Semantic Web Publishing

Dr Nicholas Gibbins
nmg@ecs.soton.ac.uk
32/4037
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?

- What identifiers do we use?
Semantic Web Principles

• Anyone can make assertions about anything
• Entities are referred to using Uniform Resource Identifiers
• Based on XML technologies
• Formal semantics
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?
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.
URIs, URLs and URNs

• Classical view – early to mid 1990s
  • Uniform Resource Locators specify the location of a resource (machine name, etc)
    • http:
  • Uniform Resource Names specify the name of a resource, independent of its name
    • isbn:

• Uniform Resource Identifiers are either URLs or URNs
URIs, URLs and URNs

• URL resolution is (usually) well-defined

• URNs don’t necessarily have well-defined resolution semantics
  • Resolving names depends on context
  • What does resolution mean for URIs which do not refer to network resources?
Representational State Transfer (REST)

- Architectural principle for Web systems
  - Resources referred to by global identifiers (URIs)
  - Manipulated via a standard interface (http)
  - Network components (clients and servers) exchange representations of the resources
  - Connectors (caches, proxies) mediate requests
httpRange-14

- W3C Technical Architecture Group issue
  - “What is the range of the HTTP range dereference operation?”
  - Raised in March 2002
  - Closed in Jun 2005

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

graph describing ‘sw’

The Semantic Web

publishedIn

sci am

title

graph describing ‘tbl’

tbl

date

tbl

creator

tbl

name

Tim Berners-Lee

graph describing ‘jh’

jh

creator

jh

name

James Hendler

graph describing ‘ora’

ora

creator

ora

name

Ora Lassila

graph describing ‘sci am’

sci am

title

Scientific American
RDF Publishing Example

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
Defining 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
Minimal Hash Namespace
Minimal Slash Namespace

[Diagram showing a client making GET requests to a server, with responses of 200 OK and 303 See Other, indicating interaction with RDF resources.]
Extended Slash Namespace
Cool URIs

Resource identifier (URI)

ID

Semantic web applications

RDF

RDF document URL

Web browsers

HTML

HTML document URL
Cool URIs – 303 Pattern

http://www.acme.com/id/alice

ID

303 redirect

Accept: application/rdf+xml

RDF

http://www.acme.com/data/alice

HTML

http://www.acme.com/people/alice
Cool URIs – Hash Pattern

http://www.acme.com/about#alice

ID

Automatic truncation of fragment

RDF

http://www.acme.com/about
Cool URIs

Diagram:
- ID
  - uses RDF
  - mentions HTML
- RDF
- HTML
- homepage

Labels:
- 303
- uses
- mentions
- meta

School of Electronics and Computer Science
Cool URIs in ECS

ID URI: http://id.ecs.soton.ac.uk/person/1269
RDF URI: http://rdf.ecs.soton.ac.uk/person/1269
HTML URI: http://www.ecs.soton.ac.uk/people/nmg
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 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 easier 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.
It’s not quite that simple...
• 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?
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)
• **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
Further Reading

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


• Uniform Resource Identifiers (URI): Generic Syntax
  IETF RFC 2396
  http://www.ietf.org/rfc/rfc2396.txt

• Hypertext Transfer Protocol - HTTP/1.1
  IETF RFC 2616
  http://www.ietf.org/rfc/rfc2616
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
Further Reading

• Cool URIs for the Semantic Web  
  http://www.w3.org/TR/2007/WD-cooluris-20071217/

• Best Practice Recipes for Publishing RDF Vocabularies  
  http://www.w3.org/TR/swbp-vocab-pub/
Embedding Semantic Web Data
Embedded Data

- Publishing patterns for linked data that we’ve already considered make the assumption that we’re publishing directly in RDF/OWL
- What if the data already exists in a web resource, in some form?
- Republishing the data separately introduces redundancy, and the possibility of inconsistency - embed our SW data in the web resource
- Two main approaches:
  - GRDDL
  - RDFa
GRDDL

- Given a document in some XML format, how can we extract the relevant portions and make them available to SW agents?
- GRDDL (Gleaning Resource Descriptions from Dialects of Languages) uses XSLT stylesheets to transform documents.
Using GRDDL with XML

• In well-formed XML, link to XSLT transformation using grddl:transformation attribute

• Need to introduce grddl: namespace

```html
<html xmlns='http://www.w3.org/1999/xhtml'
     xmlns:grddl='http://www.w3.org/2003/g/data-view#
     grddl:transformation="glean_title.xsl">
  <head>
    <title>Are You Experienced?</title>
    [...]
  </head>
</html>
```
Using GRDDL with XHTML

• Link to XSLT transformation using link element

• Introduce GRDDL in profile

```xml
<html xmlns="http://www.w3.org/1999/xhtml">
    <head profile="http://www.w3.org/2003/g/data-view">
        <title>Some Document</title>
        <link rel="transformation" href="http://www.w3.org/2000/06/dc-extract/dc-extract.xsl" />
        <meta name="DC.Subject" content="ADAM; Simple Search; Index+; prototype" />
        [...
    </head>
    [...
</html>
```
• Yet another syntax for RDF...
• Designed for embedding structured data in web pages
• Stored structure in attributes (the ‘a’ in ‘RDFa’)
I'm holding one last summer Barbecue, on September 16th at 4pm.

You can contact me via email.
I'm holding one last summer Barbecue, on September 16th at 4pm.

Showing an instance of a class

_:a <rdf:type> <cal:Event> .
I'm holding one last summer Barbecue, on September 16th at 4pm.

I'm holding <span property="cal:summary">one last summer Barbecue</span>, on September 16th at 4pm.

_:a <cal:summary> “one last summer Barbecue” .
I'm holding one last summer Barbecue, on September 16th at 4pm.

Using properties:

I'm holding one last summer Barbecue, on September 16th at 4pm.

_:a <cal:start> "20070916T1600-0500" .
<p class="contactinfo">Jo Smith. Web hacker at <a href="http://example.org">Example.org</a>. You can contact me via email.</p>

<p class="contactinfo" about="http://example.org/staff/jo">Jo Smith. Web hacker at <a href="http://example.org">Example.org</a>. You can contact me via email.</p>
Identity

<p class="contactinfo" about="http://example.org/staff/jo">Jo Smith. […]

<p class="contactinfo" about="http://example.org/staff/jo"><span property="foaf:name">Jo Smith</span>. […]</p>

<http://example.org/staff/jo> <foaf:name> “Jo Smith” .
Using existing links

<p about="http://example.org/staff/jo"> [...] <a href="http://example.org">Example.org</a>. You can contact me <a href="mailto:jo@example.org">via email</a>.</p>

<p about="http://example.org/staff/jo"> [...] <a rel="foaf:homepage" href="http://example.org">Example.org</a>. You can contact me <a rel="foaf:mbox" href="mailto:jo@example.org">via email</a>.</p>

<http://example.org/staff/jo> <foaf:mbox> <mailto:jo@example.org> ; <foaf:homepage> <http://example.org/> .
Datatypes

<span property="dc:date" content="2007-05-12" datatype="xsd:date">May 12th, 2007</span>
Further Reading

• Gleaning Resource Descriptions from Dialects of Languages
  W3C Recommendation 11 September 2007
  http://www.w3.org/TR/grddl/

• RDFa in XHTML: Syntax and Processing
  W3C Recommendation 14 October 2008
  http://www.w3.org/TR/rdfa-syntax/

• RDFa Primer
  W3C Working Group Note 14 October 2008
  http://www.w3.org/TR/xhtml-rdfa-primer/