http://eprints.ecs.soton.ac.uk/id/eprint/23054

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

Seme4

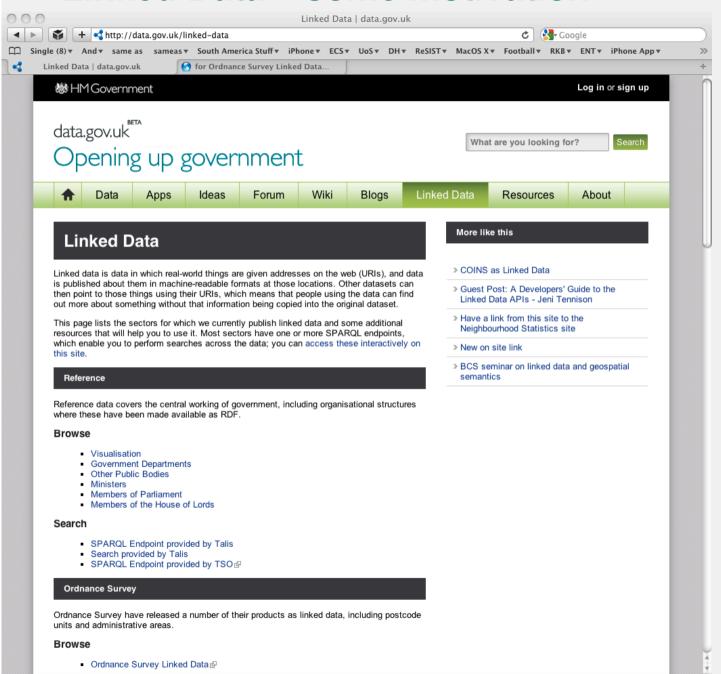
18 Soho Square London W1 3QL

T:+44 20 7060 1590





Linked Data – some motivation





Seme4.com



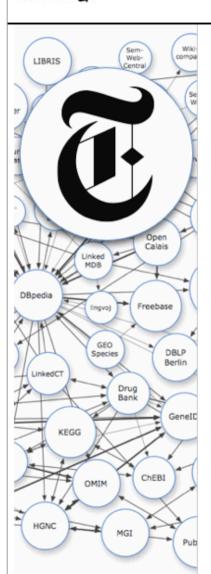




The New Hork Times

Linked Open Data BETA





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.

Туре	Manually Mapped Tags	Automatically Mapped Tags	Total
People	4,978	0	4,978
Organizations	1,489	1,592	3,081
Locations	1,910	0	1,910
			9,969

Browse individual data records:



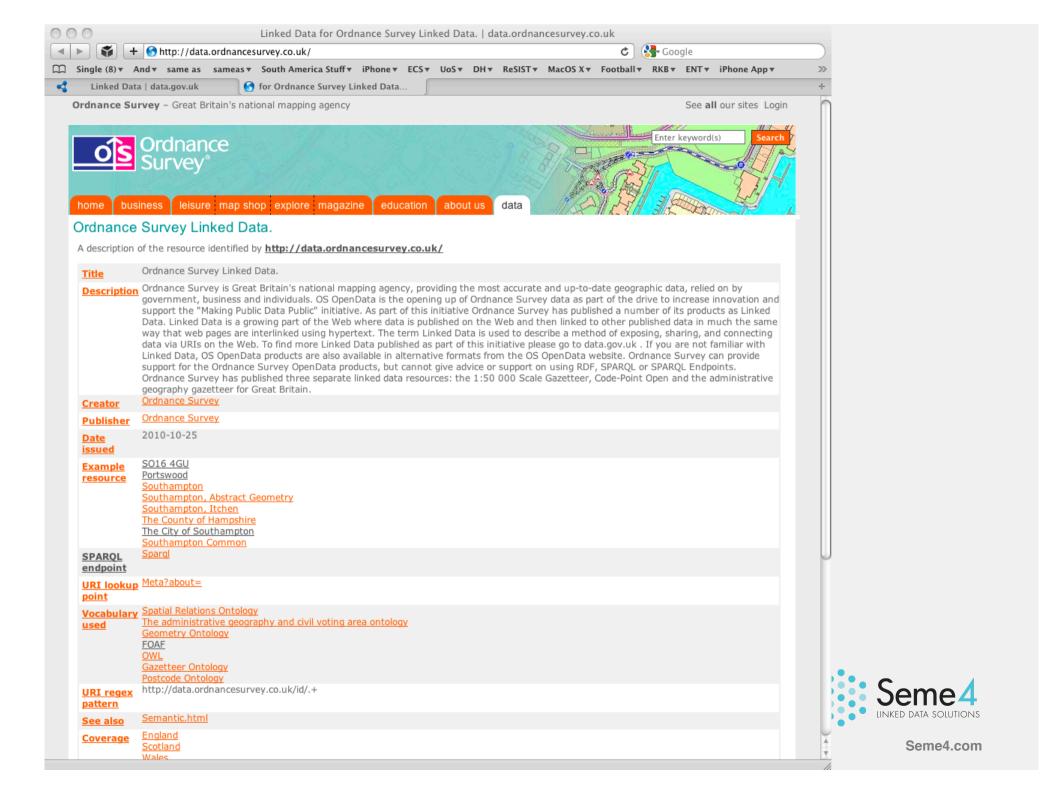
SKOS Files

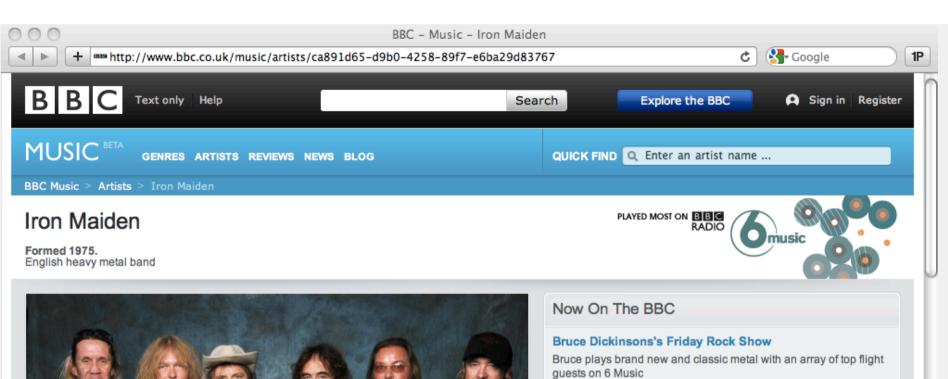
Download all of the data records as SKOS Files.

- 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.





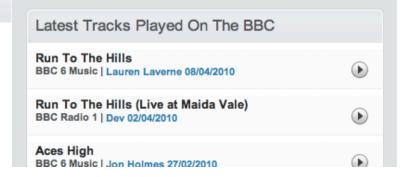






Iron Maiden are an English heavy metal band from Leyton in East London, formed in 1975. The band is directed by founder, bassist and songwriter Steve Harris. Since their inception, the group has released a collective total of thirty albums: fourteen studio albums, seven live albums, four EPs and five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





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.



Staff from the BBC's online and technology teams talk about BBC Online, BBC iPlayer, and the BBC's digital and mobile services. The blog is reactively moderated. Posts are normally closed for comment after three months. Your host is Nick Reynolds (NR).

Subscribe to BBC Internet Blog

You can stay up to date with BBC Internet Blog via these feeds.

BBC Internet Blog Feed (RSS)

BBC Internet Blog Feed (ATOM)

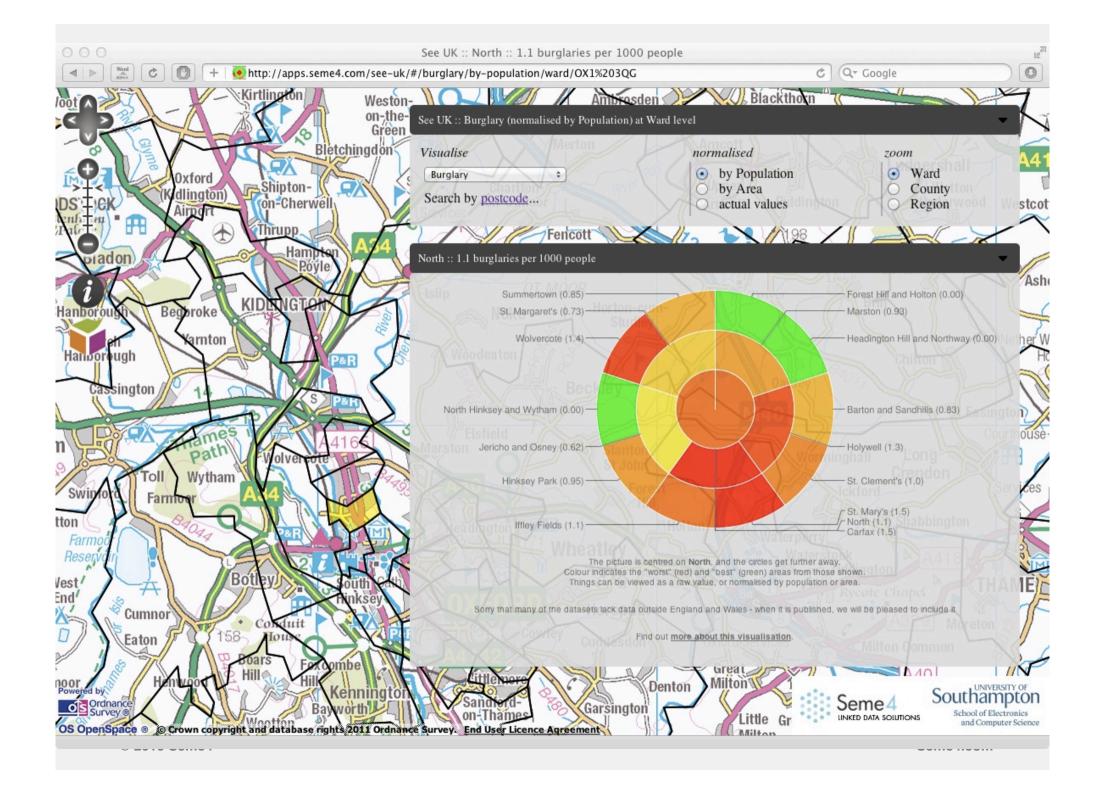
If you aren't sure what RSS is you'll find our beginner's guide to RSS useful.

BBC on Blogs

 BBC iPlayer Gets More S4C VOD, Will Also Link Out

15

- From paidContent (NR)
- Whatever happened to surround sound for streaming Radio?



CERN DD/OC
Tim Berners-Lee, CERN/DD
Information Management: A Proposal

March 1989

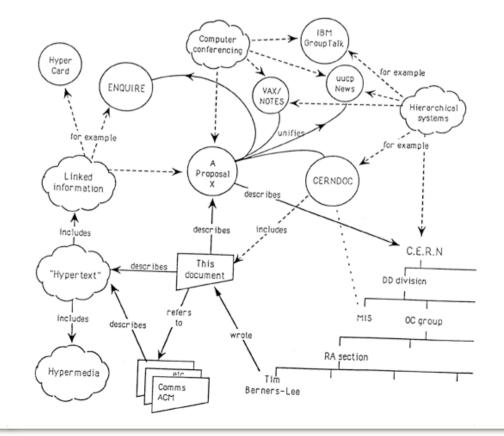
http://www.w3.org/History/1989/proposal.html

Information Management: A Proposal

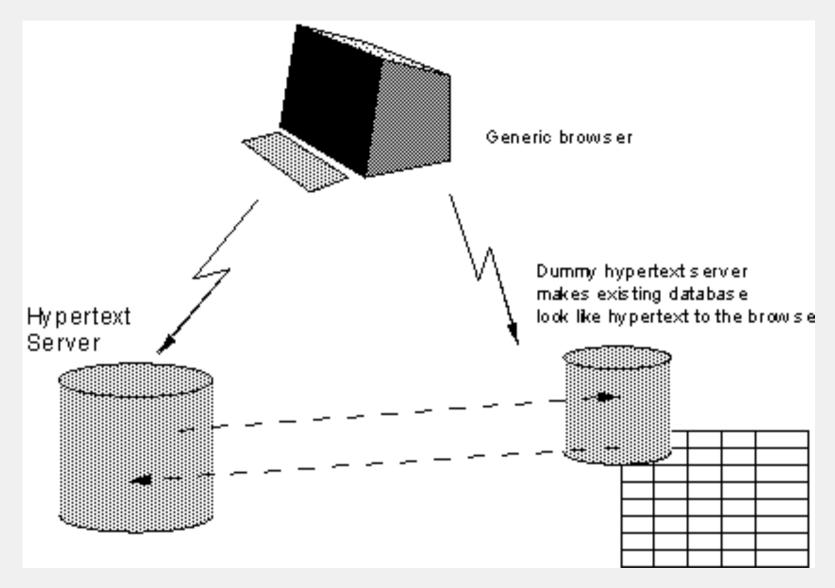
Abstract

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



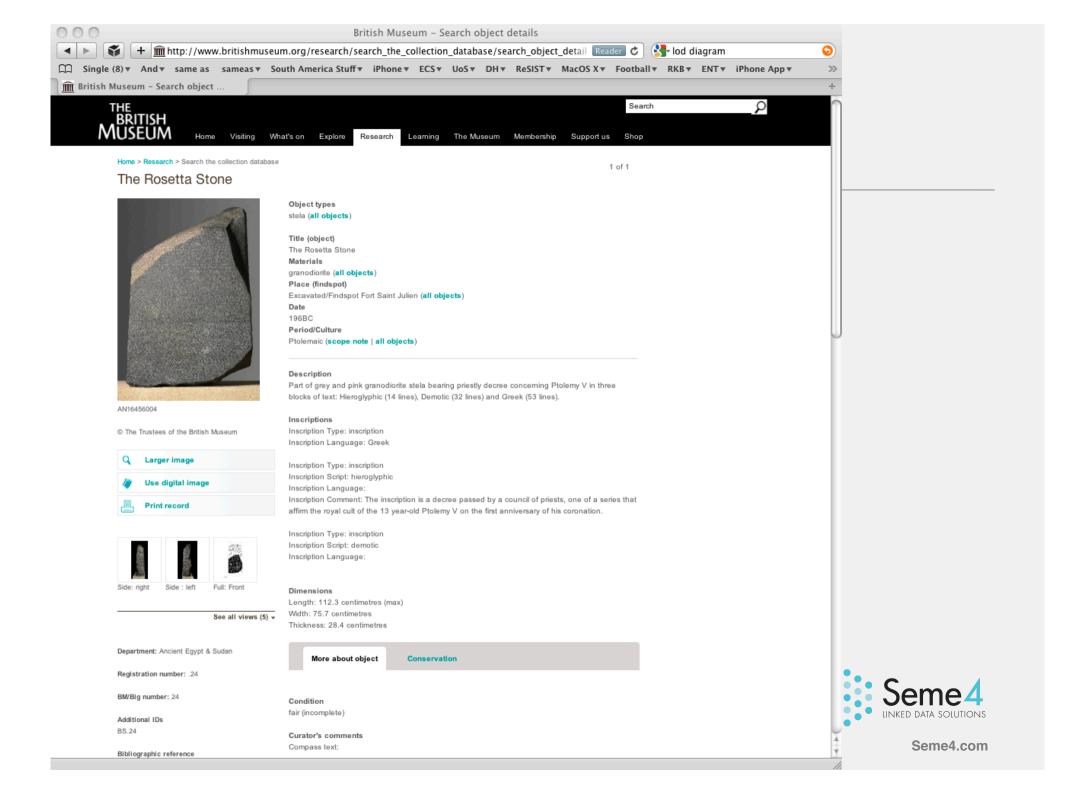


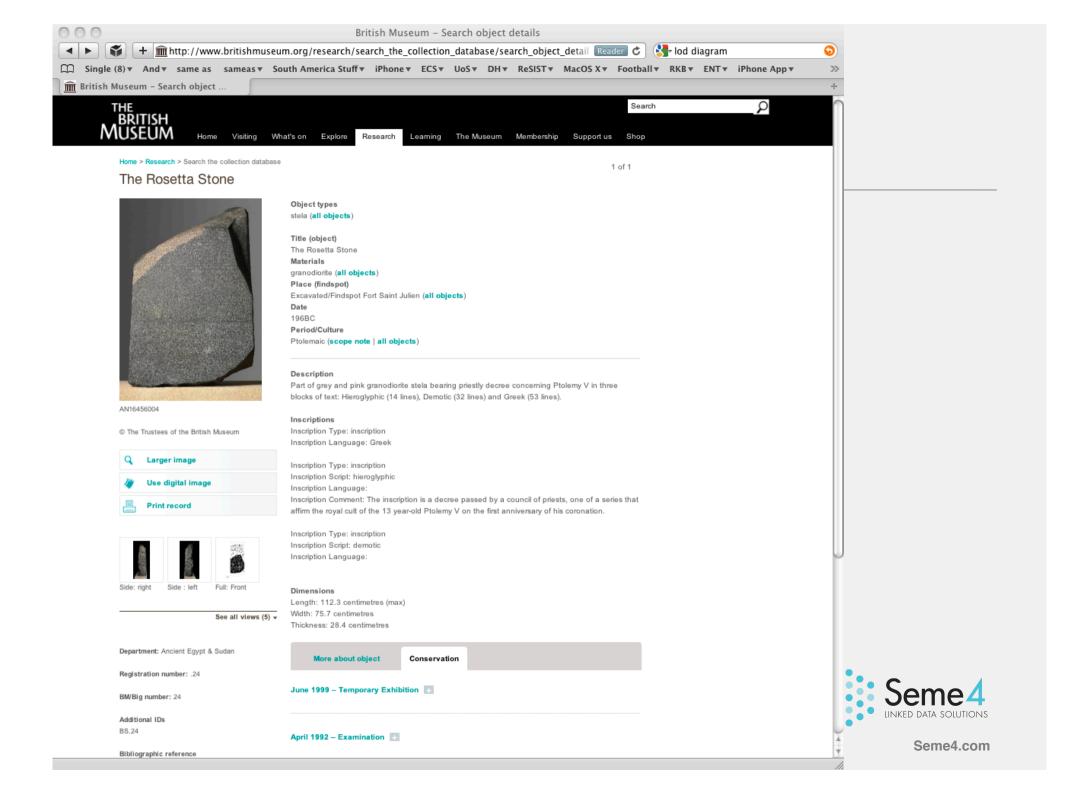


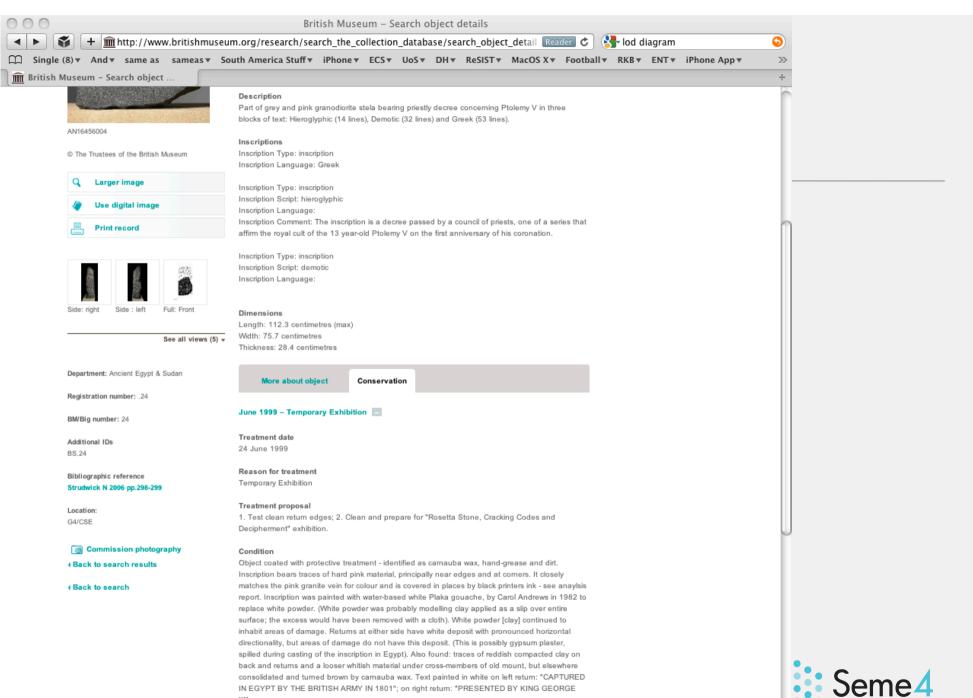


- "The ability of the semantic web to cheaply but effectively integrate data and breakdown data silos provides museums with a long awaited opportunity to present a richer, more informative and interesting picture."
 - » Dominic Oldman, British Museum









Treatment details

Samples were taken of: pink material and printers ink in inscription; surface coating from various

locations; white deposit on returns; reddish compacted clay on back and returns; loose whitish



Seme4.com

British Museum

That was a while ago

- Science data
- Conservation data
- Linked to the Catalogue
- Live on the web site

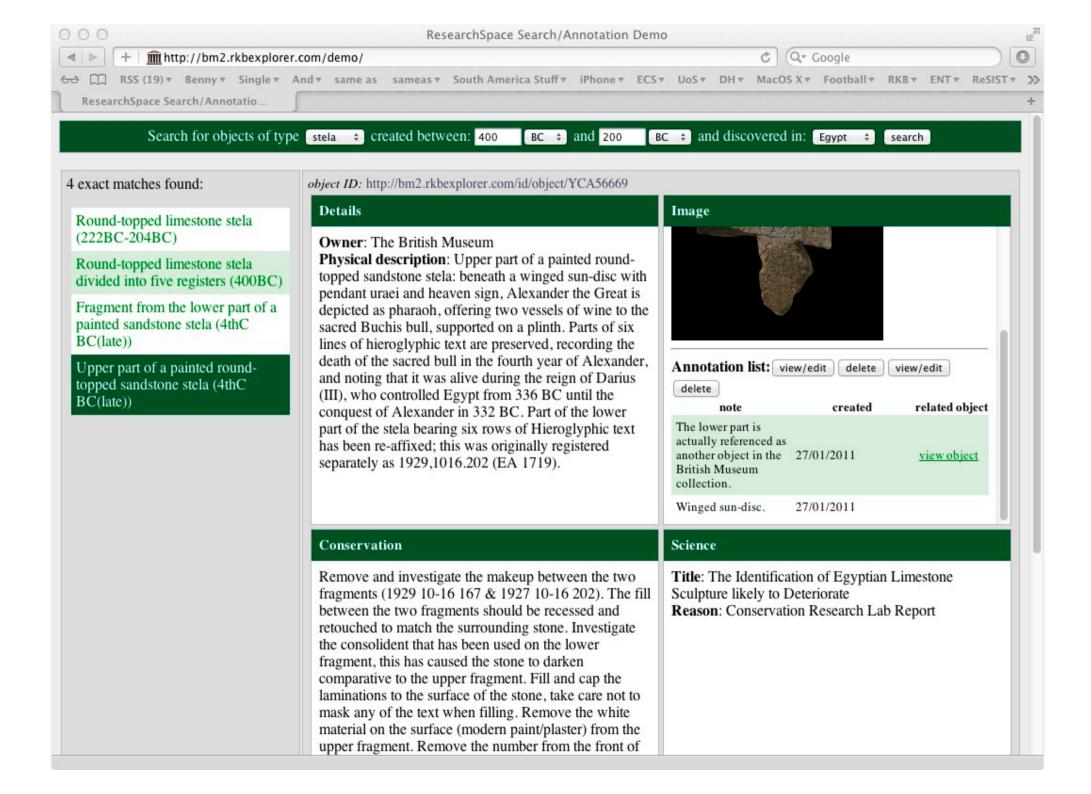
Then

Catalogue (CIDOC/CRM)

ResearchSpace.org

- More institutions
- Research environment



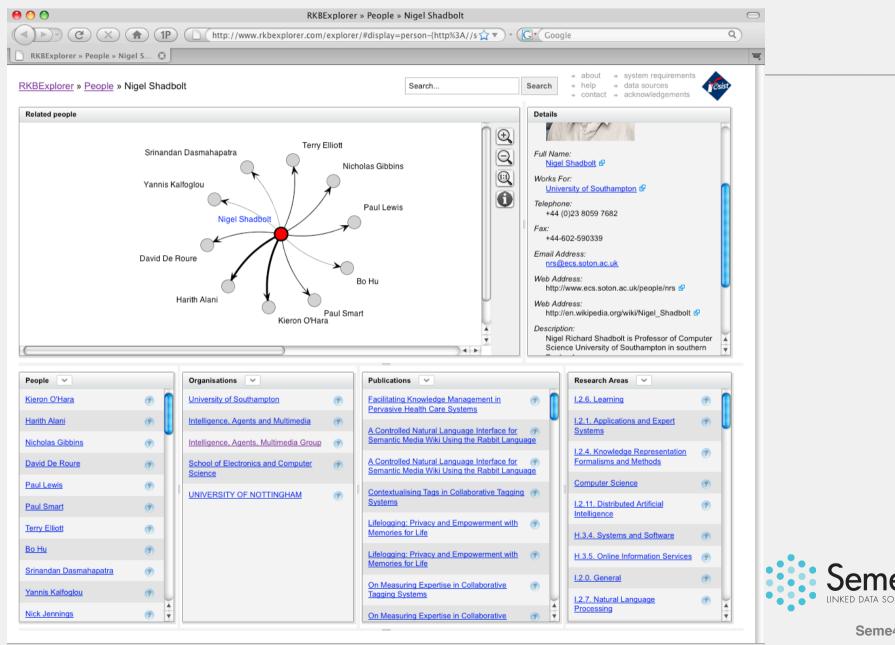


Our Example: RKB and RKBExplorer

- 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



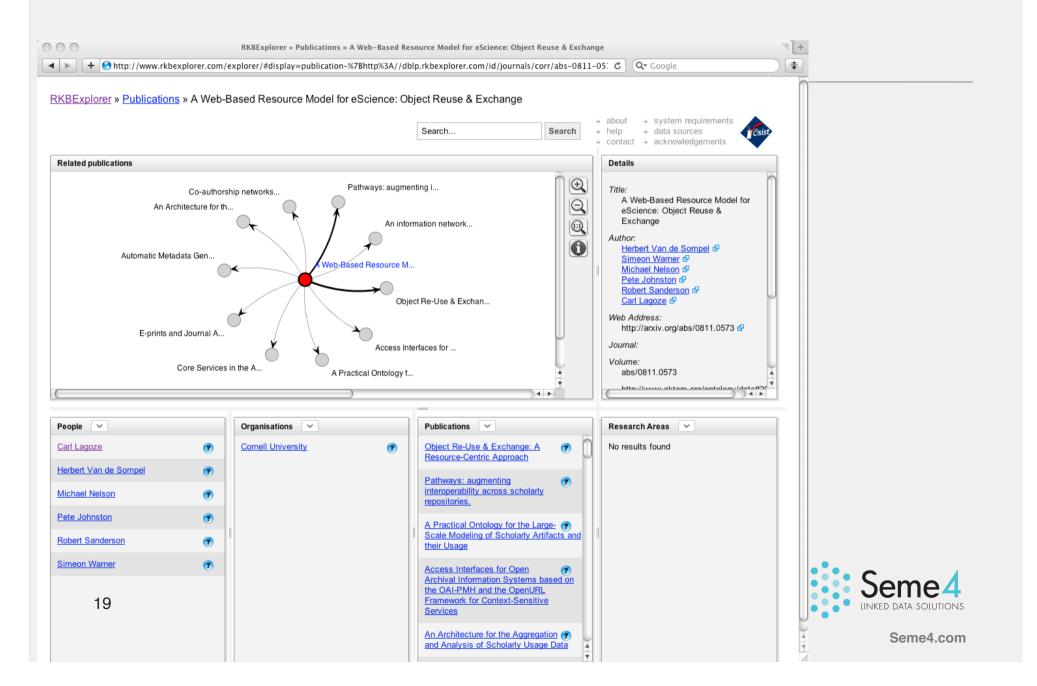
RKBExplorer.com – a person



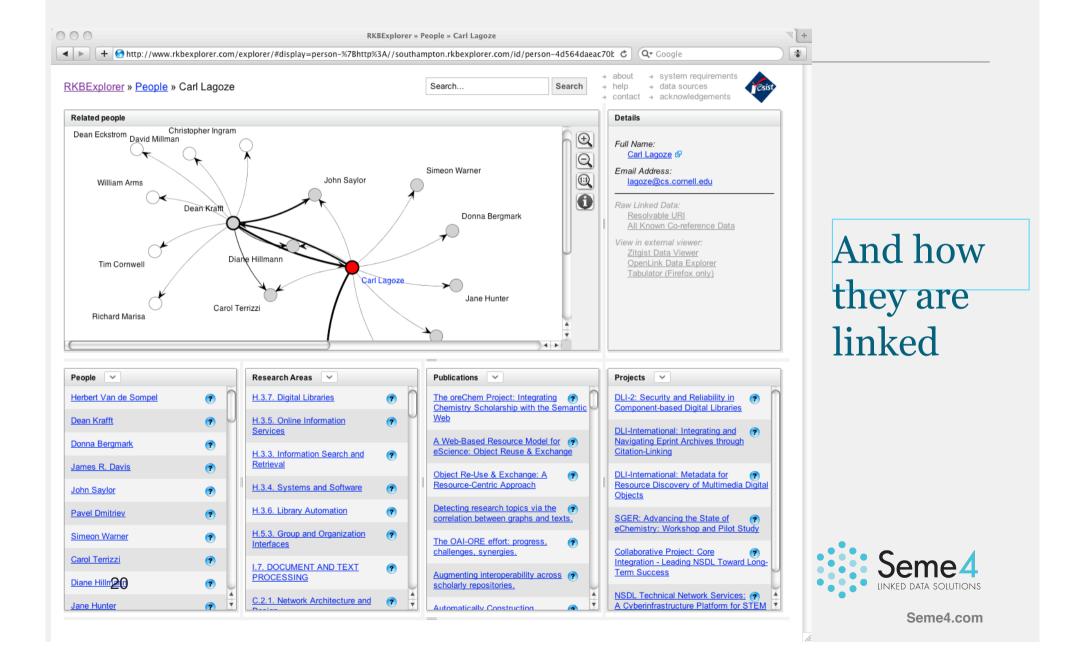


Seme4.com

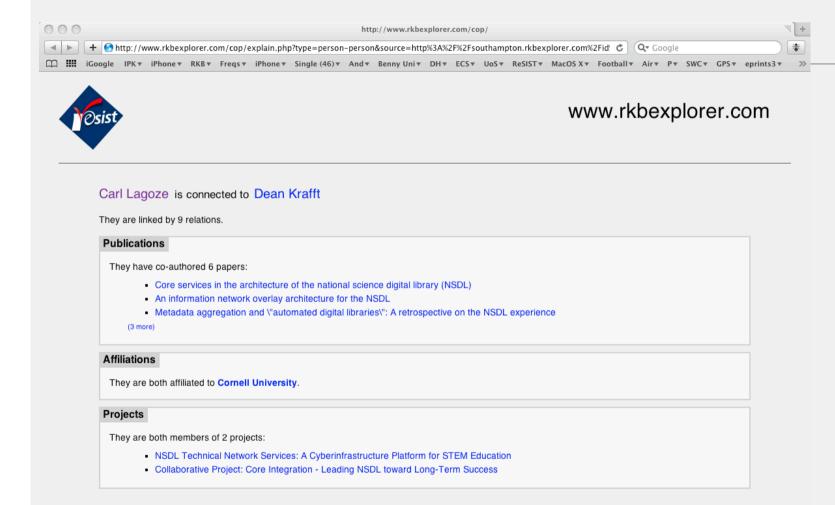
Or a paper



Or a couple of people

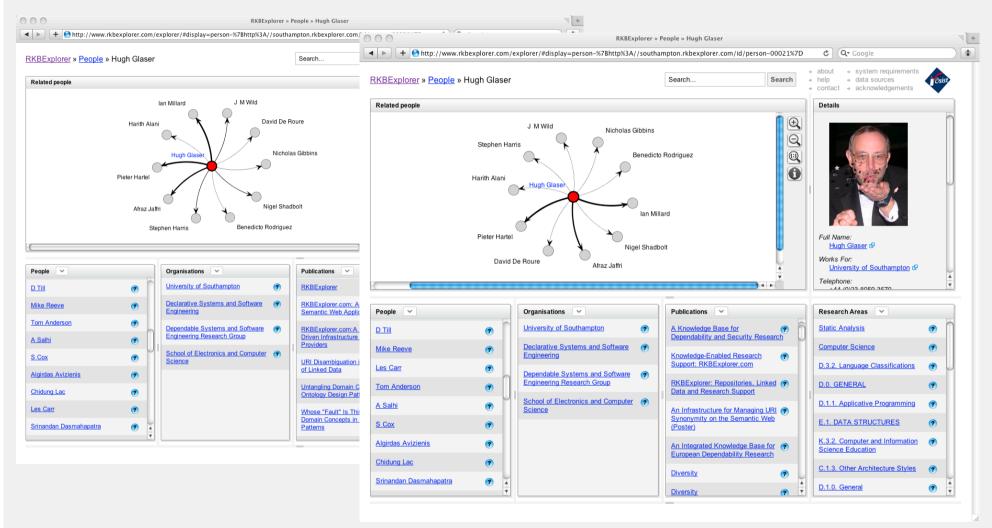


And why are they linked?





Update in Action



Before and after inserting apaper in the Southampton ePrints repository and RKB has noticed

Note the position of Les Carr in Hugh's related People

Seme4.com

One of Many KBs, including other ePrints

acm.rkbexplorer.com

budapest.rkbexplorer.com

citeseer.rkbexplorer.com

cordis.rkbexplorer.com

courseware.rkbexplorer.com

darmstadt.rkbexplorer.com

dblp.rkbexplorer.com

dbpedia.org

deepblue.rkbexplorer.com

deploy.rkbexplorer.com

epsrc.rkbexplorer.com

eurecom.rkbexplorer.com

ft.rkbexplorer.com

ibm.rkbexplorer.com

ieee.rkbexplorer.com

irit.rkbexplorer.com

italy.rkbexplorer.com

kaunas.rkbexplorer.com

kisti.rkbexplorer.com

laas.rkbexplorer.com

lisbon.rkbexplorer.com

newcastle.rkbexplorer.com

nsf.rkbexplorer.com

pisa.rkbexplorer.com

rae2001.rkbexplorer.com

resex.rkbexplorer.com

roma.rkbexplorer.com

southampton.rkbexplorer.com

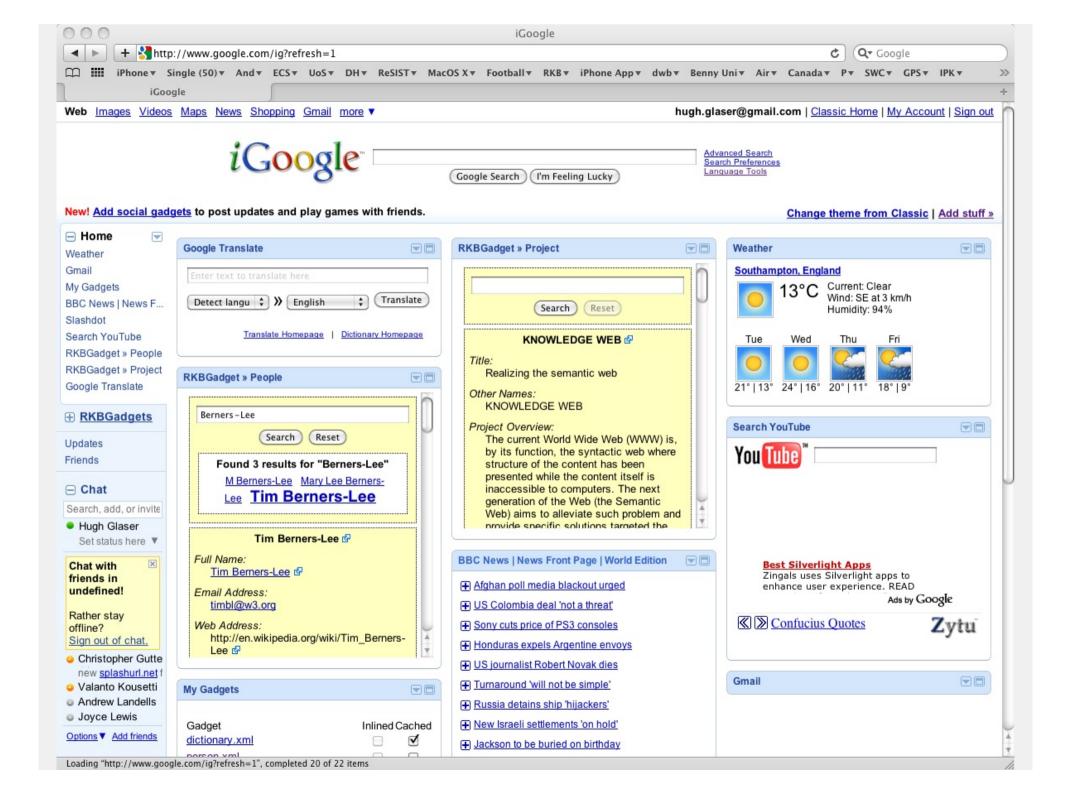
ulm.rkbexplorer.com

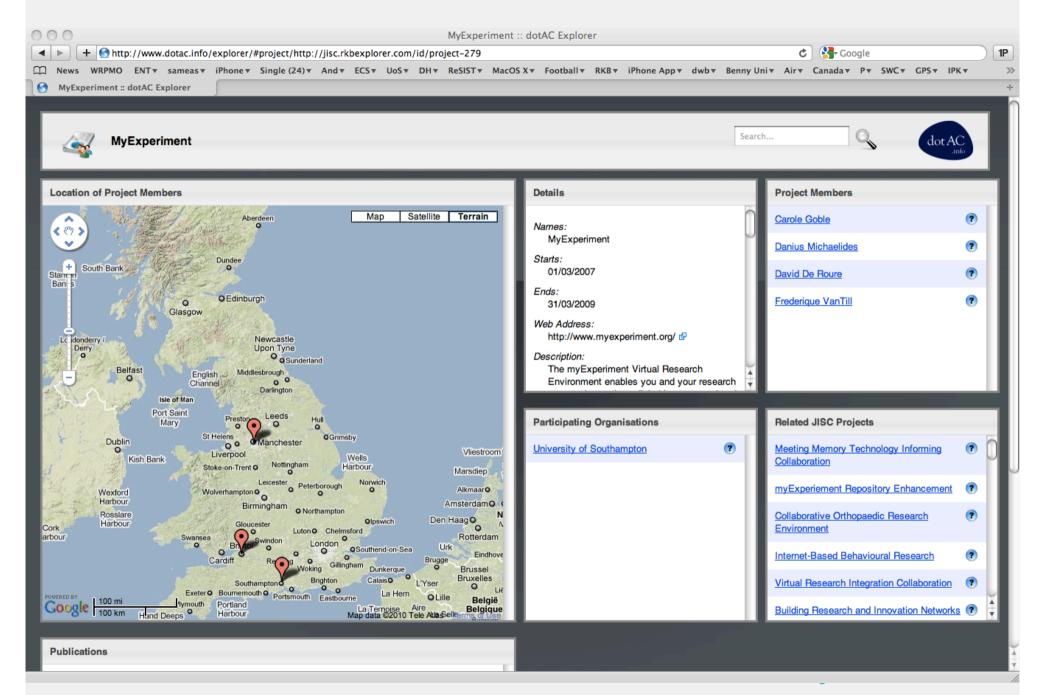
unlocode.rkbexplorer.com

wiki.rkbexplorer.com

XXX.yyy.ZZZ

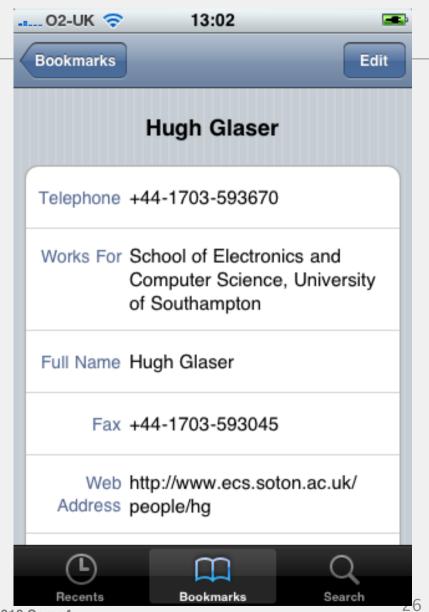
Range from a few 100 to more than 10,000,000 "facts

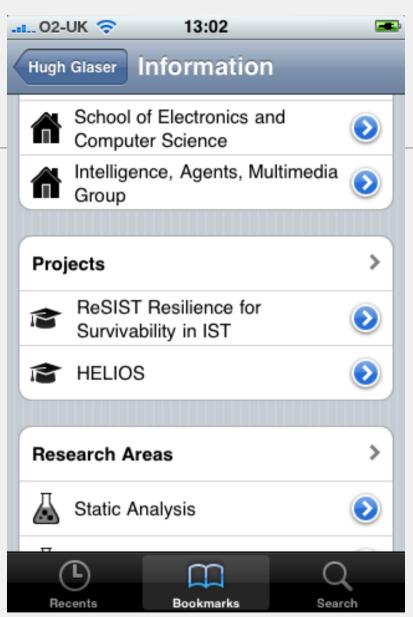




© 2010 Seme4 Seme4.com

iPhone App

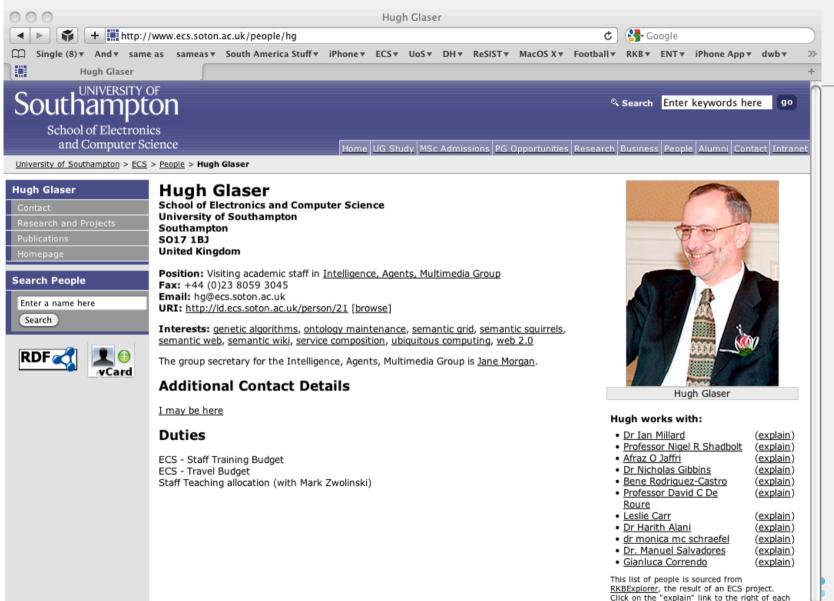






© 2010 Seme4

Semantic Web is not "All or Nothing"



Seme4
LINKED DATA SOLUTIONS

person to find out how they are connected.

Note: the explain link requires heavy processing and may take some time to load.

Search Enter keywords here

Home UG Study MSc Admissions PG Opportunities Research Business People Alumni Contact Intranet

University of Southampton > ECS > People > Professor David C De Roure

Professor David C De Roure

Contact and Biography Research and Projects

Search People

Enter a name here

Search





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

+ | http://www.ecs.soton.ac.uk/people/dder

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, amorphous computing, arts and humanities applications, audio, autonomic computing, chemoinformatics, collaborative tools, computers David works with: and music, decentralised systems, distributed programming languages, distributed systems, dr who, e-learning, e-research, e-science, e-social science, electronic lab notebooks, electronic systems design, environmental monitoring, folksomonies, grid, grid computing, grid standards, hardware-software codesign, hypertext, information systems, intelligent sensor networks, interaction, jpv6, 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, selforganising 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 OMII-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.



Professor David C De Roure

Yang Yang

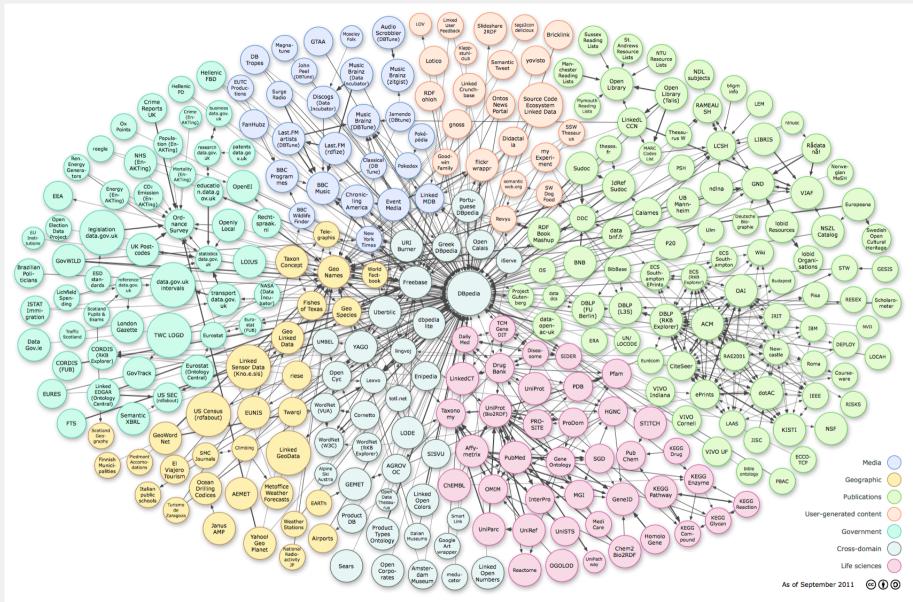
Saad A Alahmari

David works with:				
 Professor Dame Wendy 	(explain)			
<u>Hall</u>				
 Dr Danius T Michaelides 	(explain)			
Dr Les A Carr	(explain)			
 Dr Mark J Weal 	(explain)			
 Dr Don Cruickshank 	(explain)			
· Professor Nigel R Shadbolt	(explain)			
Kevin R Page	(explain)			
Dr Gary B Wills	(explain)			
 Dr David E Millard 	(explain)			
 Professor Luc Moreau 	(explain)			
 David R Newman 	(explain)			
Hugh Glaser	(explain)			
 Dr Stuart E Middleton 	(explain)			
 Dr Nicholas Gibbins 	(explain)			
 Professor Nicholas 	(explain)			
Jennings				
 Professor Hugh C. Davis 	(explain)			
 dr monica mc schraefel 	(explain)			
Ken E Meacham	(explain)			
 Professor AJG "Tony" Hey 	(explain)			
 Dr Robert J Walters 	(explain)			
 Christopher Gutteridge 	(explain)			
Dr Harith Alani	(explain)			

(explain)

(explain)

The LOD "cloud"



Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/



29

Linked Data Principles



30

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/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



RDF (Resource Description Framework)

"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> <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
- http://www4.wiwiss.fu-berlin.de/bizer/pub/linkeddatatutorial/



5 stars of Open Linked Data

From "Raw Data Now" to Real Value



One Star

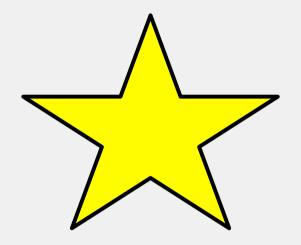
Put your Data on the Web

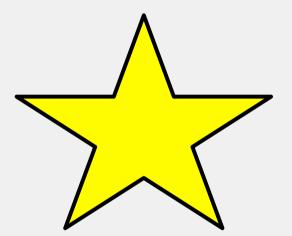
With an Open License



Two Stars

Make it Available in a Machine Readable Format





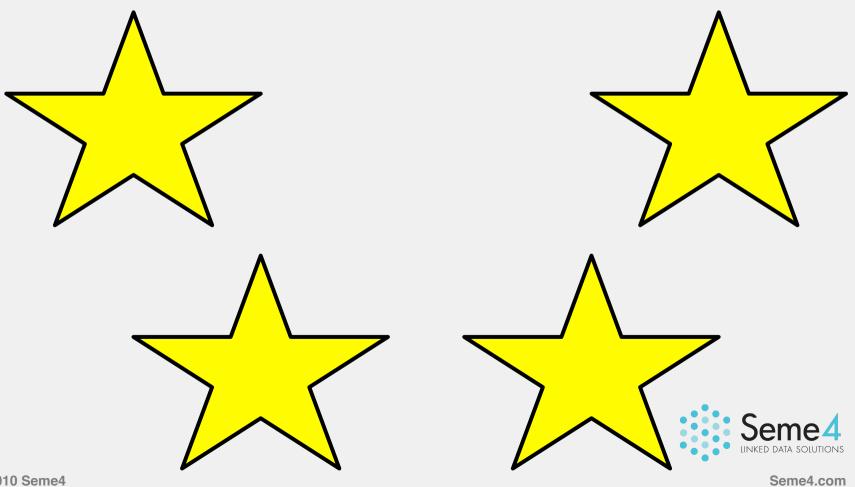


Three Stars

Use an Open, Standard, Format

Four Stars

Use an Open, Linked Data, Format

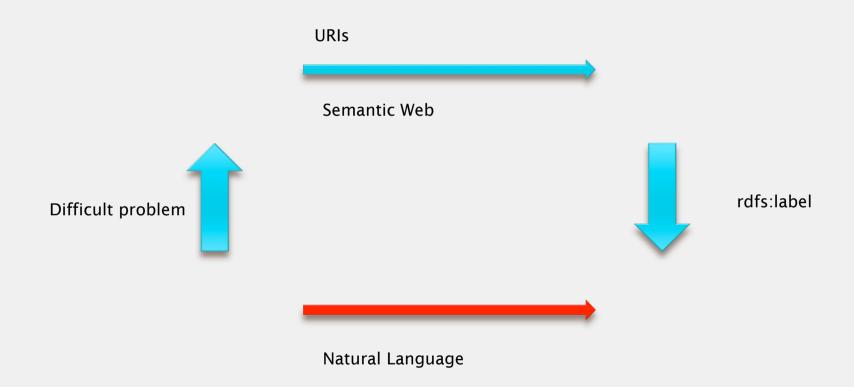


Five Stars

Link Your Data to Other People's Data

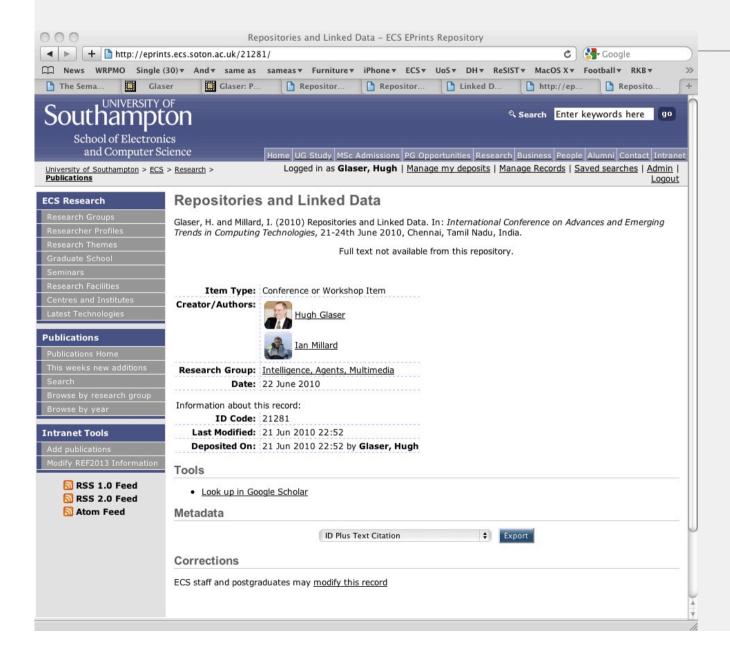


Making Processing "safer"





Example – Presentation in HTML





Presentation in RDF

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

```
bibo:authorList <a href="http://eprints.ecs.soton.ac.uk/id/eprint/21281#authors">http://eprints.ecs.soton.ac.uk/id/eprint/21281#authors</a>;
bibo:presentedAt epid:event/ext-7db7153daf85bc7459505d6c30e743ff;
dct:creator epid:person/ext-21,epid:person/ext-4860;
dct:date "2010-06-22";
dct:isPartOf <a href="http://eprints.ecs.soton.ac.uk/id/repository">http://eprints.ecs.soton.ac.uk/id/repository</a>;
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 <a href="http://eprints.ecs.soton.ac.uk/21281/">http://eprints.ecs.soton.ac.uk/21281/</a>.
```

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

HOW STANDARDS PROLIFERATE: (SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)

SITUATION: THERE ARE 14 COMPETING STANDARDS.



SOUL LATEC

SITUATION: THERE ARE 15 COMPETING STANDARDS.

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.

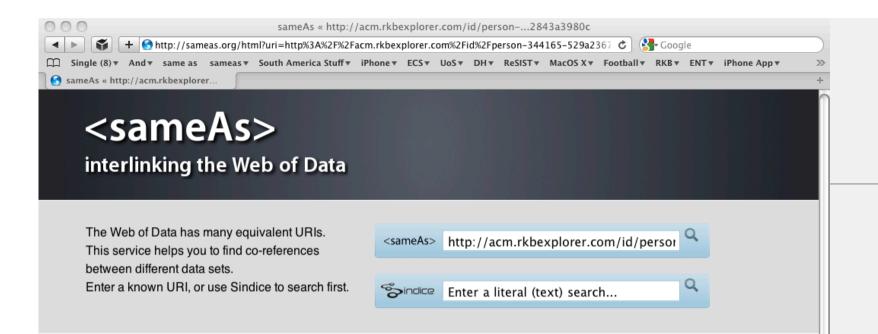
<sameas></sameas>	Enter a Linked Data URI	Q
Sindice	Enter a literal (text) search	Q

Why not try searching for the string "<u>Southampton</u>", (which we will look up for you on <u>Sindice</u> first) or finding other equivalent identifiers for <u>http://transport.data.gov.uk/id/local-authority/1755</u>?

Currently serving 39 221 073 URIs in 13 112 833 bundles!

about · contact · get the widget





Equivalent URIs for http://acm.rkbexplorer.com/id/person-344165-529a2367796b9822d610a72843a3980c -

http://data.semanticweb.org/person/atanas-kiryakov

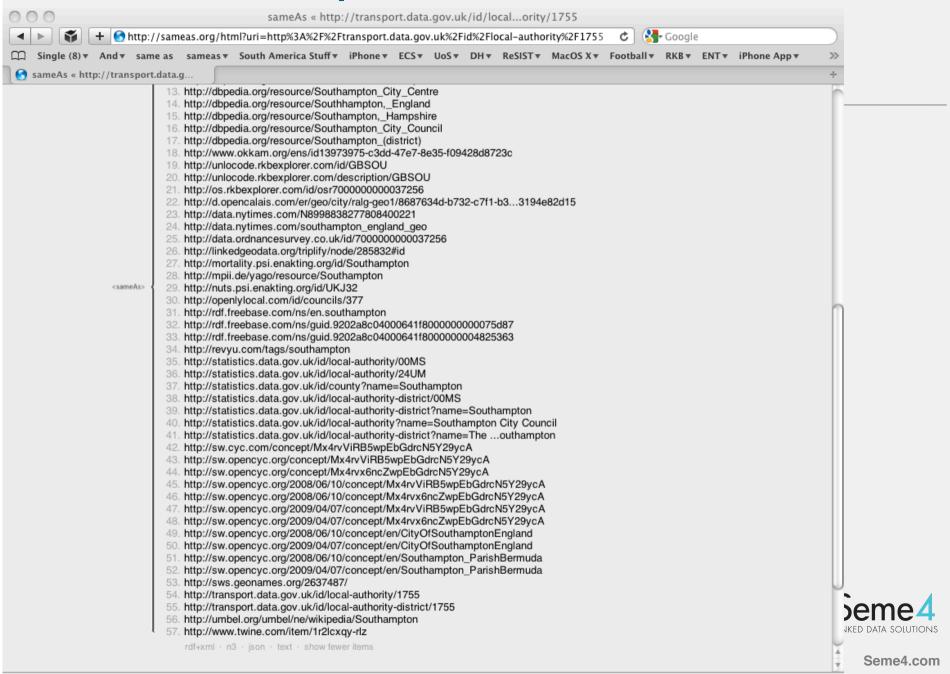
<sameAs>

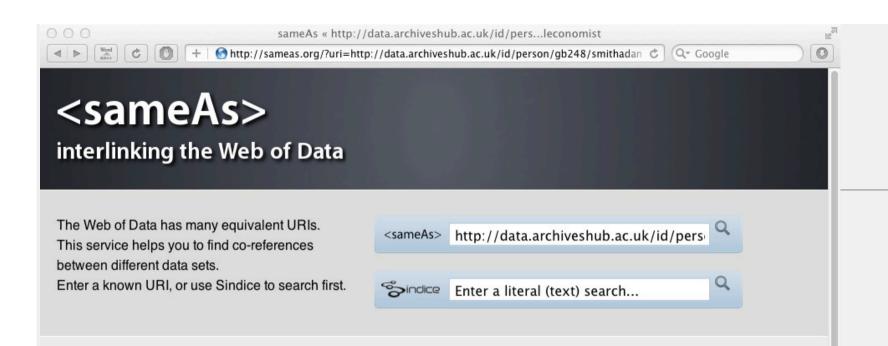
- http://acm.rkbexplorer.com/id/person-344165-529a2367796b9822d610a72843a3980c
- 3. http://acm.rkbexplorer.com/id/person-344165-be4c940af0b483a2d1675df9fa0bab6f
- 4. http://acm.rkbexplorer.com/id/person-344165-c23ba3f792faeca763400f197b89adb1
- http://acm.rkbexplorer.com/id/person-344165-e0b9eea0f283efab1c5ae86ad61bf111
- http://acm.rkbexplorer.com/id/person-344165-fa8e9aed948670c2accad97a4320122d
- 7. http://dblp.rkbexplorer.com/id/people-2ceb8136f69a235014d3d46c881a1...86ce3c9dc7
- 8. http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...f7185e51d3
- 9. http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...88ed85ab7e
- http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...601d0b0236
- 11. http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...b93fb59e9f
- 12. http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...872498927e
- http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...f41b54ff75
- http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...3621e49944
- http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...3a2580ac89
- http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...b8848ff4e0
- 17. http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...6922936d8c
- 18. http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...c42f7e7ace
- 19. http://citeseer.rkbexplorer.com/id/resource-CSP162937-13aaf289f2945...80cea8e74c
- 20. http://citeseer.rkbexplorer.com/id/resource-CSP162937-2897e9670649a...00d1d4c62c
- 21. http://citeseer.rkbexplorer.com/id/resource-CSP162937-4bfe00e2ed7ef...904f25f2f2
- 22. http://citeseer.rkbexplorer.com/id/resource-CSP162937-64595ee407c94...7867a8505d
- 23. http://citeseer.rkbexplorer.com/id/resource-CSP162937-6b4779db34ec9...afd234cdc4
- http://citeseer.rkbexplorer.com/id/resource-CSP162937-7896446de27c6...8f39db880b
- http://citeseer.rkbexplorer.com/id/resource-CSP162937-8ac5adb3359f3...213751b546



Seme4.com

What is Southampton?





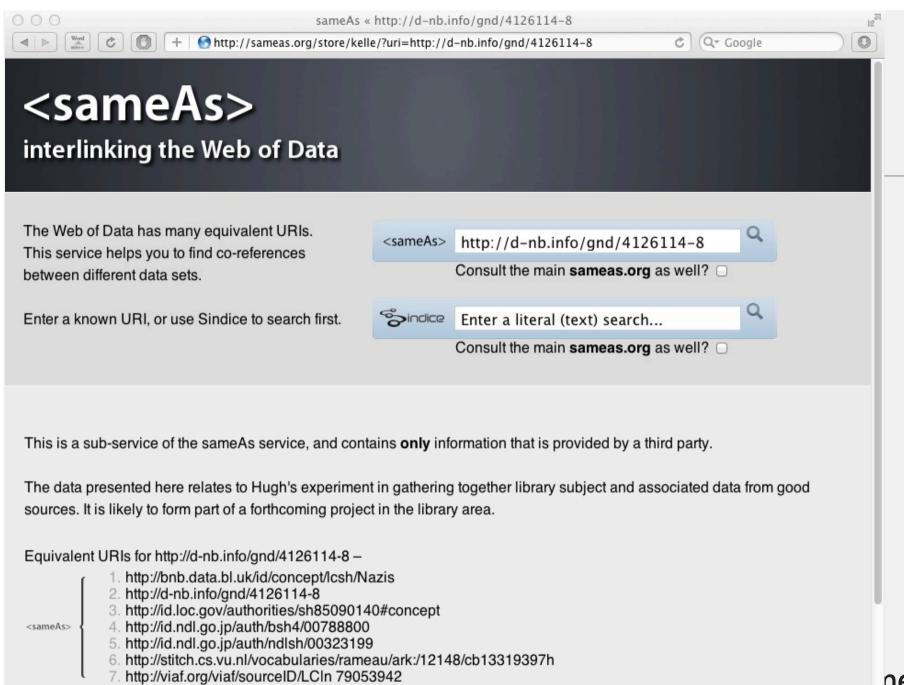
Equivalent URIs for http://data.archiveshub.ac.uk/id/person/gb248/smithadam1723-1790pol...leconomist -

- 1. http://dbpedia.org/resource/A_Smith
- 2. http://dbpedia.org/resource/Adam_Smith
- 3. http://dbpedia.org/resource/Adam_smith
- 4. http://dbpedia.org/resource/Vile_maxim
- 5. http://dbpedia.org/resource/Smith,_Adam
- 6. http://dbpedia.org/resource/Smith,_Adam
- 7. http://dbpedia.org/resource/Adam_Smith#Quotes
- 8. http://d-nb.info/gnd/118615033

<sameAs>

- 9. http://data.archiveshub.ac.uk/id/person/gb248/smithadam1723-1790pol...leconomist
- 10. http://mpii.de/yago/resource/Adam Smith
- 11. http://guotationsbook.com/author/6779
- 12. http://rdf.freebase.com/ns/guid.9202a8c04000641f80000000000065ab
- 13. http://sw.opencyc.org/2008/06/10/concept/en/AdamSmith
- 14. http://sw.opencyc.org/2009/04/07/concept/en/AdamSmith
- 15. http://sw.opencyc.org/concept/Mx4rwSxnQZwpEbGdrcN5Y29ycA
- 16. http://sw.opencyc.org/2008/06/10/concept/Mx4rwSxnQZwpEbGdrcN5Y29ycA
- 17. http://sw.opencyc.org/2009/04/07/concept/Mx4rwSxnQZwpEbGdrcN5Y29ycA
- 18. http://umbel.org/umbel/ne/wikipedia/Adam_Smith
- 19. http://viaf.org/viaf/49231791
- 20. http://www4.wiwiss.fu-berlin.de/gutendata/resource/people/Smith_Adam_1723-1790

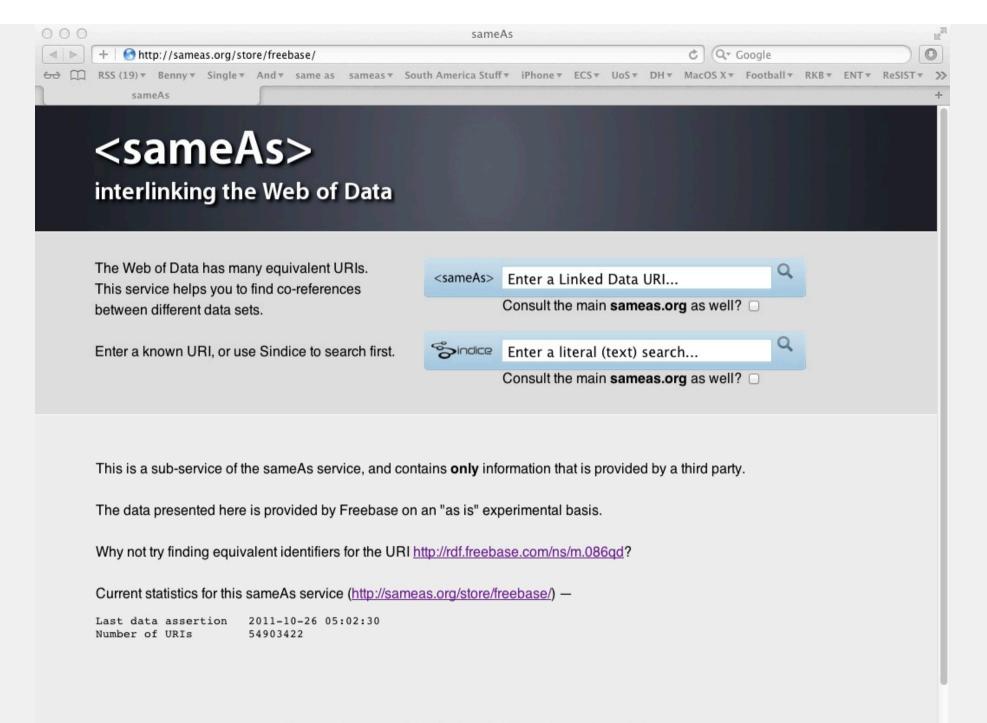




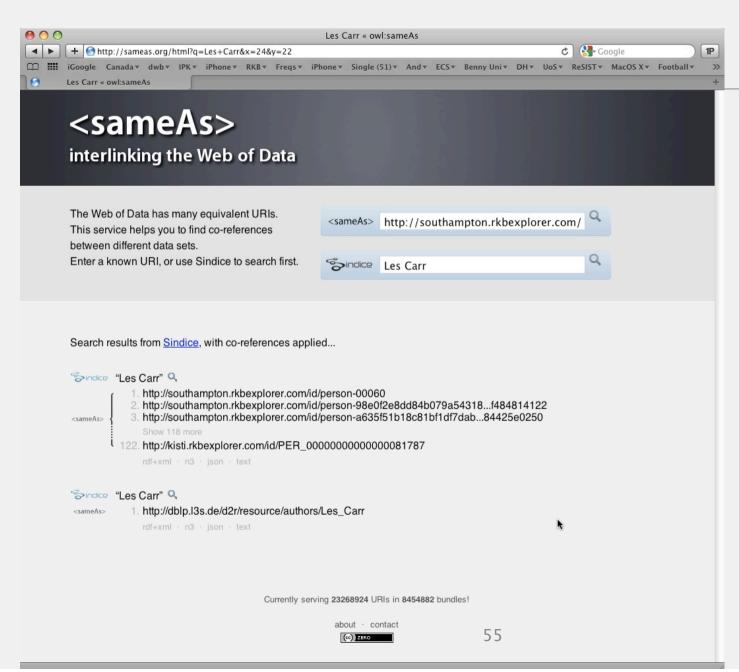
rdf+xml · n3 · json · text · show fewer items



A SOLUTIONS



Finding Co-reference





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 -> { URI₁, ..., URI_i, ..., 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.britishmuseum.org/research/search_the_collection_database/ search_object_details.aspx?objectid=117631&partid=1&searchText=YCA62958
- http://www.dotac.info/explorer/
- http://www.rkbexplorer.com/gadgets/

You are Never Alone

- Ian Millard
- ReSIST Project
- AKT Project
- EnAKTing Project

60 - 10 years of collaborators

