampton_University> .

<http://dbpedia.org/resource/Southampton_University>
  <http://www.w3.org/2000/01/rdf-schema#label>
  "The University of Southampton" .
RDF (Resource Description Framework)

- RDF Triples constitute a graph
  - <subject-uri> <predicate-uri> <object-uri>
  - <subject-uri> <predicate-uri> “String”

- ontologies define vocabularies, types and relationships

- Agreed URIs facilitate linkage between datasets
4. Include links to other URIs. so that they can discover more things

- “Foreign” URIs
  - dbpedia:Southampton_University

- Equivalence
  - owl:sameAs
  - skos:exactMatch
  - ...
Where to put the RDF?

- In a file – this is the Web!
- In a store
  - Knowledge Base
  - With a way of selecting and packaging up the RDF when a HTTP URI request comes in
  - Or querying using SPARQL
    - Similar to a Data Base and SQL

5 stars of Open Linked Data

• From “Raw Data Now” to Real Value
One Star

Put your Data on the Web
With an Open License
Two Stars

Make it Available in a Machine Readable Format
Three Stars

Use an Open, Standard, Format
Four Stars

Use an Open, Linked Data, Format
Five Stars

Link Your Data to Other People’s Data
Making Processing “safer”

- URIs
- Semantic Web
- Difficult problem
- Natural Language
- rdfs:label
Example – Presentation in HTML
Presentation in RDF

http://eprints.ecs.soton.ac.uk/id/eprint/21281

bibo:authorList <http://eprints.ecs.soton.ac.uk/id/eprint/21281#authors>;
bibo:presentedAt epid:event/ext-7db7153daf85bc7459505d6c30e743ff;
dct:creator epid:person/ext-21, epid:person/ext-4860;
dct:date "2010-06-22";
dct:isPartOf <http://eprints.ecs.soton.ac.uk/id/repository>;
dct:title "Linked Data: Publishing and Consuming Data on the Semantic Web"^^xsd:string;
rdf:type
  bibo:AcademicArticle
  bibo:Article,
  ep:ConferenceItemEPrint,
  ep:EPrint; rdfs:seeAlso <http://eprints.ecs.soton.ac.uk/21281/>.

http://eprints.ecs.soton.ac.uk/id/eprint/21281#authors

rdf:_1 epid:person/ext-21;
rdf:_2 epid:person/ext-4860.
Down & Dirty - URIs

http://imgs.xkcd.com/comics/standards.png
The Web of Data has many equivalent URIs. 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 "Southampton", (which we will look up for you on Sindice first) or finding other equivalent identifiers for http://transport.data.gov.uk/id/local-authority/1755?
interlinking the Web of Data

The Web of Data has many equivalent URIs. 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://acm.rbexplorer.com/id/person-344165-529a23877796b9822d610a72843a3980c

2. http://acm.rbexplorer.com/id/person-344165-529a23877796b9822d610a72843a3980c
3. http://acm.rbexplorer.com/id/person-344165-be4c940af0b483a2d1675df9fa0bab6f
4. http://acm.rbexplorer.com/id/person-344165-c23ba3f7f92faec763400f197b99a0d1
5. http://acm.rbexplorer.com/id/person-344165-e0b9eaa0f283efab1c5ea86e681bf111
7. http://dblp.rbexplorer.com/id/people-2eb813899a235014ed48c881a1...86ce3c9dc7
8. http://dblp.rbexplorer.com/id/people-cd930ab4bba08a7318516e5f5f9...f7185e51d3
9. http://dblp.rbexplorer.com/id/people-cd930ab4bba08a7318516e5f5f9...88d685ab7e
10. http://dblp.rbexplorer.com/id/people-cd930ab4bba08a7318516e5f5f9...601d0b0236
11. http://dblp.rbexplorer.com/id/people-cd930ab4bba08a7318516e5f5f9...6b93b5e9e9f
12. http://dblp.rbexplorer.com/id/people-cd930ab4bba08a7318516e5f5f9...872498927e
13. http://dblp.rbexplorer.com/id/people-cd930ab4bba08a7318516e5f5f9...41b544f7f5
14. http://dblp.rbexplorer.com/id/people-cd930ab4bba08a7318516e5f5f9...326149944
15. http://dblp.rbexplorer.com/id/people-cd930ab4bba08a7318516e5f5f9...3a2580ac89
16. http://dblp.rbexplorer.com/id/people-cd930ab4bba08a7318516e5f5f9...28848ff4e0
17. http://dblp.rbexplorer.com/id/people-cd930ab4bba08a7318516e5f5f9...6922936d8c
18. http://dblp.rbexplorer.com/id/people-cd930ab4bba08a7318516e5f5f9...c427fe97ace
19. http://citeseer.rbexplorer.com/id/resource-CSP162937-13ea2982945...80ce6ae74c
20. http://citeseer.rbexplorer.com/id/resource-CSP162937-2897e970649a...00dd146e2c
21. http://citeseer.rbexplorer.com/id/resource-CSP162937-4bf600e2ed7e9...901425f212
22. http://citeseer.rbexplorer.com/id/resource-CSP162937-645954ee407c94...7867a850d
23. http://citeseer.rbexplorer.com/id/resource-CSP162937-8b4779db34e9...af2d24ddc4
24. http://citeseer.rbexplorer.com/id/resource-CSP162937-78664e6e27c06...809dbb880b
25. http://citeseer.rbexplorer.com/id/resource-CSP162937-8ac5adb3355bf...213751b546
What is Southampton?
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://data.archiveshub.ac.uk/id/person/gb248/smitherman1723-1790pol...leconomist –
1. http://dbpedia.org/resource/A_Smith
2. http://dbpedia.org/resource/Adam_Smith
7. http://dbpedia.org/resource/Adam_Smith#Quotes
10. http://mpii.de/yago/resource/Adam_Smith
12. http://rdf.freebase.com/ns/guid.9202a8c0400064180000000000065ab
15. http://sw.opencyc.org/concept/Mx4rwSxnQZwpEbGdrcN5Y29ycA
17. http://sw.opencyc.org/2009/04/07/concept/Mx4rwSxnQZwpEbGdrcN5Y29ycA
<sameAs>
interlinking the Web of Data

The Web of Data has many equivalent URIs. This service helps you to find co-references between different data sets.

Enter a known URI, or use Sindice to search first.

This is a sub-service of the sameAs service, and contains only information that is provided by a third party.

The data presented here relates to Hugh’s experiment in gathering together library subject and associated data from good sources. It is likely to form part of a forthcoming project in the library area.

Equivalent URIs for http://d-nb.info/gnd/4126114-8 –

1. http://bnb.data.bl.uk/ld/concept/csh/Nazis
3. http://id.loc.gov/authorities/sh5090140#concept

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

The Web of Data has many equivalent URIs. This service helps you to find co-references between different data sets.

Enter a known URI, or use Sindice to search first.

This is a sub-service of the sameAs service, and contains only information that is provided by a third party.

The data presented here is provided by Freebase on an "as is" experimental basis.

Why not try finding equivalent identifiers for the URI http://rdf.freebase.com/ns/m.086qd?

Current statistics for this sameAs service (http://sameas.org/store/freebase/) —

Last data assertion 2011-10-26 05:02:30
Number of URIs 54903422

These services are hosted by the School of Electronics and Computer Science at the University of Southampton (UK) in collaboration with Seme4 Limited.
Finding Co-reference

<sameAs> interlinking the Web of Data

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

Search results from Sindice, with co-references applied...

Currently serving 23268924 URIs in 8454882 bundles!
Co-Reference

• **Lots of resources with multiple URIs**
  – This is the nature of the Beast
  – Legal, Sociological and Technical reasons

• **This is a Big Problem**
  – Everything is a URI (not title, name, number…)
  – Identifying multiple URIs for one resource
  – Rejecting incorrectly conflated resources
  – Publishing
  – Using

• **Solution**
  – Co-reference is just Knowledge, but quite special, and crucial, knowledge
  – Generating a new URI is not the solution to already having too many
  – The web is anarchic – don’t try to impose impossible structure on it
  – Embrace the anarchy and multiple URIs in your systems ab initio
Co-Reference Service (CRS)

• **CRS Subsystem**
  – Find co-references
  – Store them
  – Publish them
    • Essentially:
      • URI \( i \rightarrow \{ \text{URI}_1, \ldots, \text{URI}_i, \ldots, \text{URI}_n \} \)
    – Recommend a “Canon”

• **Published by the Data Publisher**
  – And possibly others

• **Middleware aggregates co-references from recognised CRSes**

• **Coldstart**
  – A serious problem
  – Nothing is linked to anything
  – Not even (reliably) within most repositories
Concluding Remarks

• Linked Data works pretty well
• RDF works pretty well
• RKBExplorer uses CRS services successfully
  – Often 100s of URIs per person
• sameAs.org turns out to be pretty useful
  – > 10K hits/day
• British Museum uses a CRS service for Data Fusion on its web site
• Don’t stop at the catalogue – there is a lot of other interesting stuff
• Organisations can go ahead and build systems, then do the identity management
  – Both internally and externally
  – Doing identity management first requires cross-organisation agreement, a delay of months to years
  – It then freezes the business processes to conform to the external requirement
More Concluding Remarks

• **Worry about your co-reference**
  – Do you have strong IDs?

• **Build systems that accept the way the world is, not what you would like it to be**

• **Consume things yourself!**
  – The only way to find out if things are any good
  – Then you might be able to:

• **Provide appropriate services in appropriate formats**
  – Consumable by “normal” systems
  – This means
    • Useful – doing useful things
    • RESTful – easy to access
    • JSON – easy to process
    • ...

• **A little Ontology goes a long way**
Endnotes
http://eprints.ecs.soton.ac.uk/id/eprint/23054

• Some URIs
  – http://sameas.org/
  – http://sameas.org/store/kelle
  – http://www.rkbexplorer.com/
  – http://www.rkbexplorer.com/services/
  – http://apps.seme4.com/see-uk/crime/by-population/ward/OX1%203QG
  – http://www.ecs.soton.ac.uk/people/dder
  – http://www.dotac.info/explorer/
  – http://www.rkbexplorer.com/gadgets/

• You are Never Alone
  – Ian Millard
  – ReSIST Project
  – AKT Project
  – EnAKTing Project
  – 10 years of collaborators

© 2010 Seme4