

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



- 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



- 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



- Organisation
- My time
- Coffee and lunch
- Tonight's beer









Linked Data, What and Why?



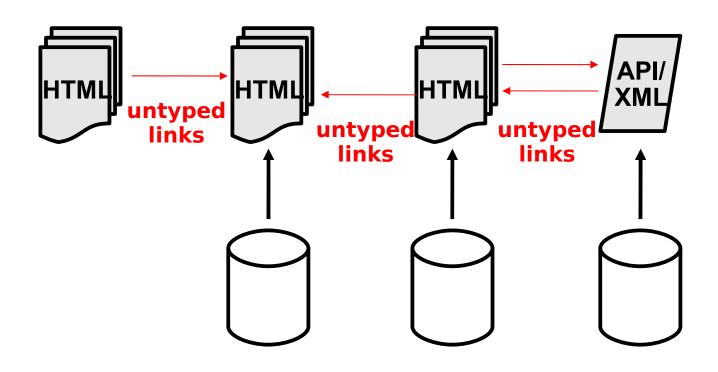
The Web of Documents

- 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





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?

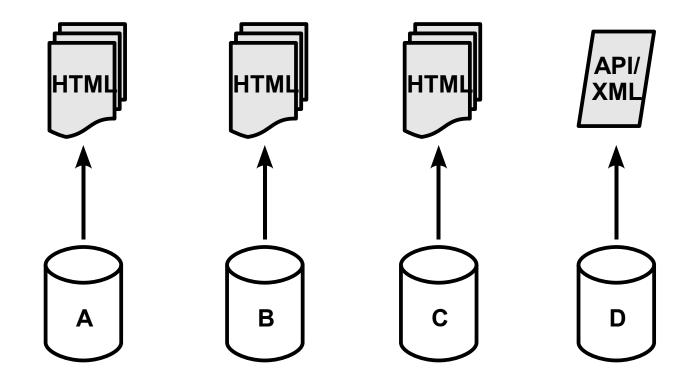


Data Silos on the Web





Data Silos on the Web





The Web of Linked Data

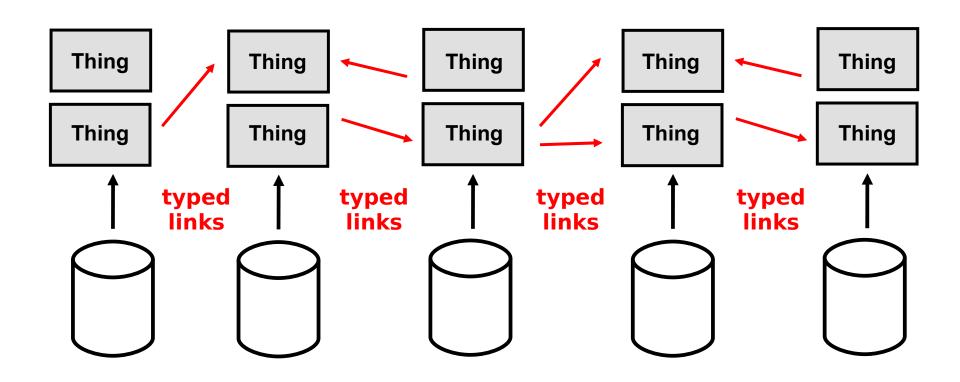
- Analogy
 - a global database
- Designed for
 - machines first, humans later
- Primary objects
 - things (or descriptions of things)
- Links between
 - things
- Degree of structure in (descriptions of) things
 - high
- Semantics of content and links
 - explicit





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/



- 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
```

 $Tom \rightarrow worksFor \rightarrow Talis$

Talis \rightarrow basedIn \rightarrow Birmingham

```
\langle uri \rangle \rightarrow \langle uri \rangle \rightarrow \langle uri \rangle 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/>

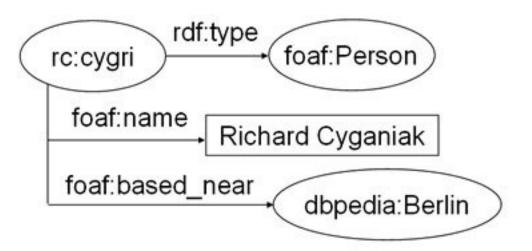


Data Merging with RDF

- 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/>

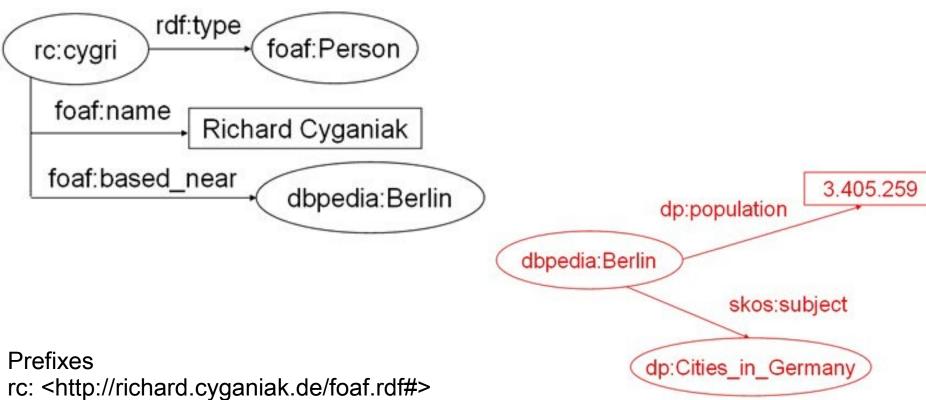
dbpedia: http://dbpedia.org/resource/

dp: <http://dbpedia.org/property/>

skos: http://www.w3.org/2004/02/skos/core#>



Data Merging with RDF



rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#>

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

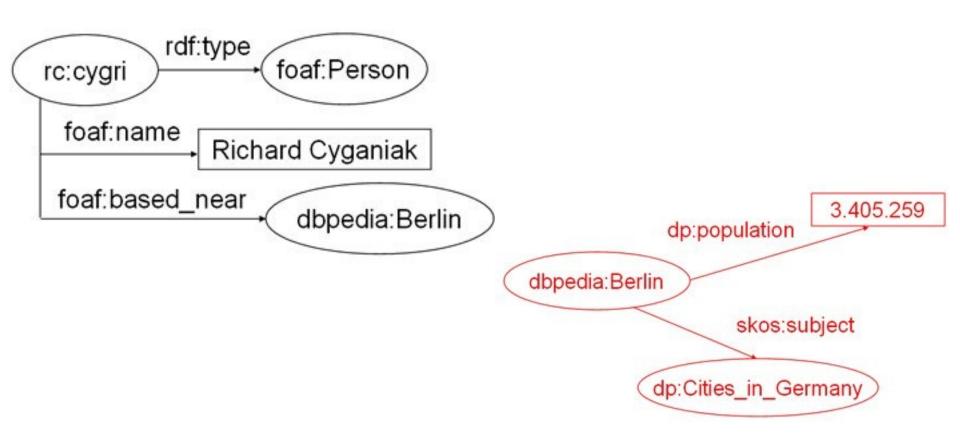
dbpedia: http://dbpedia.org/resource/

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



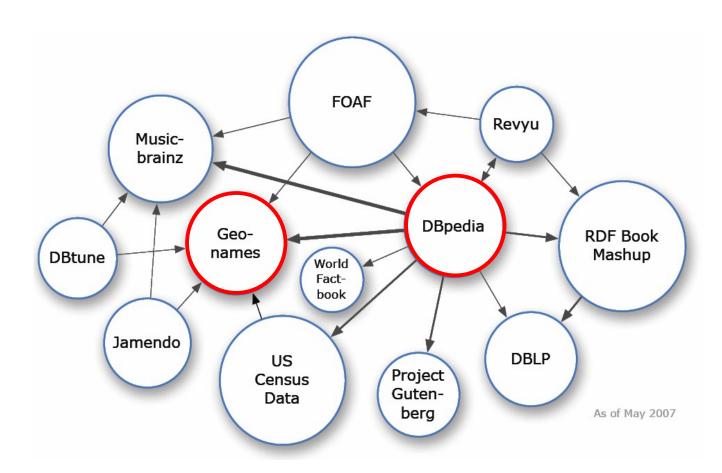
Participants

- 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)

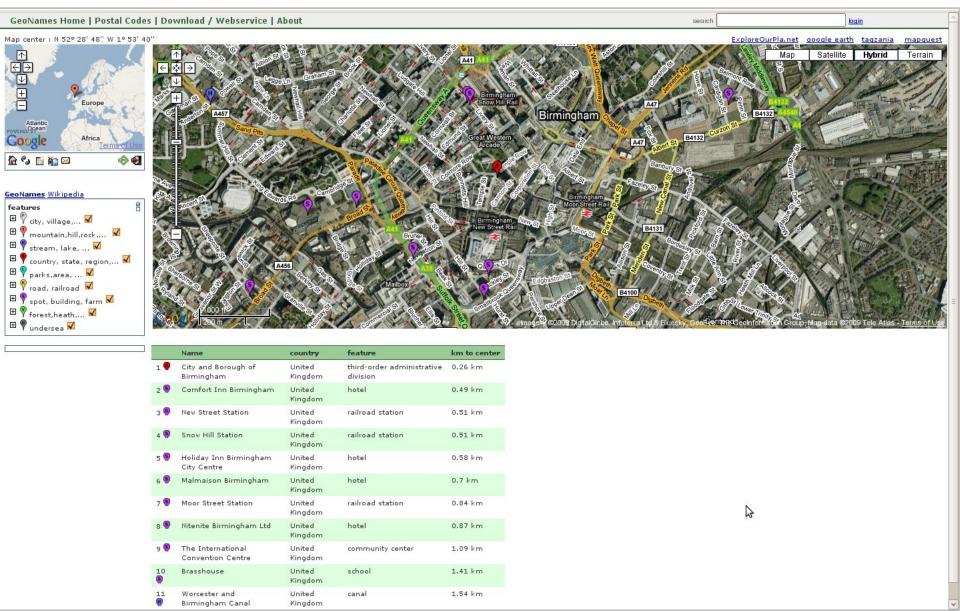
- BBC (UK)
- Talis (UK)
- Garlik (UK)
- OpenLink (UK)
- Thomson Reuters (US)
- Zitgist (US)
- Mondeca (FR)
- Cyc Foundation (US)



The LOD "Cloud" - May 2007











navigation

Main page

- Contents
- Featured content
- Current events
- Random article

search



interaction

- About Wikipedia
- Community portal Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

toolbox

- What links here
- 2.1 Geology Related changes
- Upload file

languages

العربية ■

 Avañe'ë Aymar aru

Bosanski

Чавашла

Cymraeg

Català

Česky

Dansk

Eesti

Deutsch

Ελληνικά

Български

Afrikaans

- Special pages
- Printable version
- Permanent link 4 Places of interest
- Cite this page 5 Economy

1 History

2 Geography

2.2 Climate

3 Demographics

2.3 Nearby places

- 6 Politics
- 7 Transport
- 8 Education
- 9 Sport
- 10 Food & drink
- 11 Culture and arts
 - 11.1 Arts
 - 11.2 Festivals and shows

Contents [hide]

- 11.3 Media
- 11.4 Leisure
- 11.5 Architecture
- 12 Environment
- 13 Crime and policing
- 14 Second City
- 15 Notable residents
- Español Esneranto 16 Science and invention

Birmingham

From Wikipedia, the free encyclopedia

Coordinates: 6 52°29'1"N 1°54'23"W

Birmingham

- City and Metropolitan borough -

City of Birmingham

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 (4) pronunciation (help-info), /'bax:mɪnəm/ BIRR-ming-əm) 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 1,010,200 (2005 estimate). Often considered to be the second city of the United Kingdom^[2], the City of Birmingham forms part of the larger West Midlands conurbation, which has a population of 2,284,093 (2001 census)[3] 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 1998, 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, 'Bromwicham'. There is a distinctive Brummie dialect and accent, both of which differ from the adjacent Black Country.

Birmingham's skyline viewed from Bartley Green

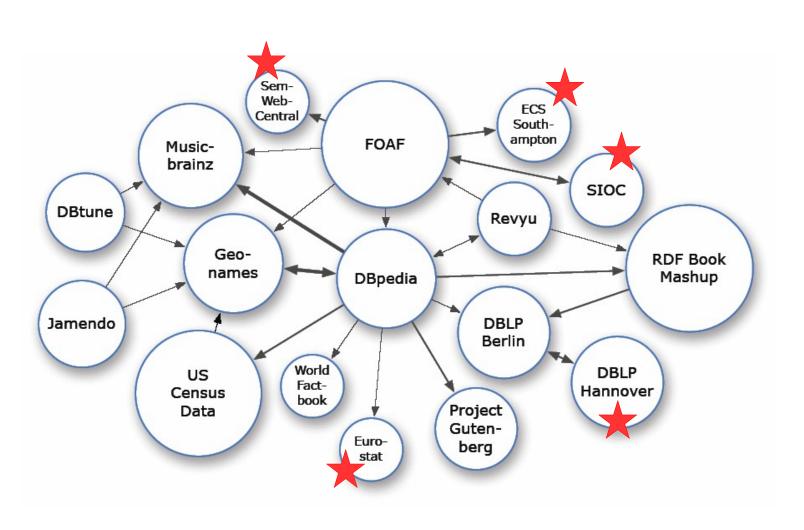


Nickname(s): "Brum", "Brummagem", "Second City", "Workshop of the World", "City of a Thousand Trades", "Canal City" Motto: Forward



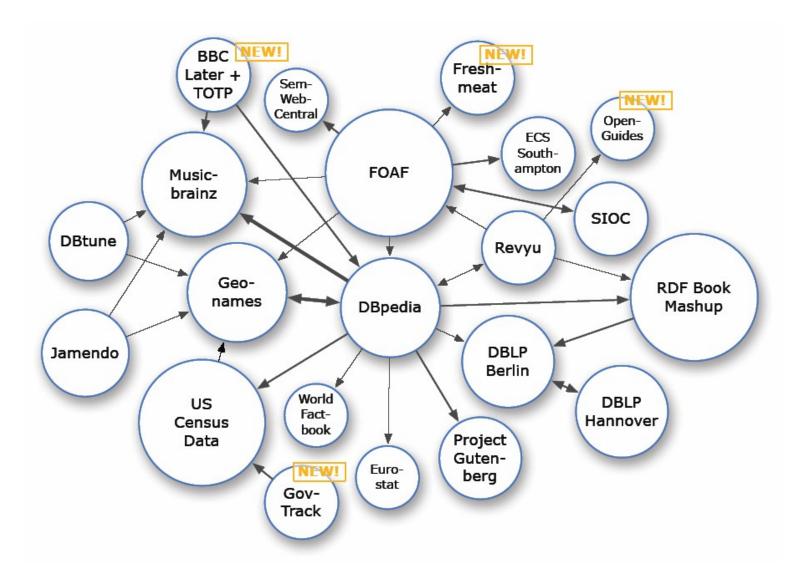


The LOD "Cloud" - July 2007



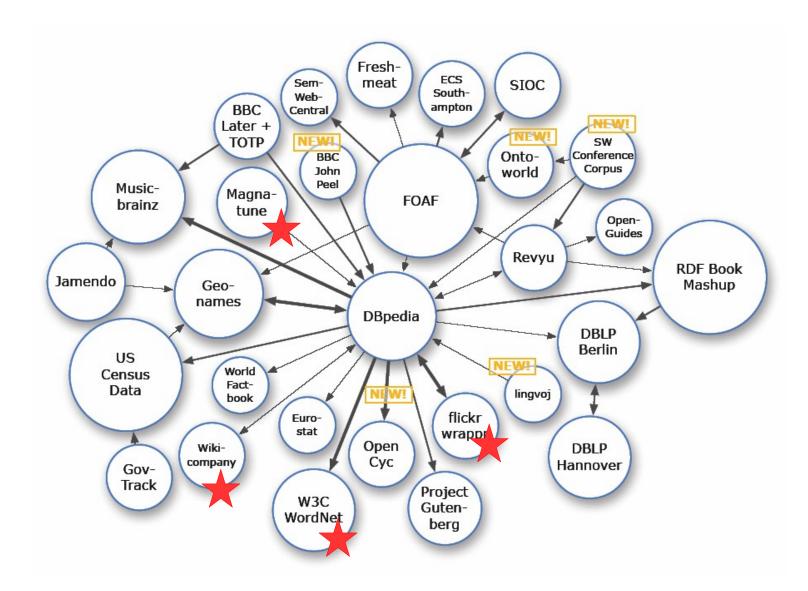


The LOD "Cloud" - August 2007



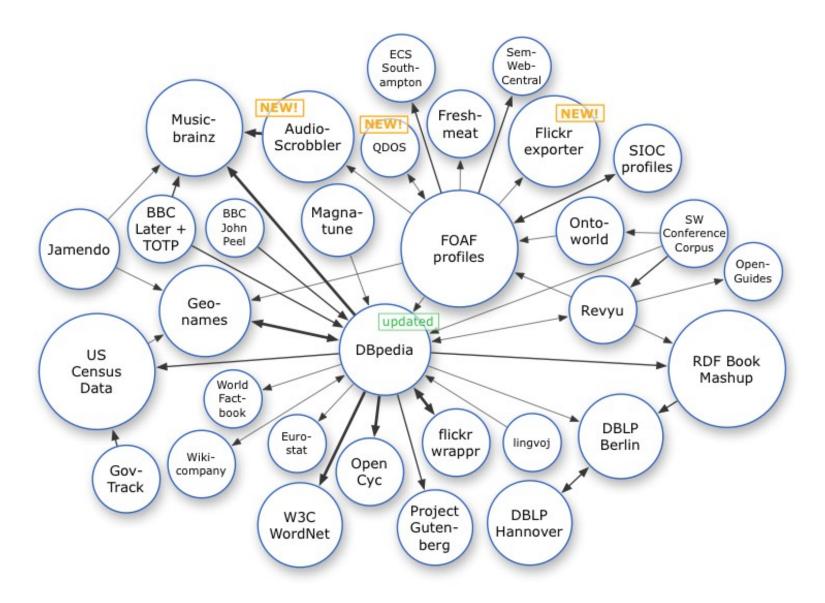


The LOD "Cloud" - November 2007



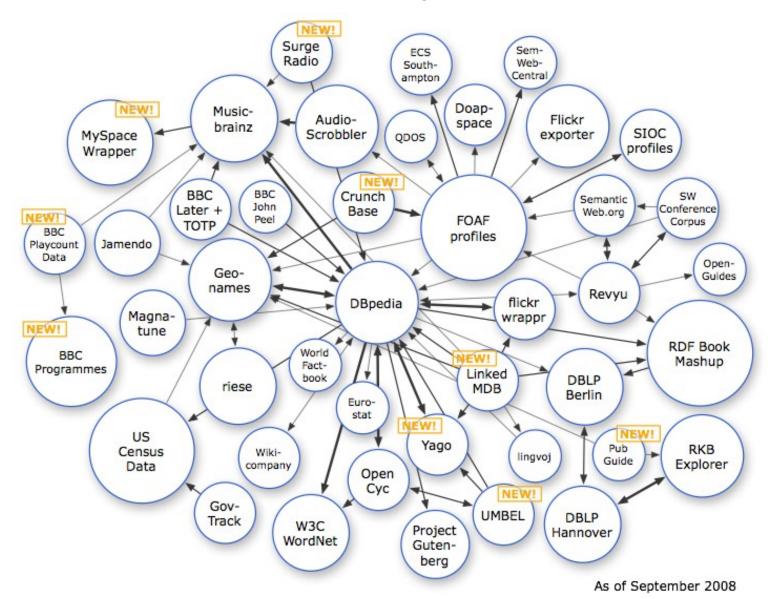


The LOD "Cloud" - Feb 2008





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

http://linkeddata.org/images-and-posters

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



- 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

talis

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?
 - **—** ...



2. Understand Your Data

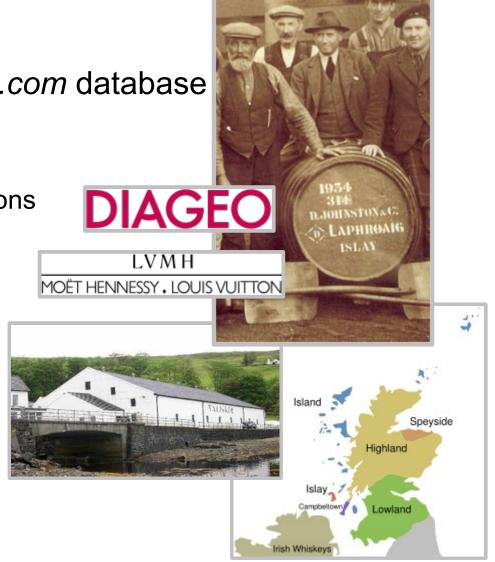
Things in the Wiskii.com database

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

talis

3. Choosing URIs: Principles

- Use HTTP URIs
- Keep out of other peoples' namespaces
 - 1. http://www.imdb.com/title/tt0441773/
 - 2. http://www.imdb.com/title/tt0441773/thing
 - 3. http://myfilms.com/tt0441773
 - 4. http://myfilms.com/tt0441773/html
- Abstract away from implementation details
 - 1. http://dbpedia.org/resource/Berlin
 - 2. http://www4.wiwiss.fu-berlin.de:2020/demos/dbpedia/cgi-bin/resources.php?id=Berlin
- 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
- http://dbpedia.org/data/New_York_City
- http://dbpedia.org/page/New_York_City
- http://revyu.com/people/tom
- http://revyu.com/people/tom/about/rdf
- http://revyu.com/people/tom/about/html
- http://kmi.open.ac.uk/people/tom/
- http://kmi.open.ac.uk/people/tom/rdf
- http://kmi.open.ac.uk/people/tom/html
- http://mydomain.com/thing
- http://mydomain.com/thing.rdf
- http://mydomain.com/thing.html

- ← Thing
- ← RDF data
- ← HTML page
- ← Thing
- ← RDF data
- ← HTML page
- ← Thing
- ← RDF data
- ← HTML page
- ← Thing
- ← RDF data
- ← 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/products/talisker-20-yo
- http://wiskii.com/products/glenmorangie-lasanta
- http://wiskii.com/people/william-matheson
- http://wiskii.com/photos/58
- http://wiskii.com/reviews/271

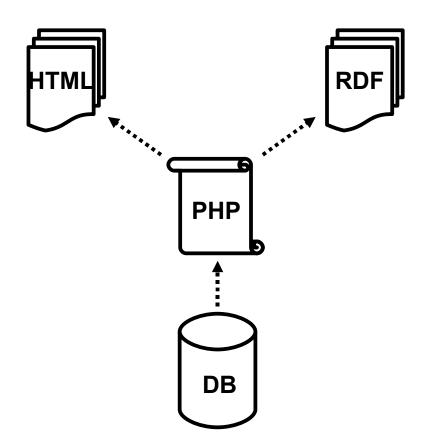


3. Choosing URIs: Wiskii.com

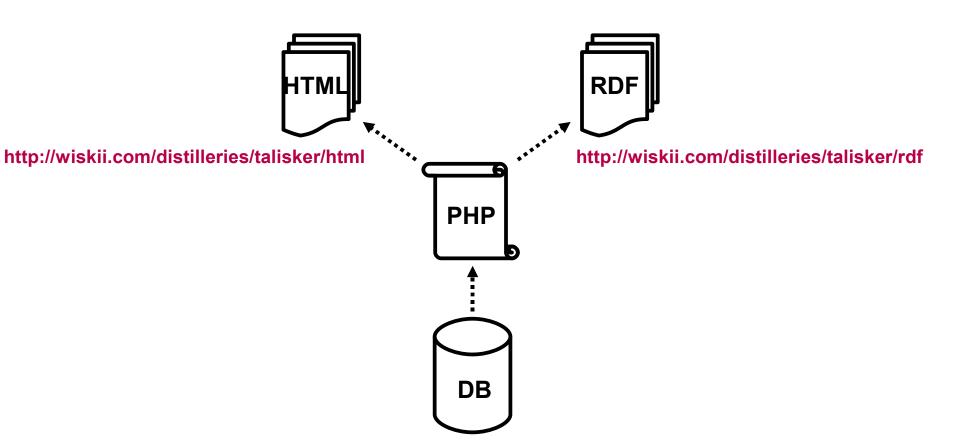
- http://wiskii.com/distilleries/talisker
- http://wiskii.com/distilleries/talisker/rdf
- http://wiskii.com/distilleries/talisker/html
- http://wiskii.com/brands/talisker
- http://wiskii.com/brands/talisker/rdf
- http://wiskii.com/brands/talisker/html
- http://wiskii.com/people/william-matheson
- http://wiskii.com/people/william-matheson/rdf
- http://wiskii.com/people/william-matheson/html
- http://wiskii.com/photos/58





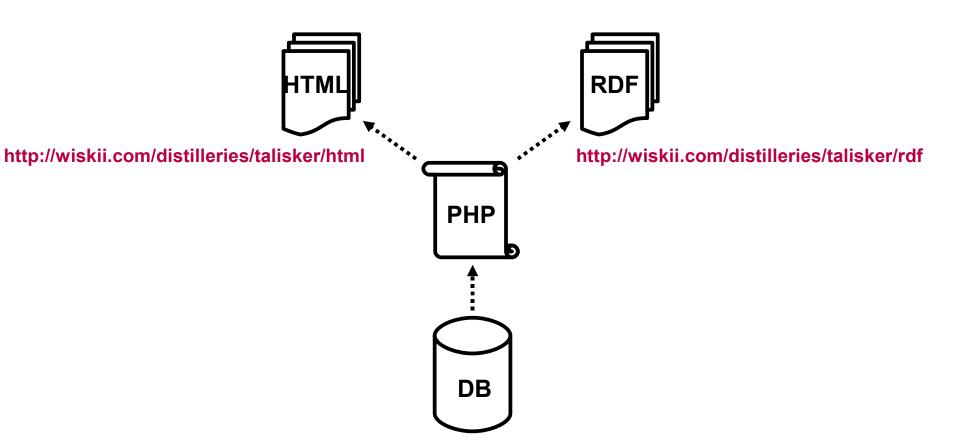






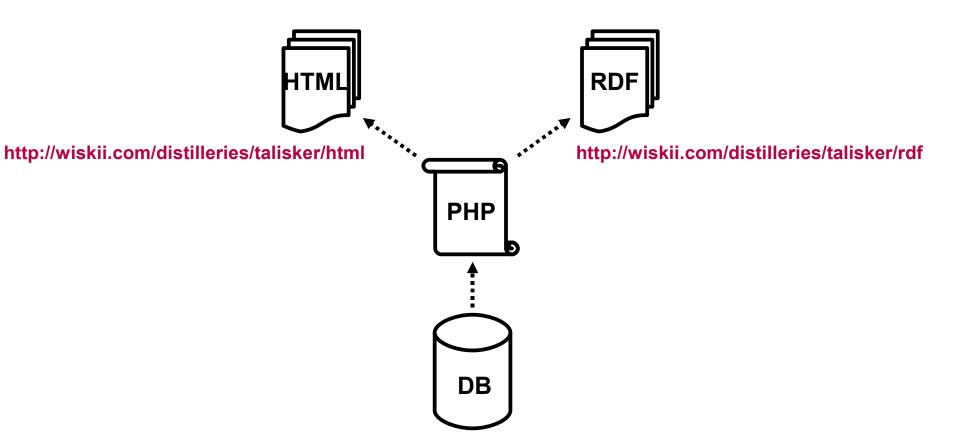


http://wiskii.com/distilleries/talisker

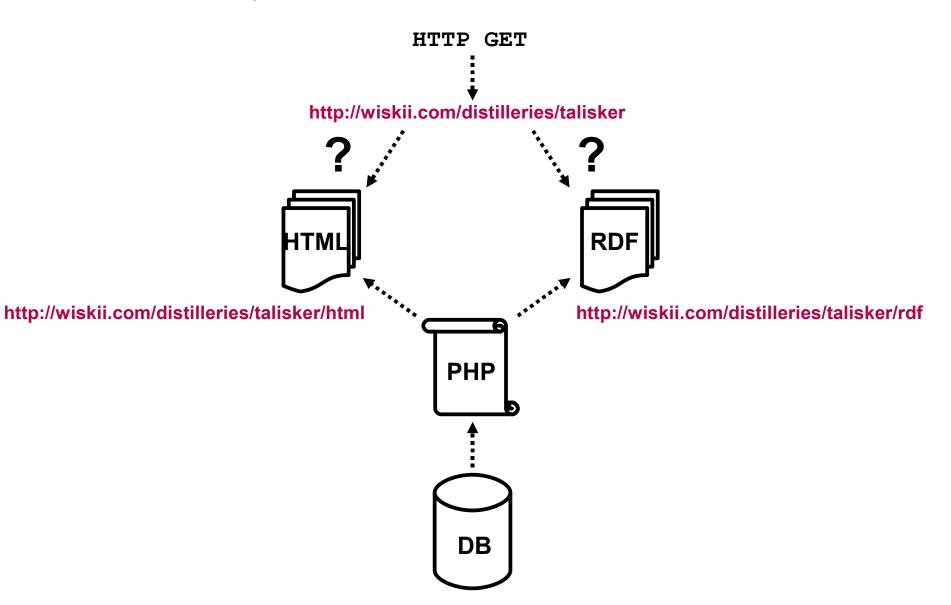






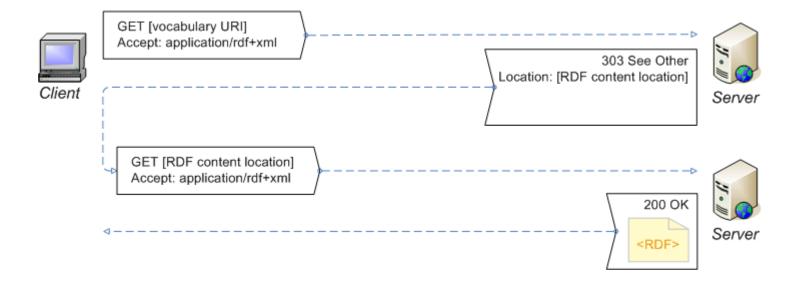




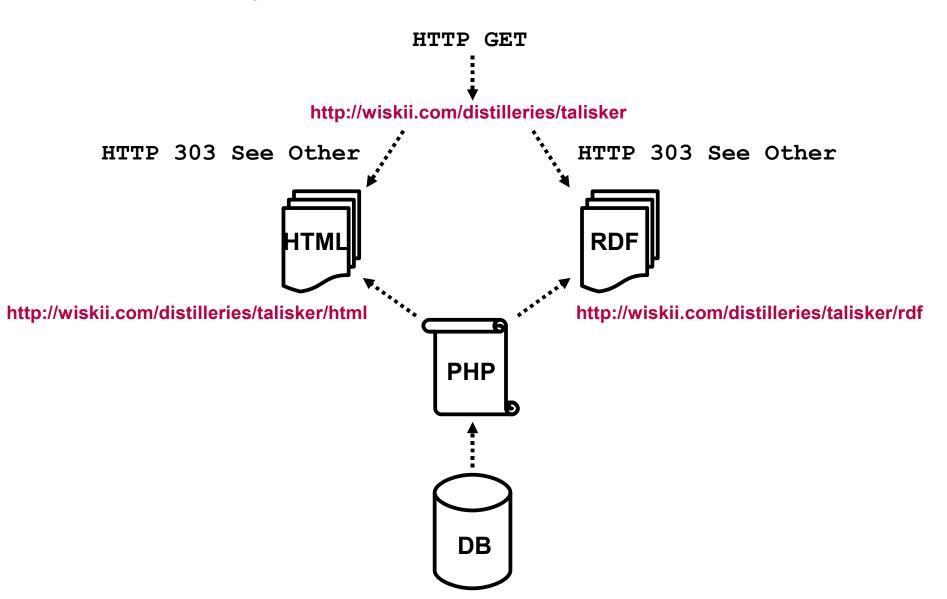


talis®

Content Negotiation



talis®





- 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-semanticweb-sites-with-curl



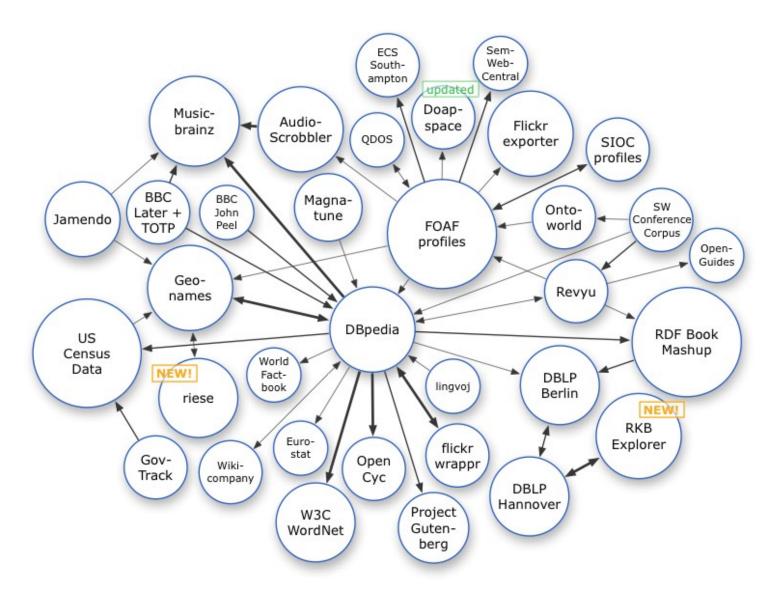
- Rolling your own is not the only option
- See Linking Open Data area of the ESW Wiki
 - http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/PublishingTools



5. Link to Other Data Sets



Other Available Data Sets



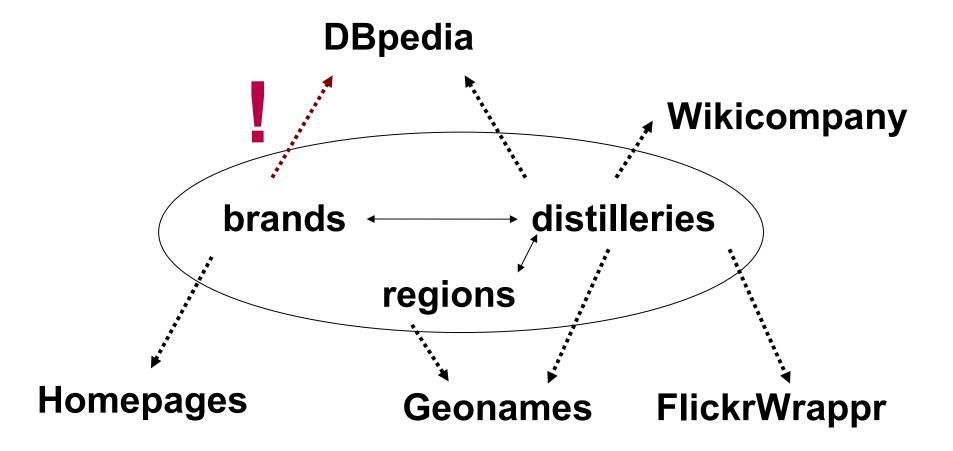


5. Link to other Data Sets

- 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





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

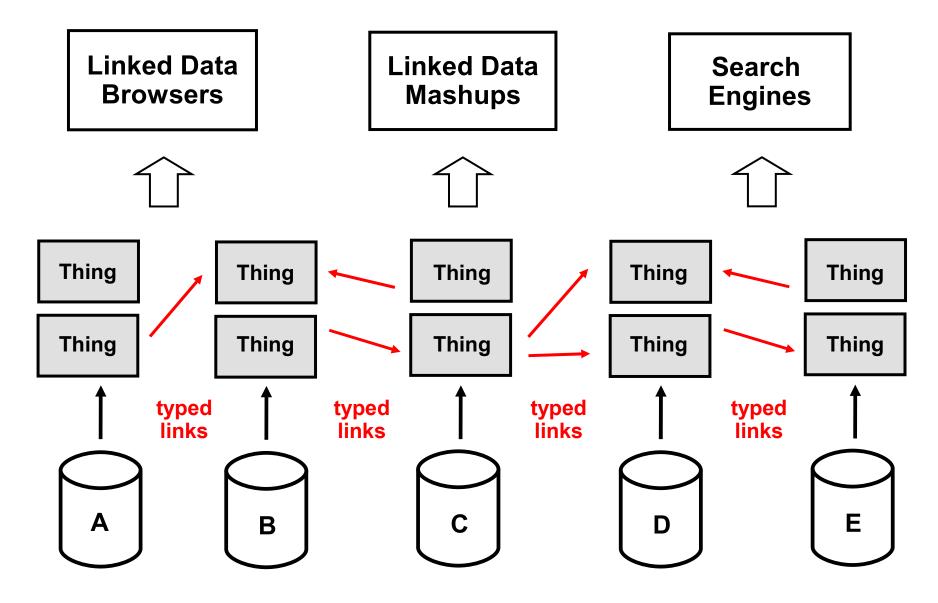


- 1. Understand the Principles
- 2. Understand your Data
- 3. Choose URIs for Things in your Data
- 4. Setup Your Infrastructure
- 5. Link to other Data Sets



Linked Data Applications







Current Linked Data Applications

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



- 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









Home | Browse Things | Search Things | Browse People

Login/Register | New Review

The Prestige

Links

Homepage: http://theprestige.movies.go.com/ See Also: http://imdb.com/title/tt0482571/

Tags

christian-bale christopher-nolan drama entertainment film hugh-jackman illusion magic michael-caine movie murder period scarlett-johansson science-fiction whodunnit

Reviews (1)

by martinp on 23 Jan 2007

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...

Revyu.com: Contact | Credits | Privacy Policy | Disclaimer

The Prestige



directed by Christopher Nolan

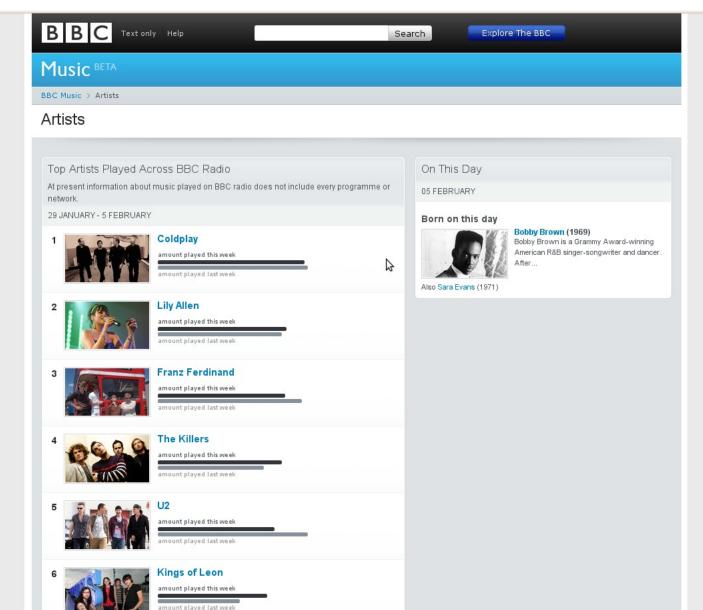
RDF Metadata About The Prestige

RDF META

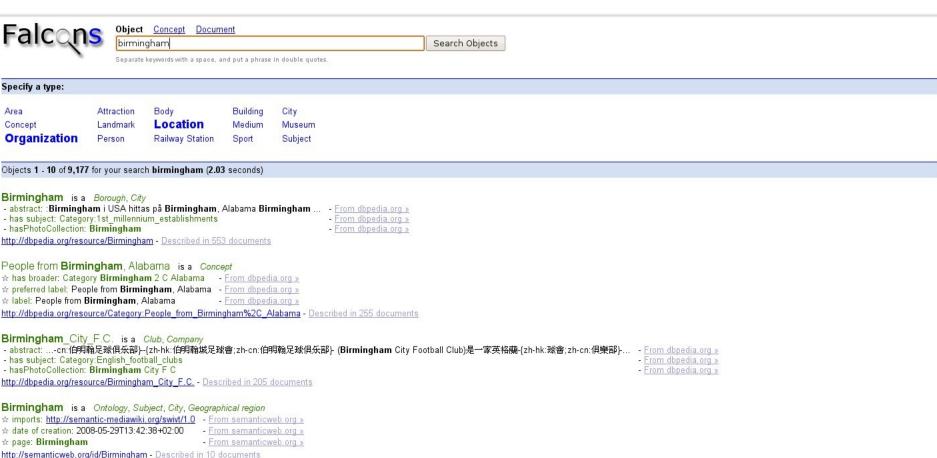
Write a Review of The Prestige

- Add to del.icio.us









- From dbpedia.org »

- From dbpedia.org »

Birmingham%2C_Alabama is a City

- hasPhotoCollection: Birmingham 2 C Alabama

Birmingham_Barons is a Sport, Team
- abstract: The Birmingham Barons are a minor league baseball team based i... - From dbpedia.org >

- has subject: Category: United States_communities_with_African_American_majority_populations - From dbpedia.org >

- has subject: Category:Minor_league_baseball_teams - From dbpedia.org >
- hasPhotoCollection: Birmingham Barons - From dbpedia.org >

http://dbpedia.org/resource/Birmingham_Barons - Described in 13 documents

- abstract: :Birmingham to także nazwa miasta w Anglii. Birmingham - mi...

http://dbpedia.org/resource/Birmingham%2C Alabama - Described in 238 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

talis®

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/
- Example
- LinkedMDB http://linkedmdb.org/

talis®

Linked Data Storage/Publishing Layers

Virtuoso

- Many things, including RDF triplestore
- SPARQL access to data
- Open source edition
- http://virtuoso.openlinksw.com/

talis®

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
 - http://www.semanticscripting.org/SFSW2005/SFSW-Toolkits.pdf
- 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