
University of
Southampton

University of Southampton

Linked Data

COMP6215 Semantic Web Technologies

Dr Nicholas Gibbins – nmg@ecs.soton.ac.uk

Linked Data

Semantic Web is the Web for machines

- Take existing data and republish it to the Web
- Rely on hypertextual nature of the Web to facilitate linking between data

- How do we publish this data?
- How does the Semantic Web interoperate with the World Wide Web?
- What identifiers do we use?

University of Southampton

Resources and Identifiers

University of Southampton

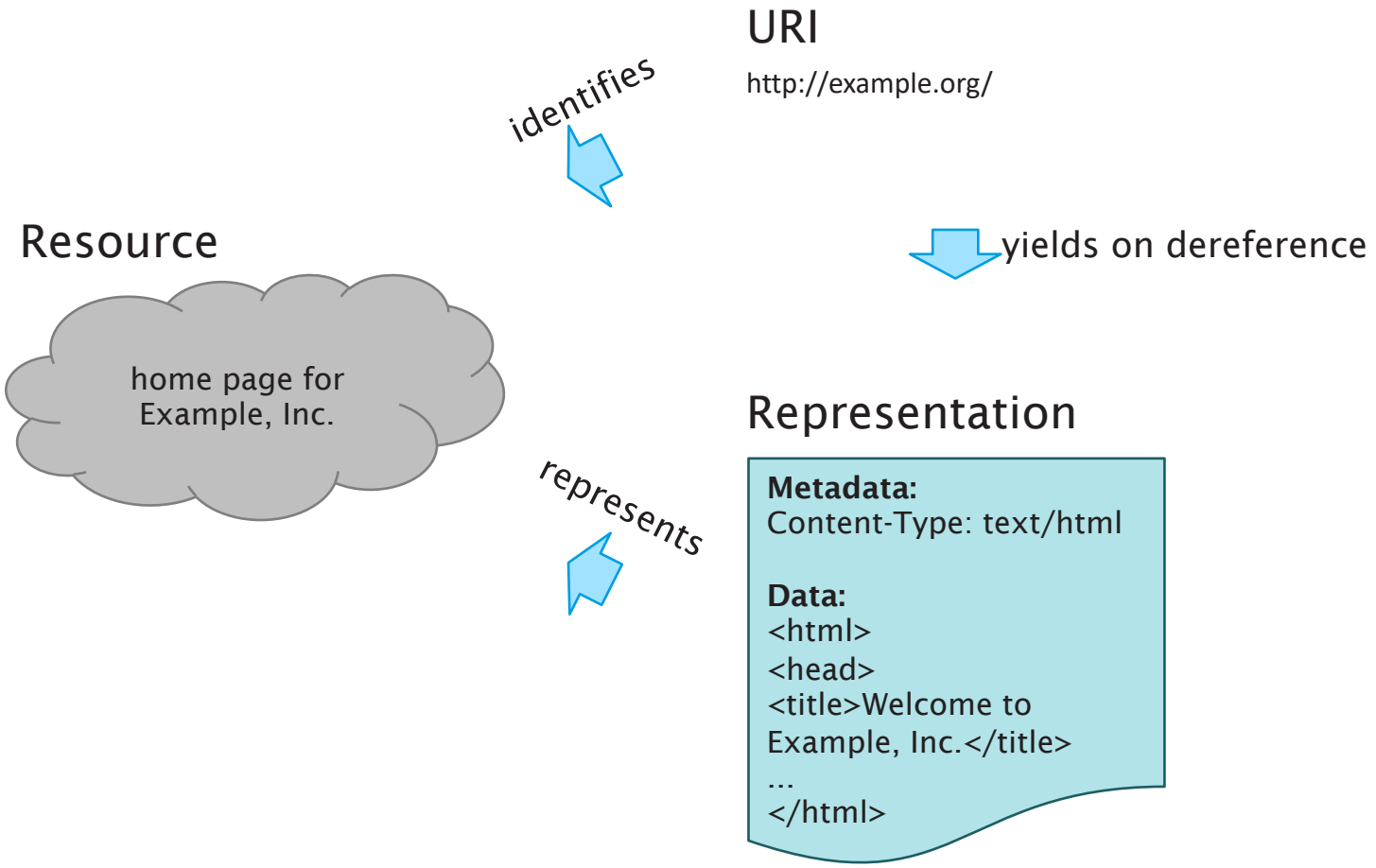
Architecture of the World Wide Web

The Web architecture has four main components:

- Resources (webpages, etc)
- Identifiers for resources (URIs)
- Protocols for interacting with resources (HTTP)
- Data formats for representing the state of resources (HTML, XML, etc)

The Semantic Web builds on this foundation

Architecture of the World Wide Web



http://www.w3.org/TR/webarch/

Uniform Resource Identifiers

What does a URI on the Semantic Web refer to?

- A real world object?
- A web page?
- Both?

- What does a URI identify in general?
- What is a resource?
- What are the implicit semantics in a URI?

University of Southampton

What is a resource?

From RFC2616 (HTTP/1.1):

“A network data object or service that can be identified by a URI [...] Resources may be available in multiple representations (e.g. multiple languages, data formats, size, and resolutions) or vary in other ways.”

http://www.ietf.org/rfc/rfc2616

University of Southampton

What is a resource?

From RFC2396 (URIs):

“A resource can be anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., "today's weather report for Los Angeles"), and a collection of other resources. Not all resources are network "retrievable"; e.g., human beings, corporations, and bound books in a library can also be considered resources.

The resource is the conceptual mapping to an entity or set of entities, not necessarily the entity which corresponds to that mapping at any particular instance in time. Thus, a resource can remain constant even when its content - the entities to which it currently corresponds - changes over time, provided that the conceptual mapping is not changed in the process.”

http://www.ietf.org/rfc/rfc2396.txt

University of Southampton

httpRange-14

W3C Technical Architecture Group issue

- “What is the range of the HTTP range dereference operation?”
- Raised in March 2002
- Closed in June 2005

TBL’s original stance: HTTP URIs (without "#") should be understood as referring to documents, not physical things

http://www.w3.org/2001/tag/group/track/issues/14

University of Southampton

All resources are equal…

…but some are more equal than others

The things identified by URIs are resources

Some resources can be retrieved by dereferencing their URIs

- Or rather, representations of some resources can be retrieved

Some resources cannot be retrieved

- People, cats, cars

Information Resources

"Information resources are resources, identified by URIs and whose essential characteristics can be conveyed in a message"

- An (abstract) document (with a URI) can be dereferenced to get an 'obvious' representation of that document
- The majority of current Web resources are information resources

University of Southampton

What makes an information resource?

Consider the case of resources identified by HTTP URIs:

- If dereferencing the URI results in a 200 OK response code, the resource is an information resource
  - From the HTTP RFC: “an entity corresponding to the requested resource is sent in the response”
- If it results in a 303 See Other response, the resource could be any resource
  - “the response to the request can be found under a different URI and SHOULD be retrieved using a GET method on that resource”
- If it results in a 4xx (client error) or 5xx (server error) response, we can’t say either way

University of Southampton

The Linked Data Principles

University of Southampton

Linked Data Principles

Set of publishing practices for SW data:

1. Use URIs as names for 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

Effectively, putting the hypertext back into the Semantic Web

Simplifies integration between datasets while maintaining loose coupling

1. Use URIs as names for things

Use a unique identifier to denote things:

- Hegel, Georg Wilhelm Friedrich
  - http://dbpedia.org/resource/Georg_Wilhelm_Friedrich_Hegel
  - http://viaf.org/viaf/89774942
  - …

- Hegel, Georg Wilhelm Friedrich: Gesammelte Werke / Vorlesungen über die Logik
  - urn:isbn:978-3-7873-1964-0

University of Southampton

2. Use HTTP URIs

- Enables “lookup” of URIs via Hypertext Transfer Protocol
- Piggy-backs on hierarchical Domain Name System to guarantee uniqueness of identifiers
- Uses established infrastructure
- Connects logical level (thing) with physical level (source)
- Important distinction between name/“thing URI” and location/“source URI”
  - Also called “other resource“/“non-information resource“ vs. “information resource“
  - See also httpRange-14

University of Southampton

3. Provide useful information

When somebody looks up a URI, return data using the standards (RDF*, SPARQL)

University of Southampton

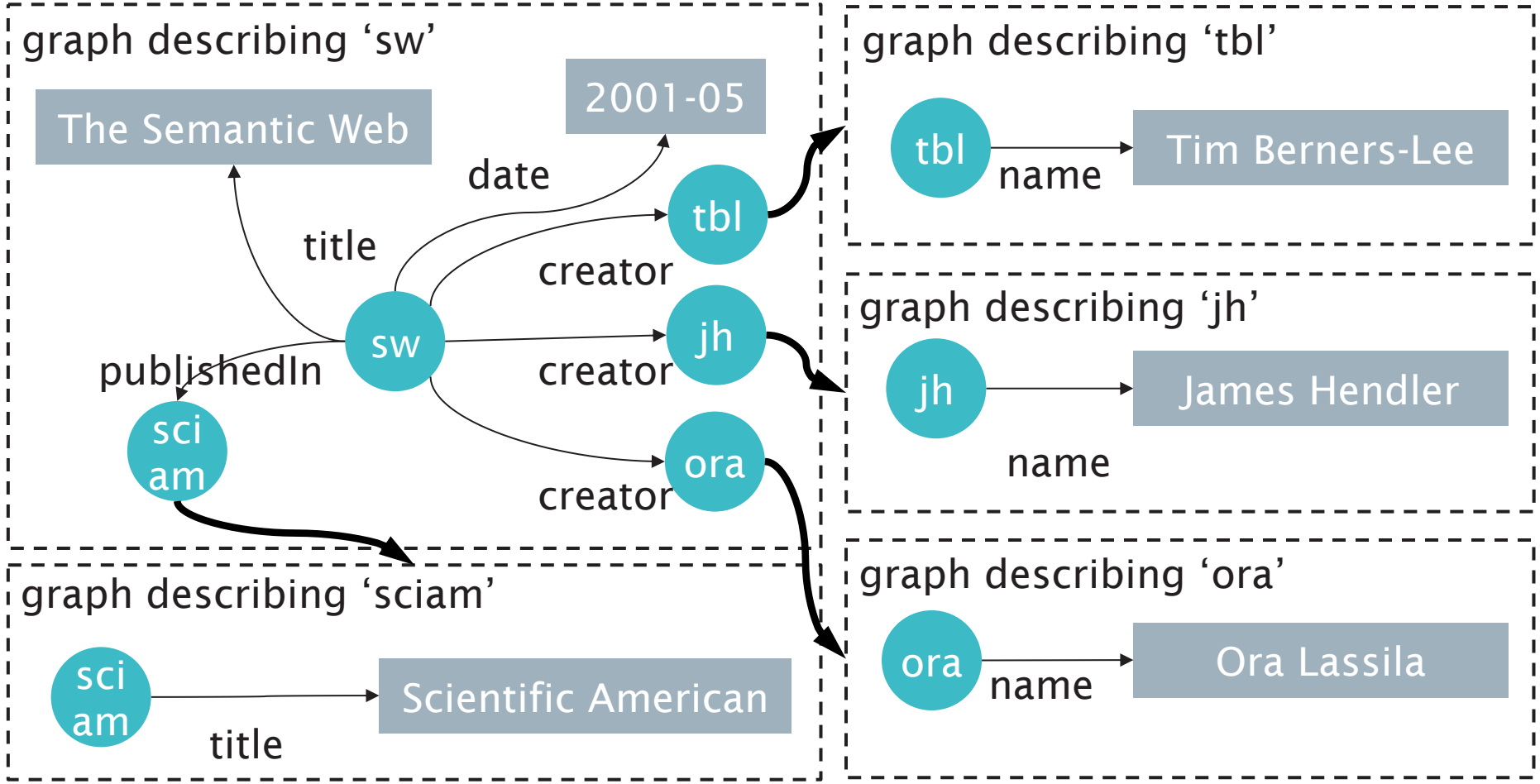
4. Link to other URIs

- Enable people (and machines) to jump from server to server
- External links vs. internal links (for any predicate)

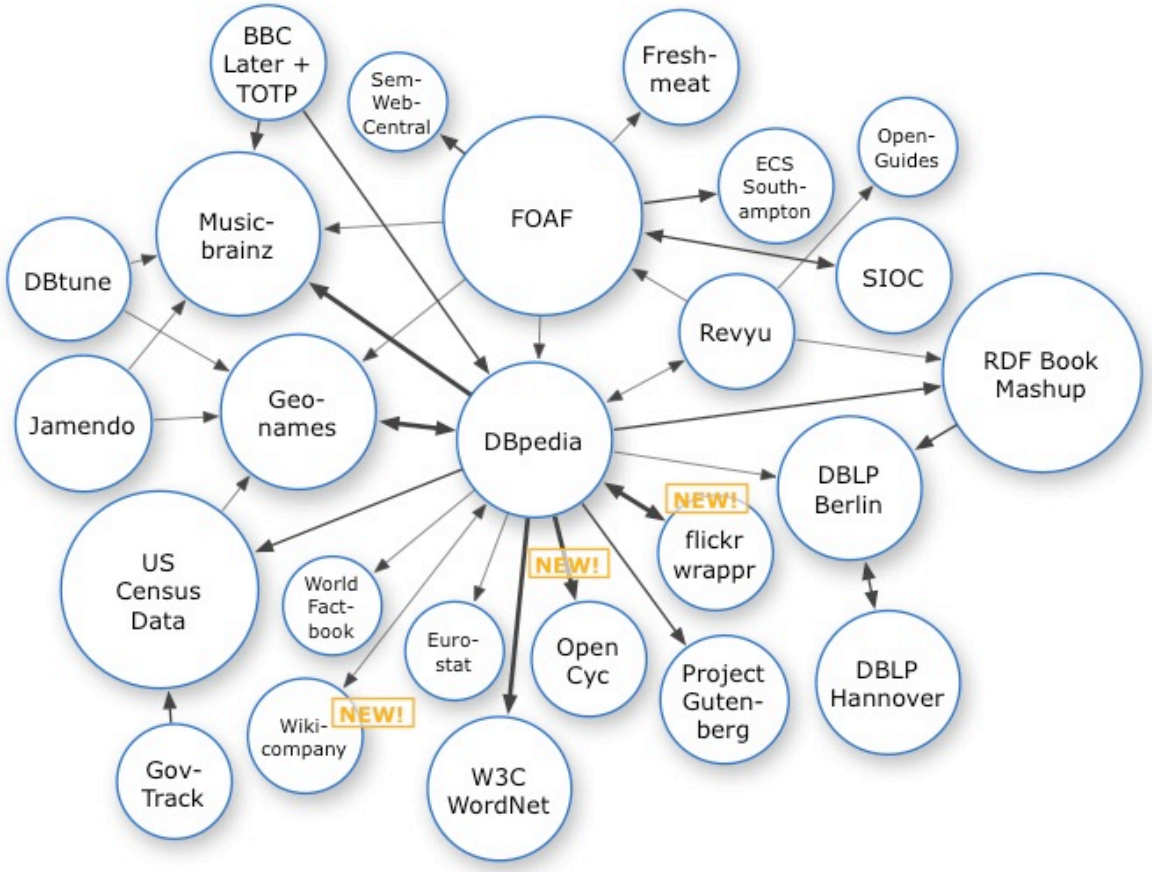
- Using external vocabularies enables linking
- Vocabularies might be interlinked, too

- Special owl:sameAs links to denote equivalence of identifiers (useful for data merging)
- Other types of links are possible as well

Example

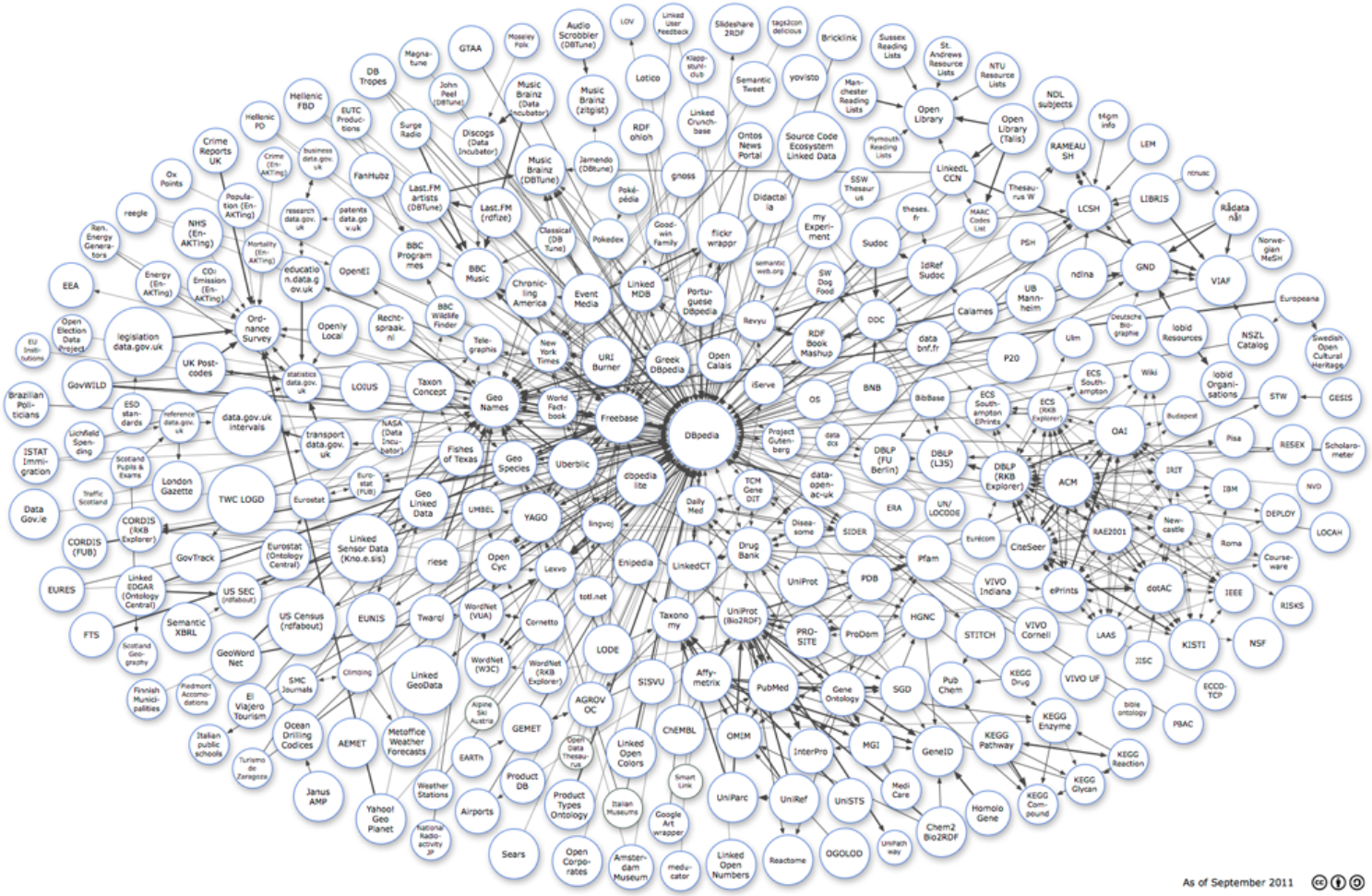

graph describing ‘sw’
The Semantic Web
2001-05
date
title
tbl
creator
sw
jh
creator
publishedIn
sci
am
ora
creator
graph describing ‘sciam’
sci
am
Scientific American
title
graph describing ‘tbl’
tbl
name
Tim Berners-Lee
graph describing ‘jh’
jh
James Hendler
name
graph describing ‘ora’
ora
name
Ora Lassila

Linked Data on the Web: 2007



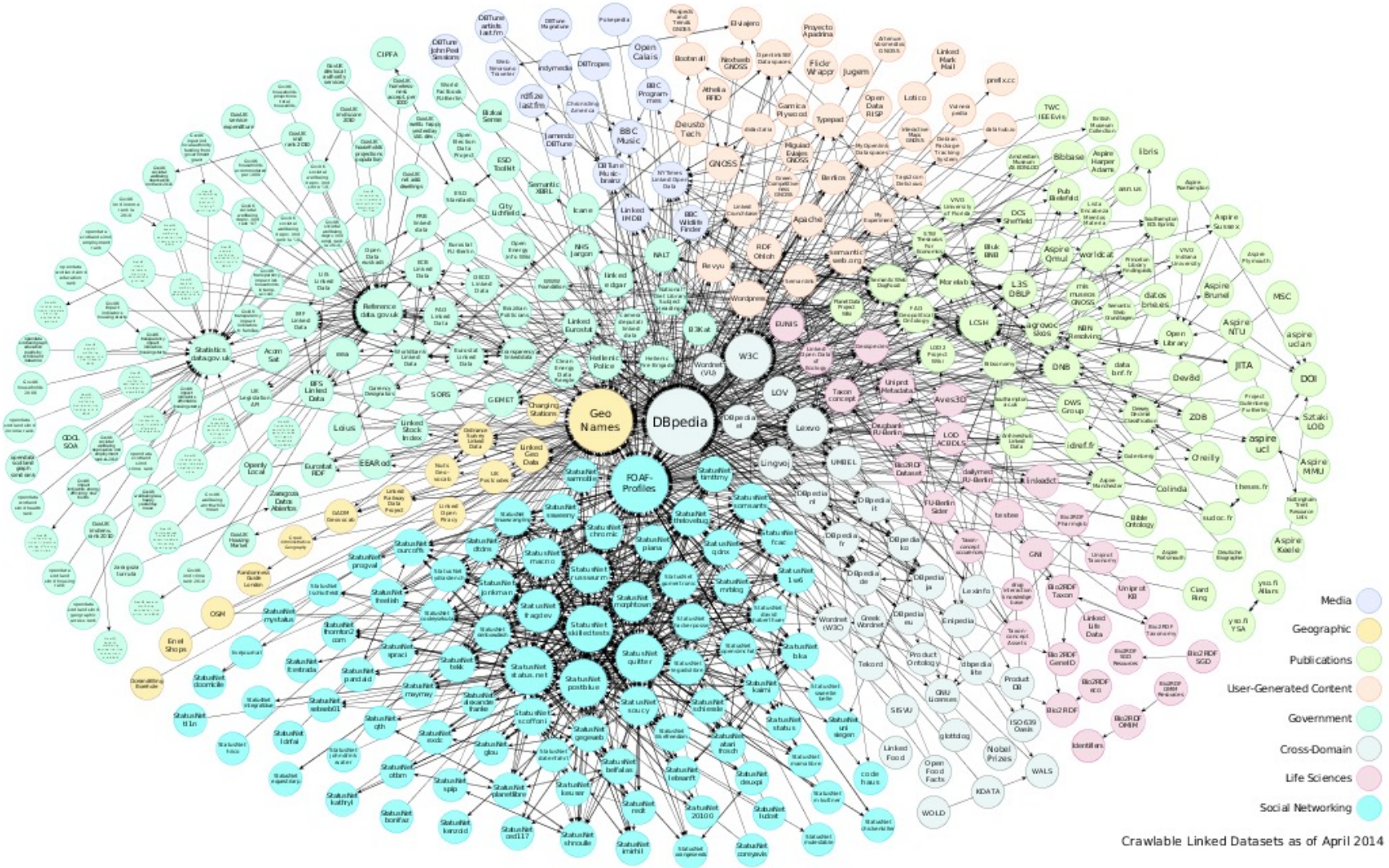
University of Southampton

2011



University of Southampton

2014



http://lod-cloud.net/state/state_2014/

Analysis of the LOD Cloud: 2014

Datasets by topical domain.

Topic	Datasets	%
Government	183	18.05%
Publications	96	9.47%
Life sciences	83	8.19%
User-generated content	48	4.73%
Cross-domain	41	4.04%
Media	22	2.17%
Geographic	21	2.07%
Social web	520	[illegible]%
Total		1014

Category	Vocabulary	Usage	Category	Vocabulary	Usage
social web	foaf	86.12%	life sciences	dct	66.29%
	dct	40.65%		foaf	41.57%
	wgs84	36.99%		void	31.46%
publications	dct	81.73%	government	dct	63.98%
	foaf	69.23%		cube	60.75%
	bibo	41.34%		odc*	46.24%
user-generated content	dct	81.91%	geographic	dct	82.93%
	foaf	74.55%		foaf	65.85%
	sioc	43.63%		skos	48.78%
media	foaf	75.67%	crossdomain	dct	72.73%
	dct	54.05%		foaf	72.73%
	mo	18.91%		skos	38.63%

http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/

Interlinking in the LOD Cloud 2014

Categorization by number of linked datasets

Number of linked datasets	Number of datasets
more than 10	79 (7.79%)
6 to 10	81 (7.99%)
5	31 (3.06%)
4	
3	
2	
1	
0	

Datasets with the ten highest indegrees

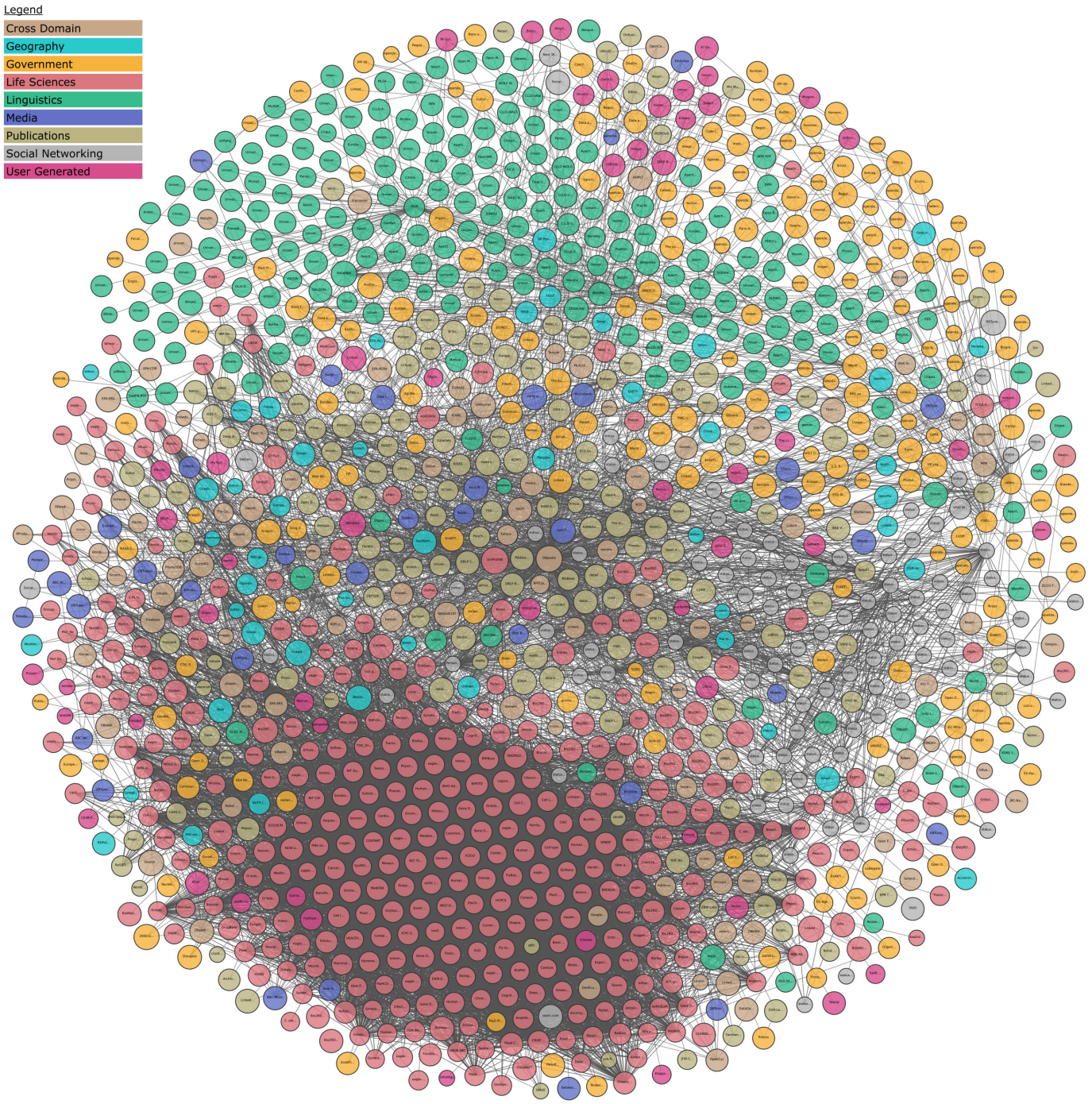
Dataset	Category	Indegree
dbpedia.org	cross-domain	207
geonames.org	geographic	141
w3.org	cross-domain	117
quitter.se	social web	64
status.net	social web	63
postblue.info	social web	56
skilledtests.com	social web	55
reference.data.gov.uk	government	45
data.semanticweb.org	publications	44
fragdev.com	social web	41
lexvo.org	cross-domain	37

Datasets with the ten highest outdegrees

Dataset	Category	Outdegree
bibsonomy.org	publications	91
semanlink.net	user-generated content	88
deri.org	social web	71
harth.org	social web	68
quitter.se	social web	67
semanticweb.org	user-generated content	64
skilledtests.com	social web	60
postblue.info	social web	59
status.net	social web	47
w3.org	crossdomain	45
data.semanticweb.org	publications	45

http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/

2020



http://lod-cloud.net/

Publishing Semantic Web Data

http://www.flickr.com/photos/cibergaita/97220057/lightbox/

Creating Semantic Web resources

In `http://example.org/data.rdf` :

```
@prefix foaf: <http://xmlns.com/foaf/0.1/>
<#fred> <foaf:name> “Fred Smith”.
```

We have a new resource: `http://example.org/data.rdf#fred`

University of Southampton

Publishing RDF Vocabularies

SW Best Practice Recipes for Publishing RDF Vocabularies

Distinguishes between ‘hash’ and ‘slash’ namespaces

- `http://example.org/ontology#foo`
- `http://example.org/ontology/foo`

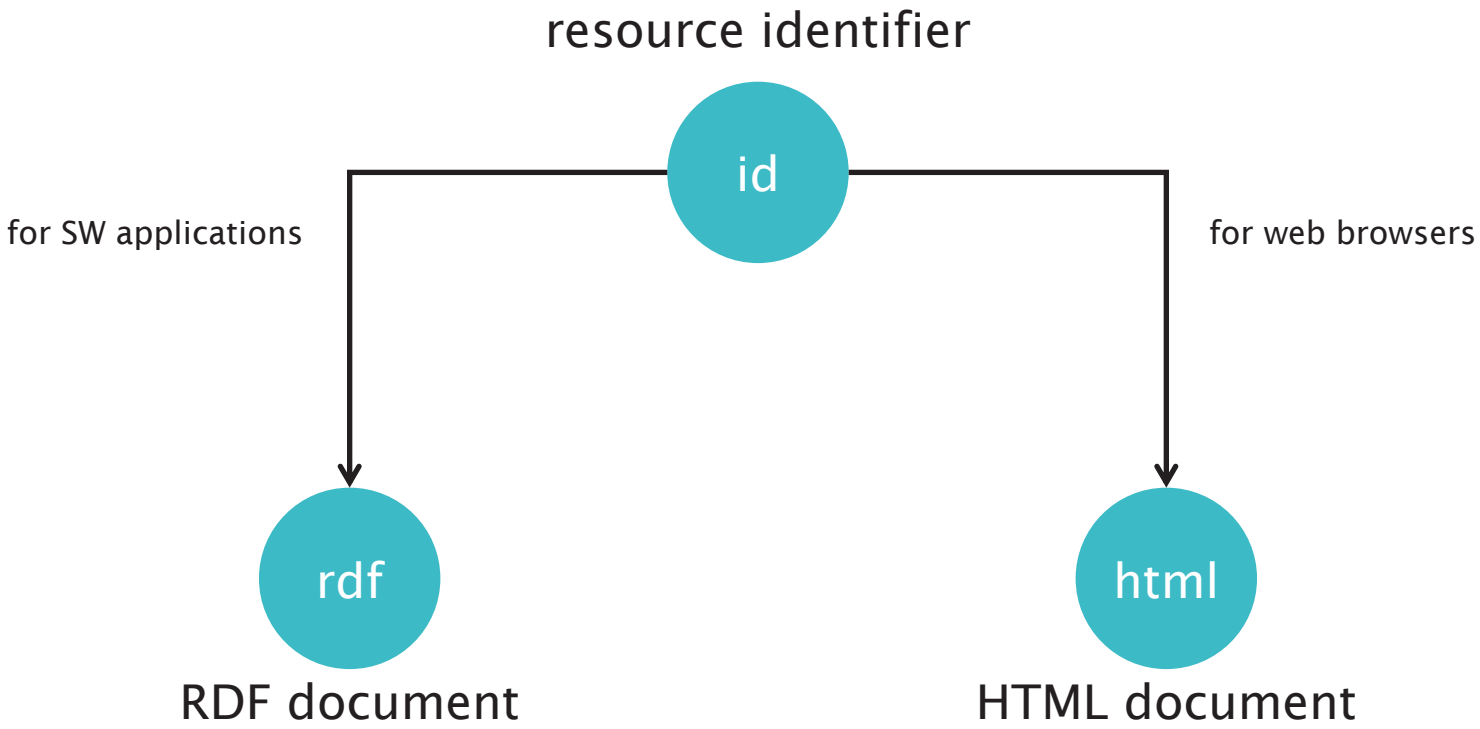
Uses content negotiation (HTTP Accept: header) to serve different representations of resources

- Machine-readable RDF vs human-readable HTML

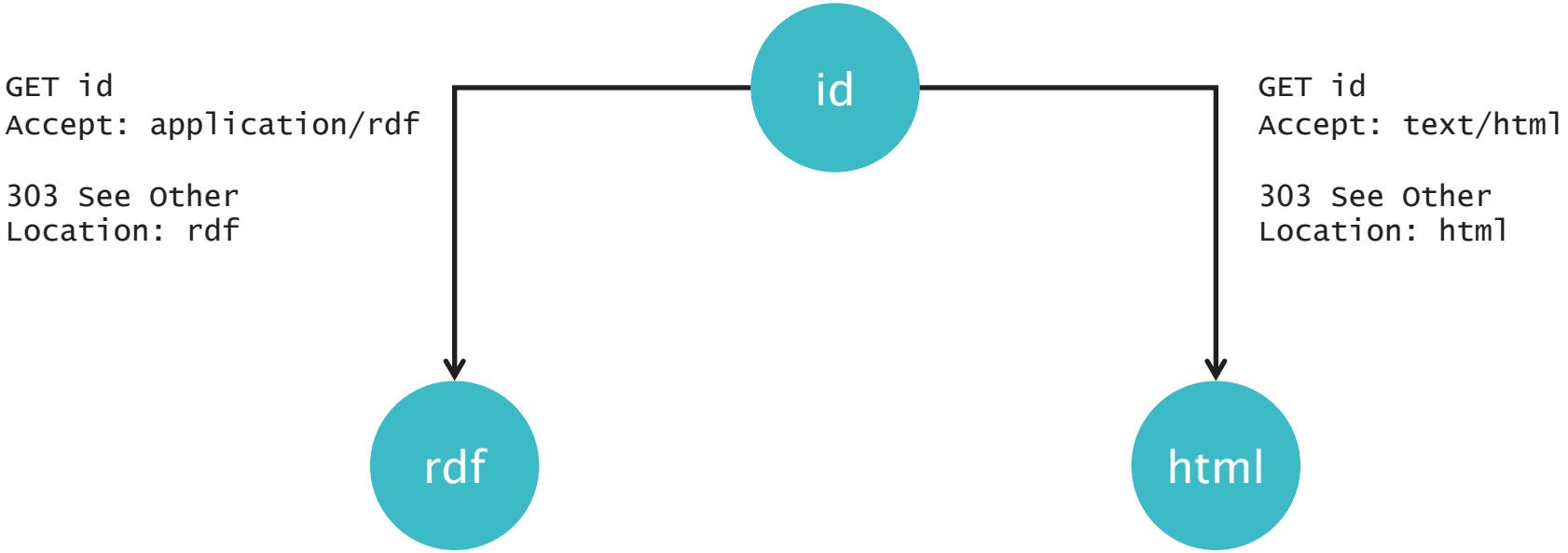
https://www.w3.org/TR/swbp-vocab-pub/

University of Southampton

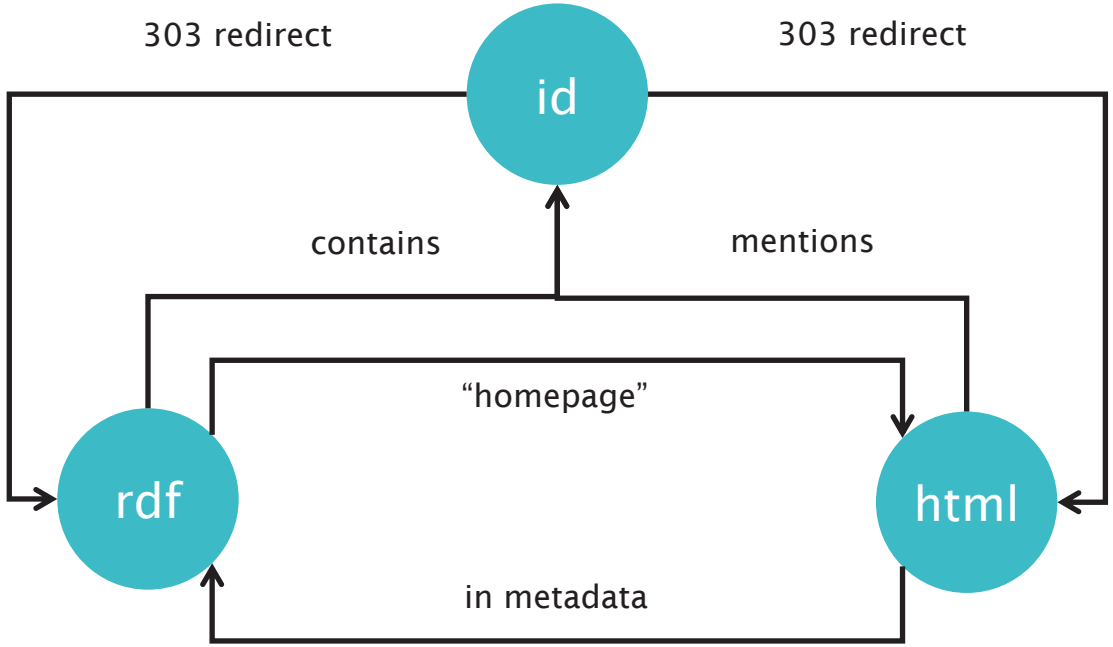
Cool URIs for the Semantic Web



Cool URIs – 303 Redirects



Cool URIs


303 redirect
id
303 redirect
contains
mentions
“homepage”
rdf
html
in metadata

Open Data Service beta

UNIVERSITY OF Southampton

#opensoton

University of Southampton

- Open Data Homepage
- 5★ Data
- Frequently Asked Questions
- Apps
- Data Catalogue
  - Places
  - Phonebook
  - Academic Programmes
  - Organisation
  - Jargon
  - Products & Services
- SPARQL Endpoint
- Feedback
  - Suggestions
  - Report a Problem
  - Register an App
- Credits

University of Southampton > Open Data

University of Southampton Open Data

The University of Southampton provides open access to some of our administrative data.

We believe that this will be of benefit to our own members and visitors, and increase the transparency of our operations.

Featured App:

Open Data Map

This tool is under development by Postgraduate students, it's a work in progress and so may break now and then, but it looks amazing. You can search for services on and near our campuses!

Linked Open Data

The executive summary: There's data we have which isn't in any way confidential which is of use to our members, visitors, and the public. If we make the data available in a structured way with a license which allows reuse then our members, or anyone else, can build tools on top of it without needless bureaucracy. That's common sense. We call it "Open Data".

For more on Open Data and it's benefits see these presentations by Southampton's Nigel Shadbolt and Tim Berners-Lee. They helped establish data.gov.uk the UK Government's Open Data site and are members of the Coalition Government's Transparency Board.

We publish our data in RDF format and link our identifiers to other sites in the Linked Open Data Web. This makes it much eaiser to merge data from multiple sources and other sites can link their datasets up with ours. Like the HTML Web, the whole is much greater than the sum of its parts, that's "Linked Data".

Show me the data!

Browse the list of datasets or view the links on the left to explore some of our data.

Southampton Data Blog
2011-04-14, 16:02h

Interview with Christopher Gutteridge
There's an interview with Christopher Gutteridge (me!) on this weeks Ubuntu UK Podcast. (If you're wondering, data.soutampton.ac.uk runs on virtual machine running Ubuntu) Actually, it's worth giving a shout out to the technologies we use, but I'll save that for a future post.

[April 1st Gag] PDF selected as Interchange Format
The following article is our prank for April 1st. Just to be clear PDF is a dreadful format to exchange data in. It was inspired, in part, by The Register wesbsite running the following picture and quote. Yes, I did say that, but I was talking about research and data communication. It was fun working [...]

New Formats
New ways to enjoy our data. We've added some links to the "Get the Data" box which let you see what formats are available. Some pages let you download RDF, others you can get back as tabular data, suitable for loading into Excel, amongst other things. Roughly speaking, pages about things have RDF versions, pages [...]

Grasping the nettle and changing some URIs
We've realised that using UPPER CASE in some URIs looked fine in a

It’s not quite that simple...

University of Southampton

rdfURIMeaning-39

W3C Technical Architecture Group issue

- Raised in July 2003
- Currently open

Is a given inference engine expected to take into account a given document under given circumstances?

How does one avoid having to commit to things one does not trust?

http://www.w3.org/2001/tag/group/track/issues/39

University of Southampton

HttpRedirections-57

W3C Technical Architecture Group issue

- “Mechanisms for obtaining information about the meaning of a given URI”
- Raised in July 2007
- Currently open

Further consideration of the use of:

- 303 HTTP status codes (and interaction with caching)
- Other possible mechanisms for obtaining a description of a (non-information) resource (HTTP Link: header – see RFC2068)

http://www.w3.org/2001/tag/group/track/issues/57

University of Southampton

UniformAccessToMetadata-62

W3C Technical Architecture Group issue

- “Given the URI of an HTTP-accessible information resource R, how can an agent learn the URIs of metadata documents about R authorized by the owner of the original URI”
- Raised in March 2009
- Currently open

http://www.w3.org/2001/tag/group/track/issues/62

University of Southampton

Further Reading

Architecture of the World Wide Web
http://www.w3.org/TR/webarch/

R.T. Fielding and R.N. Taylor, Principled Design of the Modern Web Architecture, ACM Transactions on Internet Technology 2 (2): 115–150
http://www.ics.uci.edu/~taylor/documents/2002-REST-TOIT.pdf

Uniform Resource Identifiers (URI): Generic Syntax
IETF RFC2396
http://www.ietf.org/rfc/rfc2396.txt

Hypertext Transfer Protocol - HTTP/1.1
IETF RFC2616
http://www.ietf.org/rfc/rfc2616

University of Southampton

Further Reading

What do HTTP URIs identify?
http://www.w3.org/DesignIssues/HTTP-URI

W3C TAG issue httpRange-14
http://www.w3.org/2001/tag/group/track/issues/14

W3C TAG Issue rdfUriMeaning-39
http://www.w3.org/2001/tag/group/track/issues/39

W3C TAG issue httpRedirections-57
http://www.w3.org/2001/tag/group/track/issues/57

W3C TAG issue UniformAccessToMetadata-62
http://www.w3.org/2001/tag/group/track/issues/62

Dereferencing HTTP URIs
http://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14

Further Reading

Cool URIs for the Semantic Web
https://www.w3.org/TR/cooluris/

Best Practice Recipes for Publishing RDF Vocabularies
http://www.w3.org/TR/swbp-vocab-pub/

Data on the Web Best Practices
https://www.w3.org/TR/dwbp/

University of Southampton

Next Lecture: SPARQL