An Introduction to Linked Data

Dr Tom Heath
Platform Division
Talis Information Ltd

tom.heath@talis.com
http://tomheath.com/id/me

13/14 February 2009
Austin, Texas
Objectives

• Introduce the concept, principles, and key features of Linked Data
• Provide hands-on experience of creating Linked Data
• Provide a broad technical understanding of how to publish Linked Data
• Highlight some of the tools available for publishing and consuming Linked Data
Schedule

- 09:00 Welcome and Introductions
- 09:15 Linked Data, What and Why?
- 10:00 The Web of Data
- 10:30 Coffee
- 11:00 Linked Data Hands-on
- 12:30 Lunch
- 13:30 How to Publish Linked Data on the Web
- 15:00 Coffee
- 15:30 Current Linked Data Applications
- 16:00 Linked Data Toolbox
- 16:30 Discussion
- 17:00 Close
Sponsors

- Organisation
- My time
- Coffee and lunch
- Tonight's beer
Linked Data, What and Why?
The Web of Documents

- Analogy
  - a global **filesystem**
- Designed for
  - **human** consumption
- Primary objects
  - **documents**
- Links between
  - **documents** (or sub-parts of)
- Degree of structure in objects
  - fairly **low**
- Semantics of content and links
  - **implicit**
The Web of Linked Documents

[Diagram showing HTML documents linked with untyped links.]
The Web of Documents: Issues

• Simplicity
  • Loosely structured data, untyped links, disconnected data

• Integration
  • Show me all the publications by publicly-funded PhD students

• Querying
  • Which papers have I written with people from European institutions outside the UK?
Data Silos on the Web
Data Silos on the Web

- A
- B
- C
- D
The Web of Linked Data

- Analogy
  - a global database
- Designed for
  - machines first, humans later
- Primary objects
  - things (or descriptions of things)
- Links between
  - things
- Degree of structure in (descriptions of) things
  - high
- Semantics of content and links
  - explicit
The Web of Linked Data

Don't just link the *documents*, link the *things*
Linked Data is...

- ...a way of publishing data on the Web that:
  - encourages reuse
  - reduces redundancy
  - maximises its (real and potential) inter-connectedness
  - enables network effects to add value to data
Linked Data Technology Stack

- URIs
- HTTP
- RDF
- (RDFS/OWL)
URIs – Not Just for Web Pages

• “A Uniform Resource Identifier (URI) provides a simple and extensible means for identifying a resource.” -- RFC 3986

• Many different schemes: http://, ftp://, tel:, urn:, mailto:

• Some URIs for “real world” things:
  – http://tomheath.com/id/me
  – http://dbpedia.org/resource/Talis_Group
  – http://sws.geonames.org/4671654/
HTTP

- Data access mechanism
- Using http:// URIs to identify things allows people to look these things up
RDF: Resource Description Framework

- Data format for describing things and their interrelations
The RDF Data Model

- **Triples**

  subject $\rightarrow$ predicate $\rightarrow$ object

  $Tom \rightarrow \text{worksFor} \rightarrow Talis$

  $Talis \rightarrow \text{basedIn} \rightarrow Birmingham$

  $<\text{uri}> \rightarrow <\text{uri}> \rightarrow <\text{uri}>$ or "literal"
“Talis is Based Near Birmingham”

<http://dbpedia.org/resource/Talis_Group>

<http://xmlns.com/foaf/0.1/based_near>

<http://sws.geonames.org/3333125/>
Data Merging with RDF

- Mix schemas/vocabularies within one document
- Less painful data merging
Data Merging with RDF

Prefixes
rc: <http://richard.cyganiak.de/foaf.rdf#>
rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
foaf: <http://xmlns.com/foaf/0.1/>
dbpedia: <http://dbpedia.org/resource/>
dp: <http://dbpedia.org/property/>
skos: <http://www.w3.org/2004/02/skos/core#>
Data Merging with RDF

Prefixes
rc: <http://richard.cyganiak.de/foaf.rdf#>
rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
foaf: <http://xmlns.com/foaf/0.1/> 
dbpedia: <http://dbpedia.org/resource/>
dp: <http://dbpedia.org/property/>
skos: <http://www.w3.org/2004/02/skos/core#>
This is Linked Data
Linked Data Principles

• Use URIs as names for things
  – anything, not just documents
  – you are not your homepage
  – information resources and non-information resources

• Use HTTP URIs
  – globally unique names, distributed ownership
  – allows people to look up those names

• Provide useful information in RDF
  – when someone looks up a URI

• Include RDF links to other URIs
  – to enable discovery of related information
Why Publish Linked Data?

• Ease of discovery
• Ease of consumption
  – standards-based data sharing
• Reduced redundancy
• Added value
  – build ecosystems around your data/content
The Web of Data
The Linking Open Data Project
The Linking Open Data Project

- Community project with W3C support
- Take existing open data sets
- Make them available on the Web in RDF
- Interlink them with other data sets
- Began early 2007
Participants

- Massachusetts Institute of Technology (US)
- University of Southampton (UK)
- Freie Universität Berlin (DE)
- DERI (IE)
- KMi, Open University (UK)
- University of London (UK)
- Universität Hannover (DE)
- University of Pennsylvania (US)
- Universität Leipzig (DE)
- Universität Karlsruhe (DE)
- Joanneum (AT)
- University of Toronto (CA)

- BBC (UK)
- Talis (UK)
- Garlik (UK)
- OpenLink (UK)
- Thomson Reuters (US)
- Zitgist (US)
- Mondeca (FR)
- Cyc Foundation (US)
The LOD "Cloud" - May 2007

Shared innovation
Birmingham

From Wikipedia, the free encyclopedia

This article is about the city in England. For the U.S. city named after it, see Birmingham, Alabama. For other uses, see Birmingham (disambiguation).

Birmingham (ˌbɜrmɪŋəm) is a city and metropolitan borough in the West Midlands county of England. Birmingham is the most populous of England's core cities, and is the second most populated British city after London, with a population of 1,030,200 (2005 estimate). Often considered to be the second city of the United Kingdom,[3] the City of Birmingham forms part of the larger West Midlands conurbation, which has a population of 2,280,093 (2001 census) and includes several neighbouring towns and cities, such as Solihull, Wolverhampton and the towns of the Black Country.

The city's reputation was forged as a powerhouse of the Industrial Revolution in England, a fact which led to Birmingham being known as "the workshop of the world" or the "city of a thousand trades".[4] Although Birmingham's industrial importance has declined, it has developed into a national commercial centre, being named as the third best place in the United Kingdom to locate a business, and the 21st best in Europe by Cushman & Wakefield in 2007.[5] It is also the fourth most visited city by foreign visitors in the UK.[6] In 1999, Birmingham hosted the G8 summit at the International Convention Centre, on the site of Bingley Hall, the world's first purpose-built exhibition hall, and remains a popular location for conventions today along with the National Exhibition Centre in nearby Solihull.[7] In 2007, Birmingham was ranked as the 55th most livable city in the world, according to the Mercer index of worldwide standards of living.[8]

People from Birmingham are known as 'Brummies', a term derived from the city's nickname of Brum. This comes in turn from the city's dialect name, Brummagem,[9] which is derived from one of the city's earlier names, Bromwich. There is a distinctive Brummie dialect and accent, both of which differ from the adjacent Black Country.

Contents

- History
- Geography
  - 2.1 Geology
  - 2.2 Climate
- 2.3 Nearby places
- Demographics
- Places of interest
- Economy
- Politics
- Transport
- Education
- Sport
- Food & drink
- Culture and arts
  - 11.1 Arts
  - 11.2 Festivals and shows
  - 11.3 Media
  - 11.4 Leisure
  - 11.5 Architecture
- Environment
- Crime and policing
- Second City
- Notable residents
- Science and innovation

Motto: Forward
The LOD "Cloud" - July 2007
The LOD "Cloud" - August 2007
The LOD "Cloud" - November 2007
The LOD "Cloud" – Feb 2008

[Diagram showing various entities connected by lines and arrows, representing a network of shared innovation and data sources.]

- Musicbrainz
- Audio-Scrobbler
- Freshmeat
- Flickr exporter
- SIOC profiles
- FOAF profiles
- ECS Southampton
- Sem-Web-Central
- BBC Later + TOTP
- BBC John Peel
- Magnatune
- Jamendo
- GeoNames
- DBpedia
- US Census Data
- World Factbook
- OpenCyc
- W3C WordNet
- Gov-Track
- Wiki-company
- Eurostat
- OpenGuides
- RDF Book Mashup
- DBLP Berlin
- DBLP Hannover
- SW Conference Corpus
- SW World
- Revyu
- lingvoj
- Project Gutenberg
- flickr wrappr
- OpenGuides
- Gov-Track
- US Census Data
The LOD "Cloud" – Sept 2008
Linked Data Hands-On: Pimp Your FOAF
FOAF: Friend of a Friend

• An RDF vocabulary for describing people:
  – identities
  – interests
  – affiliations
  – social networks
  – etc
Pimp Your FOAF

- Hands-on Exercise
- Create a basic FOAF file
- Enhance it with Linked Data
- Prizes for the best pimping
  - number of links, accuracy, diversity...
Pimp Your FOAF: Instructions

1. Create yourself a FOAF file
   - http://www.ldodds.com/foaf/foaf-a-matic
   - name your file yourname.rdf

2. Upload it
   - ftp://playground.linkeddata.org
   - user: **********, pass: **********
   - validate it: http://www.w3.org/RDF/Validator/

3. Explore the cloud

4. Create as much Linked Data in your FOAF as you can
   - Look for predicates:
     - http://xmlns.com/foaf/spec/
     - http://schemaweb.info/

5. Browse it using e.g. Marbles
   - http://beckr.org/marbles
Pimp Your FOAF: Prizes!
How to Publish Linked Data on the Web
Scenario

- Online whisky shop: Wiskii.com
- New business venture, founded by Jeff
- For the whisky connoisseur
- Detailed background information from experts
- Contributions from customers
- Custom web app, relational backend
- Simultaneous publication in HTML and RDF
5 Steps to Publishing Linked Data

1. Understand the Principles
2. Understand your Data
3. Choose URIs for Things in your Data
4. Setup Your Infrastructure
5. Link to other Data Sets
1. Understand the Principles
Linked Data Principles: Redux

- Use URIs as names for things
  - anything, not just documents
  - you are not your homepage
  - information resources and non-information resources

- Use HTTP URIs
  - globally unique names, distributed ownership
  - allows people to look up those names

- Provide useful information in RDF
  - when someone looks up a URI

- Include RDF links to other URIs
  - to enable discovery of related information
2. Understand your Data
2. Understand Your Data

- What are the key things present in your data?
  - People?
  - Places?
  - Books?
  - Films?
  - Musicians?
  - Concepts?
  - Photos?
  - Comments?
  - Reviews?
  - ...

shared innovation
2. Understand Your Data

• Things in the Wiskii.com database
  – Distilleries
  – Regions and Locations
  – Founders
  – Owners
  – Brands
  – Products
  – Photos
  – Reviews
  – Comments
  – Prices/Offers
2. Understand Your Data

• What vocabularies can be used to describe these?
  – Principles
    • Reuse, don't reinvent
    • Mix liberally
  – Potential Ontologies/Vocabularies
    • Geo
    • GoodRelations
    • FOAF
    • Review
    • SIOC
    • Whisky
3. Choose URIs for Things in Your Data
3. Choosing URIs: Principles

• Use HTTP URIs

• Keep out of other peoples' namespaces

• Abstract away from implementation details

• Hash or Slash
  1. http://mydomain.com/foaf.rdf#me
  2. http://mydomain.com/id/me
3. Choosing URIs: Common Patterns

- http://dbpedia.org/resource/New_York_City ← Thing
- http://dbpedia.org/data/New_York_City ← RDF data
- http://dbpedia.org/page/New_York_City ← HTML page
- http://revyu.com/people/tom ← Thing
- http://revyu.com/people/tom/about/rdf ← RDF data
- http://revyu.com/people/tom/about/html ← HTML page
- http://kmi.open.ac.uk/people/tom/ ← Thing
- http://kmi.open.ac.uk/people/tom/rdf ← RDF data
- http://kmi.open.ac.uk/people/tom/html ← HTML page
- http://mydomain.com/thing ← Thing
- http://mydomain.com/thing.rdf ← RDF data
- http://mydomain.com/thing.html ← HTML page
3. Choosing URIs: Wiskii.com

- http://wiskii.com/regions/speyside
- http://wiskii.com/distilleries/talisker
- http://wiskii.com/brands/talisker
- http://wiskii.com/people/william-matheson
- http://wiskii.com/photos/58
- http://wiskii.com/reviews/271
3. Choosing URIs: Wiskii.com

- http://wiskii.com/distilleries/talisker
- http://wiskii.com/distilleries/talisker/rdf

- http://wiskii.com/brands/talisker
- http://wiskii.com/brands/talisker/rdf

- http://wiskii.com/people/william-matheson
- http://wiskii.com/people/william-matheson/rdf

- http://wiskii.com/photos/58
4. Setup Your Infrastructure
4. Setup Your Infrastructure
4. Setup Your Infrastructure

http://wiskii.com/distilleries/talisker/html
http://wiskii.com/distilleries/talisker/rdf
4. Setup Your Infrastructure

http://wiskii.com/distilleries/talisker

http://wiskii.com/distilleries/talisker/html

http://wiskii.com/distilleries/talisker/rdf
4. Setup Your Infrastructure

HTTP GET

http://wiskii.com/distilleries/talisker

http://wiskii.com/distilleries/talisker/html

http://wiskii.com/distilleries/talisker/rdf
4. Setup Your Infrastructure

HTTP GET

http://wiskii.com/distilleries/talisker

HTML

PHP

DB

RDF

http://wiskii.com/distilleries/talisker/html

http://wiskii.com/distilleries/talisker/rdf
Content Negotiation

GET [vocabulary URI]
Accept: application/rdf+xml

Server

303 See Other
Location: [RDF content location]

GET [RDF content location]
Accept: application/rdf+xml

Server

200 OK
<RDF>
4. Setup Your Infrastructure

HTTP GET

http://wiskii.com/distilleries/talisker

HTTP 303 See Other

http://wiskii.com/distilleries/talisker/html

HTTP 303 See Other

http://wiskii.com/distilleries/talisker/rdf

HTTP GET

http://wiskii.com/distilleries/talisker

HTTP 303 See Other

http://wiskii.com/distilleries/talisker/html

HTTP 303 See Other

http://wiskii.com/distilleries/talisker/rdf

HTTP GET

http://wiskii.com/distilleries/talisker

HTTP 303 See Other

http://wiskii.com/distilleries/talisker/html

HTTP 303 See Other

http://wiskii.com/distilleries/talisker/rdf
4. Setup Your Infrastructure

- Testing your content negotiation
  - Install the LiveHTTPHeaders and Modify Headers extensions for Firefox
  - Try LiveHTTPHeaders against my URI
    - http://tomheath.com/id/me
    - do the same with URIs from other data sets
  - Modify your headers to ask for application/rdf+xml
  - What do you get back?
  - Do the same with cURL
    - http://dowhatimean.net/2007/02/debugging-semantic-web-sites-with-curl
4. Setup Your Infrastructure

• Rolling your own is not the only option

• See Linking Open Data area of the ESW Wiki
  – http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/PublishingTools
5. Link to Other Data Sets
Other Available Data Sets
5. Link to other Data Sets

- Popular Predicates for Linking
  - owl:sameAs
  - foaf:homepage
  - foaf:topic
  - foaf:based_near
  - foaf:maker/foaf:made
  - foaf:depiction
  - foaf:page
  - foaf:primaryTopic
  - rdfs:seeAlso
5. Link to other Data Sets

- **DBpedia**
- **WikiCompany**
- **Homepages**
- **Geonames**
- **FlickrWrappr**
5. Link to other Data Sets

- Linking Algorithms
  - String Matching
    - e.g. Lexical Distance between labels
  - Common Key Matching
    - e.g. ISBN, Musicbrainz IDs
  - Property-based Matching
    - Do these two things have the same label, type and coordinates

- Aim for reciprocal links
Summary

1. Understand the Principles
2. Understand your Data
3. Choose URIs for Things in your Data
4. Setup Your Infrastructure
5. Link to other Data Sets
Linked Data Applications
Current Linked Data Applications

- Browsing with Marbles and DBpedia Mobile
- Searching with Falcons
- Mashups, e.g. Revyu, BBC Music, Pipes
Marbles

- http://beckr.org/marbles
- plug in a URI of your choice
- browse the Web of Data/Things
- notice the effect of link density
DBpedia Mobile
The Prestige

This is a drama about intense rivalry between stage magicians in the late 19th Century. The evocation of the period, although first-rate, is not the main attraction, however. The Prestige has an incredibly clever plot including the most ingenious murder I've ever come across. It also has a deeply moving and sad love story hidden in it, which gradually emerges over the course of the film.

The film requires a strong suspension of disbelief on some key points: there is a science-fiction premise which is introduced using the real historical character of Nikola Tesla (I'd rather they had used a fictional scientist). There are a couple more implausibilities required to hold it together (something odd that goes on that none of the characters pick up on and a dead-end that by a huge coincidence turns out not to be a dead-end. I can't be more specific without spoiling the plot).

However, rather than feeling cheated by these aspects of the film, I'm hugely impressed. The writers have taken an implausible (okay, impossible) premise but created an intricate, involving and visual story that would be impossible without that premise. Scenes join up with each other in many subtle ways echoing the same writers' earlier film Memento. Even when you've seen the twist coming, the final scene which lays it all out are has a lot of impact and I suspect the final shot will haunt my dreams.

I expected the film to be about nice costumes or impressive magical trickery, but it is actually about deep emotions felt by the main characters as they deal with the situations life has dealt them, and it rather than serving up those emotions on a plate, it requires you to think and piece together what you've seen. That's got to be a good thing, in fact the best of what film a be.

What do you think of The Prestige? Write Your Own Review.
### Top Artists Played Across BBC Radio

At present information about music played on BBC radio does not include every programme or network.

**26 January - 5 February**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Artist</th>
<th>Image</th>
<th>Music Played this Week</th>
<th>Music Played Last Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Coldplay</td>
<td><img src="coldplay_image" alt="Coldplay" /></td>
<td>a mount played this week</td>
<td>a mount played last week</td>
</tr>
<tr>
<td>2</td>
<td>Lily Allen</td>
<td><img src="lily_allen_image" alt="Lily Allen" /></td>
<td>a mount played this week</td>
<td>a mount played last week</td>
</tr>
<tr>
<td>3</td>
<td>Franz Ferdinand</td>
<td><img src="franz_ferdinand_image" alt="Franz Ferdinand" /></td>
<td>a mount played this week</td>
<td>a mount played last week</td>
</tr>
<tr>
<td>4</td>
<td>The Killers</td>
<td><img src="the_killers_image" alt="The Killers" /></td>
<td>a mount played this week</td>
<td>a mount played last week</td>
</tr>
<tr>
<td>5</td>
<td>U2</td>
<td><img src="u2_image" alt="U2" /></td>
<td>a mount played this week</td>
<td>a mount played last week</td>
</tr>
<tr>
<td>6</td>
<td>Kings of Leon</td>
<td><img src="kings_of_leon_image" alt="Kings of Leon" /></td>
<td>a mount played this week</td>
<td>a mount played last week</td>
</tr>
</tbody>
</table>

### On This Day

**05 February**

**Bobby Brown** (1969)
Bobby Brown is a Grammy Award-winning American R&B singer-songwriter and dancer. After...
Falcons

Falcons

Shared innovation

Falcons

Object  Concept  Document

Birmingham

Search Objects

Specify a type:

Area  Attraction  Body  Building  City

Concept  Landmark  Location  Medium  Museum

Organization  Person  Railway Station  Sport  Subject

Objects 1 - 10 of 9,177 for your search birmingham (2.03 seconds)

Birmingham is a Borough, City

- abstract: Birmingham i USA https på Birmingham, Alabama Birmingham
- has subject: Category 1st millennium establishments
- hasPhotoCollection: Birmingham
http://wikidata.org/resource/Birmingham - Described in 653 documents

People from Birmingham, Alabama is a Concept

- has broader: Category Birmingham 2 C Alabama
- preferred label: People from Birmingham, Alabama
- label: People from Birmingham, Alabama
http://wikidata.org/resource/Category:People_from_Birmingham%2C_Alabama - Described in 255 documents

Birmingham_City_F.C. is a Club, Company

- abstract: ...-ch-hk 伯明翰城足球會 (zh-ch-hk 伯明翰城足球會) (Birmingham City Football Club)是一家英格蘭 ...-ch-hk 伯明翰城) ...-ch-cn 伯明翰城足球俱樂部)
- has subject: Category English football clubs
- hasPhotoCollection: Birmingham City F C
http://wikidata.org/resource/Birmingham_City_F.C. - Described in 205 documents

Birmingham is a Ontology, Subject, City, Geographical region

- imports: http://semantic-mediaeval.org/swat/1.0
- date of creation: 2008-05-28T13:42:38+02:00
- page: Birmingham
http://semanticweb.org/id/Birmingham - Described in 10 documents

Birmingham%2C_Alabama is a City

- abstract: Birmingham to także nazwa miasta Anglii Birmingham - mi ...-ch-cn 伯明翰城)
- has subject: Category United States communities with African American majority populations
- hasPhotoCollection: Birmingham 2 C Alabama
http://wikidata.org/resource/Birmingham%2C_Alabama - Described in 230 documents

Birmingham Barons is a Sport, Team

- abstract: The Birmingham Barons are a minor league baseball team based i ...-ch-cn 伯明翰)
- has subject: Category Minor league baseball teams
- hasPhotoCollection: Birmingham Barons
http://wikidata.org/resource/Birmingham_Barons - Described in 13 documents
Semantic Web Pipes

- Like Yahoo Pipes, but for RDF
- http://pipes.deri.org/
Outlook for Linked Data Applications

• Requirements
  – slicker interfaces
  – better backend infrastructure
  – highly focused functionality
Linked Data Toolbox
Linked Data Storage/Publishing Layers

- **D2R Server**
  - Relational Database to RDF Middleware
  - SPARQL access to RDB
  - [http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/](http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/)
  - Example
  - LinkedMDB [http://linkedmdb.org/](http://linkedmdb.org/)
Linked Data Storage/Publishing Layers

- Virtuoso
  - Many things, including RDF triplestore
  - SPARQL access to data
  - Open source edition
  - http://virtuoso.openlinksw.com/
Linked Data Storage/Publishing Layers

• Talis Platform
  – SaaS, Cloud-based storage for RDF data and binary objects
  – SPARQL access
  – REST APIs to additional services
    • Faceting, Augmentation
  – Linked Data compatible out of the box
  – http://www.talis.com/platform
Linked Data Storage/Publishing Layers

- Paget Framework
  - publishing framework for Linked Data
  - serves up RDF according to Linked Data principles
  - reduces configuration overhead
  - can serve up data from static files or the Talis Platform
  - http://code.google.com/p/paget
Consuming Linked Data

• RDF Frameworks
  – ARC (PHP) http://arc.semsol.org/
  – RAP (PHP) http://www4.wiwiss.fu-berlin.de/bizer/rdfapi/
  – Jena (Java) http://jena.sourceforge.net/
  – Summary

• Discovering more data
  – Sindice http://sindice.com/
  – SQUIN http://squin.sourceforge.net/
Discussion
More Information

• Contact Details
  – tom.heath@talis.com
  – http://tomheath.com/
  – http://www.talis.com/

• Slides
  – ...

• Tutorial
  – http://linkeddata.org/docs/how-to-publish
shared innovation