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

SITUATION: THERE ARE 14 COMPETING STANDARDS.

14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES. YEAH!





# It's all about Identification

The Andrew W. Mellon Foundation Workshop on Reconciliation of Linked Open Data New York, 1<sup>st</sup> December 2014

## **Hugh Glaser**

http://eprints.soton.ac.uk/id/eprint/372358

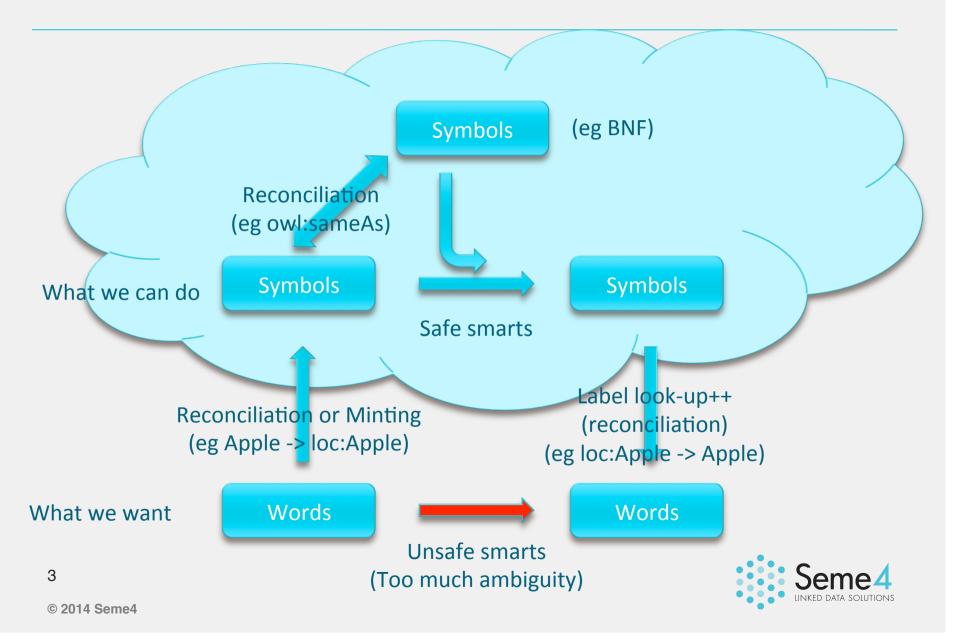
### Seme4

18 Soho Square London W1D 3QL T: + 44 20 7060 1590



© 2014 Seme4 Seme4.com

## **Overview of Identification Issues**





## **Heraclitus**

- "All entities move and nothing remains still"
- "You cannot step twice into the same stream"



# **Cratylus (Pupil of Heraclitus)**

 According to Aristotle, "ended by thinking that one need not say anything, and only moved his finger; and criticized Heraclitus for saying that one cannot enter the same river twice, for he himself held that it cannot be done even once"

- So do we want to be mute?
- We must build the assumption of "flow" into our systems



# **Identification Management**

- Specialist Knowledge Base
- No Authority
  - What price
    - ISNI
    - ORCID
    - VIAF
    - ENS (okkam)
- In fact no Agreement needed must be a principle
- Dynamic, not Static
- Distributed



## The Web

- You cannot control what others say
  - No de jure IDs
- You can say what you want to
  - de facto IDs

- You can consume what you choose
- Others consume what they choose

- Ideally there is the data to inform such decisions
- And the technology to action the consequences



# Doing the Reconciliation of Words to Symbols

## The Context of Words is ambiguous

- Project immediately to Symbols
  - Probably mint lots of Symbols
- Then use the "safe" context of the structured data

## Compare High School Algebra – "Variables"

- There are 3 times as many pears as oranges. A group of children receive 5 oranges and 8 pears each ...
- Let p be the number of pears ...
- Not: Well I guess the number of pears may be the same as Mary has, which is q ...

## By the way, People do it badly

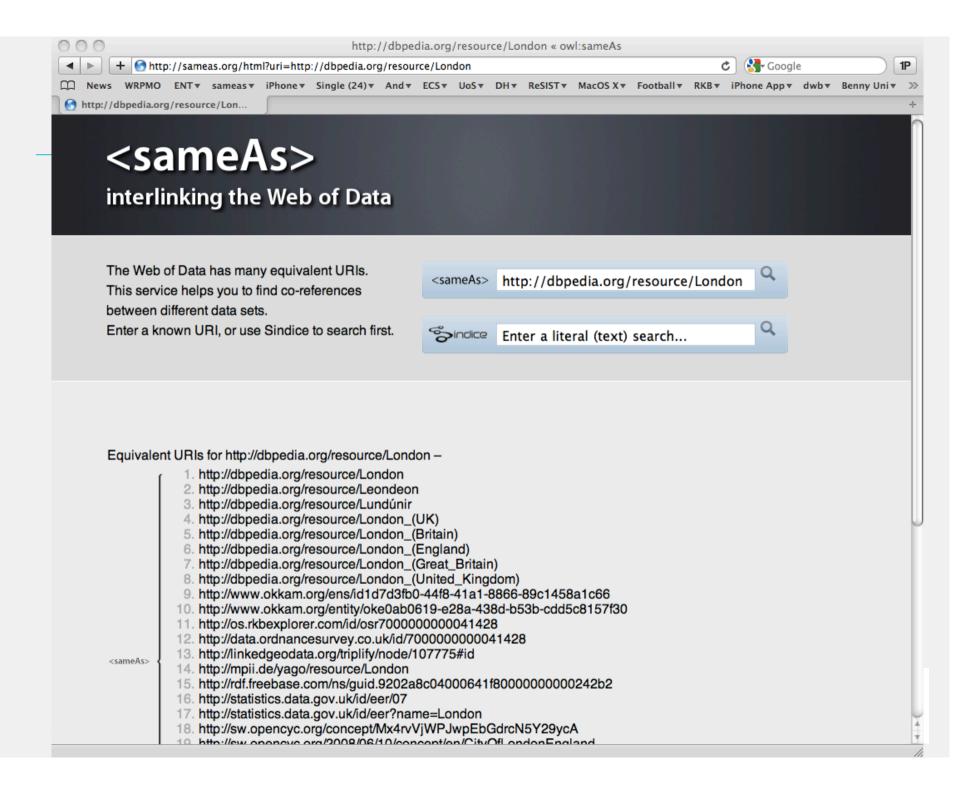
- They don't have the expertise
  - Eg Profs giving the job to their PGs
- They unknowingly step outside their expertise
  - Eg Engelbert Humperdinck



# **My Solution for Management**

- "sameAs" identification Knowledge Bases
- Supports issues around reconciliation
- Result storage
- Result publishing
  - How to make the results public?
  - How to reuse within the organisation?
- Can be a lookup resource
- Equivalence Identification
- Label Identification





## sameAs™

- I run sameAs.org as a service to the community
  - Long-standing, mature technology
- Linking without authority
  - Remember the XKCD Cartoon
- Simply:
  - Given a symbol ID, it tells you which other symbol IDs are the "same"
- The glue for the Open Data world
- Sort of like Google for Linked Data
  - In fact Google send me their sameAs data
  - As do Library of Congress, British Library, BNF, DNB, to name a few
- You too can have a sameAs store, to publish such data



## What does sameAs do?

- One Thing Well
- symbol -> {symbol<sub>1</sub>, symbol<sub>2</sub>, ..., symbol<sub>n</sub>}
- PUT symbol pair
- GET on a symbol
- DELETE symbol pair



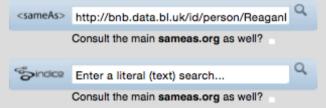


## interlinking the Web of Data

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.



This is a sub-service of the sameAs service, and contains only information that is provided by a third party.

The data presented here relates to Hugh's experiment in gathering together library subject and associated data from good sources. It is likely to form part of a forthcoming project in the library area.

You can find URIs that are on the following domains, obtained from the sources, with their help and permission:

http://bnb.data.bl.uk http://d-nb.info http://data.bibsys.no http://nektar.oszk.hu http://viaf.org http://id.loc.gov http://id.ndl.go.jp http://stitch.cs.vu.nl http://datos.bne.es http://data.bnf.fr with some links to http://dbedia.org

KELLE Store
http://sameas.org/store/kelle/

Equivalent URIs for http://bnb.data.bl.uk/id/person/ReaganRonald -

- http://dbpedia.org/resource/Ronald\_Reagan
- http://bnb.data.bl.uk/id/person/ReaganRonald
- http://d-nb.info/gnd/118598724
- http://d-nb.info/gnd/http://d-nb.info/gnd/118598724
- http://data.bibsys.no/data/notrbib/authorityentry/x90052261
- http://libris.kb.se/resource/auth/237204
- http://nektar.oszk.hu/resource/auth/523530
- http://viaf.org/viaf/76321889
- http://viaf.org/viaf/104721224
- http://viaf.org/viaf/150795308
- 11. http://viaf.org/viaf/76321889/#foaf:Person
- 12. http://www.idref.fr/027091775/id

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

## **Facilitates**

- Collaboration
  - Many reconcilers can work together
  - And can be different from the data providers
- Avoids the n<sup>2</sup> problems of owl:sameAs
- Avoids duplication of information
  - No labels for Linked Data it's at the URI
- Dynamic identification identification
- Separation of Concerns
- Multiple Reconciliation Views
  - For different contexts and uses
- Can be Used for SImple Reconciliation in Restricted Domains
  - Zip codes
  - British actors



## What To Do

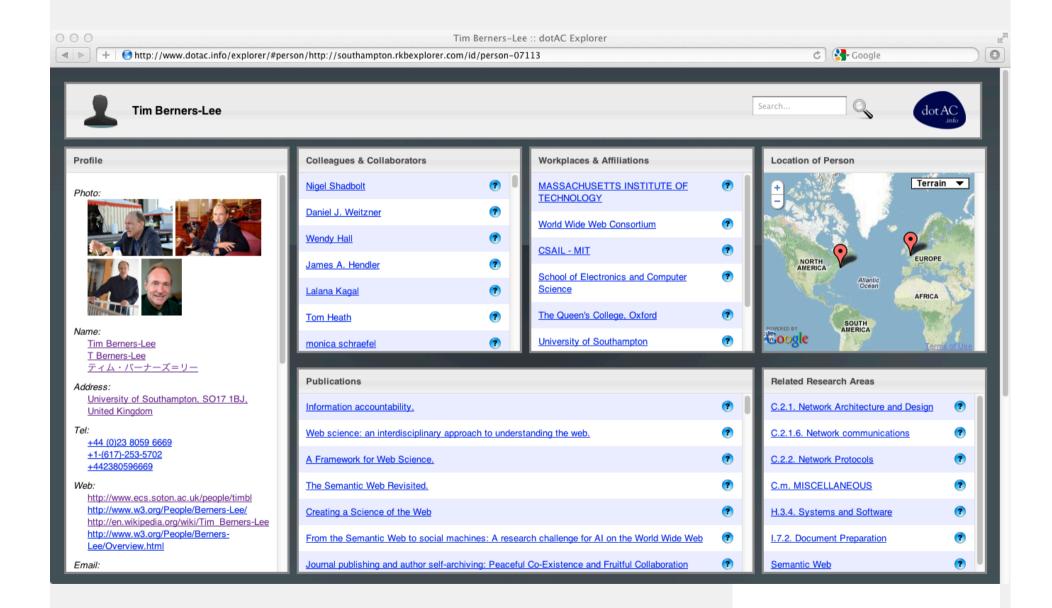
- Publish the Data (Raw Data Now)
  - Mostly with "natural" IDs/URIs
  - RDF
- Use It
  - So it is loved and maintained
- Publish Your Linkages/Reconciliations
- Make your systems sameAs aware
- Encourage and help others to Reconcile over it
- Federate sameAs stores
  - Transitive closure
- Canonise if Appropriate
  - Republish the RDF with preferred URIs
  - Deprecate the old ones they become keys but not values in the sameAs store



# You Really Can Build Complex Apps

- http://www.ukphotonics.org
- http://apps.seme4.com/see-uk/
- http://apps.seme4.com/see-soton/
- http://www.dotac.info/explorer/
- http://genealogy.ragld.com/
- http://ysgolion.com
- http://hampshire.data.seme4.com/services/ visualisation/





## differentFrom!

- Don't Forget differentFrom (differentFrom.org)
- More costly knowledge than sameAs
- It's your Regression Tests
- And You need to manage that Data
- And Publish it, just like sameAs



# **Thank You**

Next slides are about examples



## **Product: UK Photonics Portal**

- http://www.ukphotonics.org
- A dynamic portal for the UK Photonics Industry
- Linked Data integrates across multiple sources
  - avoids curating own copy of data
    - no "staleness" problem
  - bring in data where possible "from source"
- Comprehensive equivalence mapping between synonyms
- Integrated within Wordpress CMS





## **UK PHOTONICS**

The knowledge-driven photonics portal

Home

About

**Community News** 

Photonics resources

Capabilities & Techniques

Researchers

Organisations

**Equipment & Facilities** 

Useful links

**Events** 

**Portal News** 

Frequently asked Questions

Advanced Search

**EPSRC Centre for Innovative** Manufacturing in Photonics



## General Engineering Research Institute (GERI) - John Moores University

Byrom Street, Liverpool, L3 3AF, United Kingdom

w: http://www.ljmu.ac.uk/geri/

General Engineering Research Institute (GERI) - John Moores University is part of Liverpool John Moores University.

### Sub groups

- Coherent and Electro-Optics Research Group
- Electronic and Ultrasonic Engineering
- Photonics in Engineering
- The Advanced Manufacturing Technology Research Laboratory

### Members

- Mr Peter Moran
- Mr. Guang Ming Zhang
- Dr. Francis Lilley
- Mr Adam Rosowski
- Dr. Mark Murphy
- Dr Martin Sharp
- Dr Tahsin Tecelli Opoz
- Dr Michael N Morgan
- Professor Allan Hobson
- Professor Xun Chen
- · Dr. Frederic Bezombes
- · Professor Michael J. Lalor
- Dr. Paul French
- · Professor David M. Harvey
- Dr Andre Batako
- · Professor David R. Burton

### SEARCH PORTAL CONTENT

Capabilities and Techniques...

or try advanced search

PORTAL NEWS

Public launch of UK Photonics Portal

A guide to Freebase

Data sources... exposed

New data sources integrated

Refinements to search

Data in the portal is gathered with minimum editing from many sources. As a result there can be significant variations in how people describe themselves and their organisations' capabilities. Care is therefore advised when searching to produce fully inclusive results. Our hints and tips will help, as will our example queries.

### **NEWS HIGHLIGHTS**

**EPSRC** funded projects TSB funded projects Where to find Photonics in H2020 Photonics in H2020 KTN meeting Dec 2013

**Tweets** 





# Product: <sameAs>

- sameAs.org is the de-facto resource for co-reference in the Open pen Linked Data cloud
- Primarily a discovery mechanism
  - given a URI, what other URIs describe the same resource / concept?

rdd

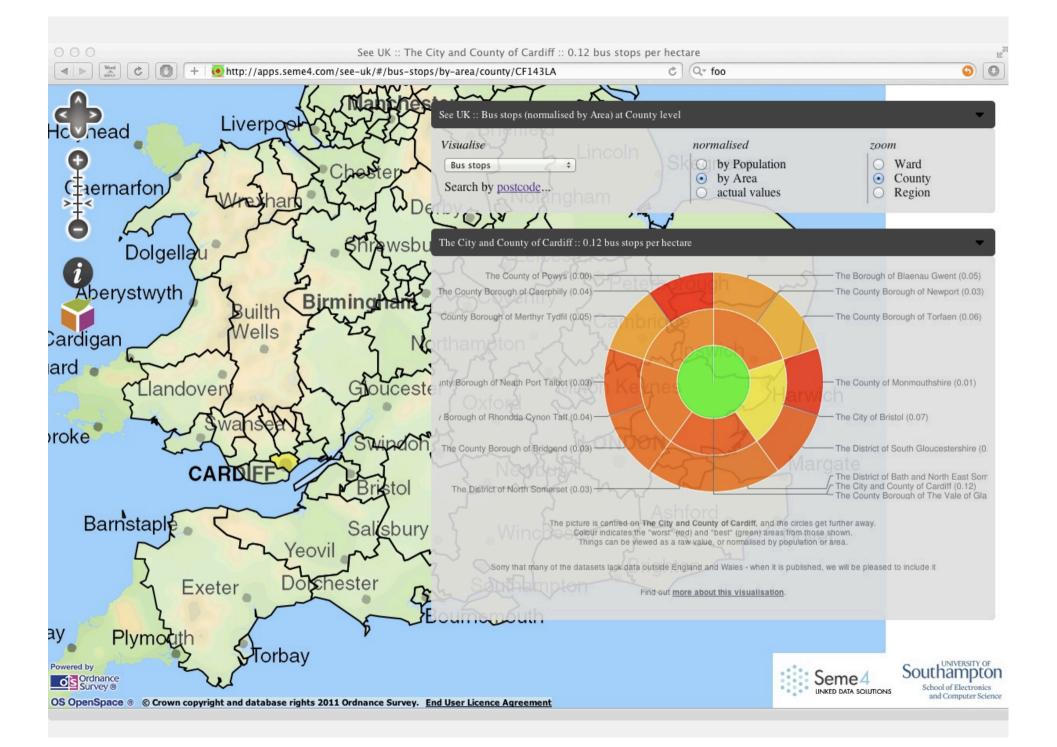
- Sub-stores for publishing 33 party data a
  - used by Google, British Library, JISC,
  - Ordnance Survey, VIAF, LATC,
  - and many others

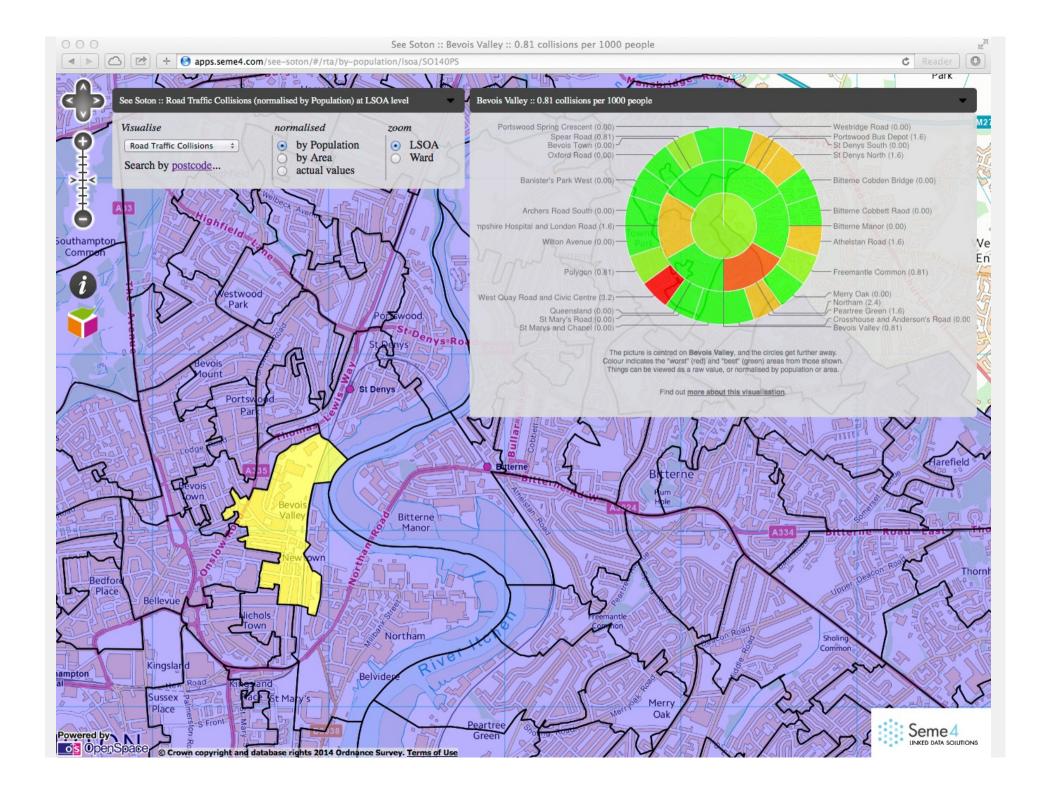


## **Demos: See-UK and See-Soton**

- http://apps.seme4.com/see-uk/
- http://apps.seme4.com/see-soton/ n/
- Wide variety of coordinate based public sector information converted dinto Linked Data
- Point locations resolved to electoral wards de
- Aggregations into higher level regions, and normalisations by area a and population
- Web application compares neighbouring regions on a relative "good" to "bad" scale
- Custom lower level view for Southampton City Council, with additional city specific data



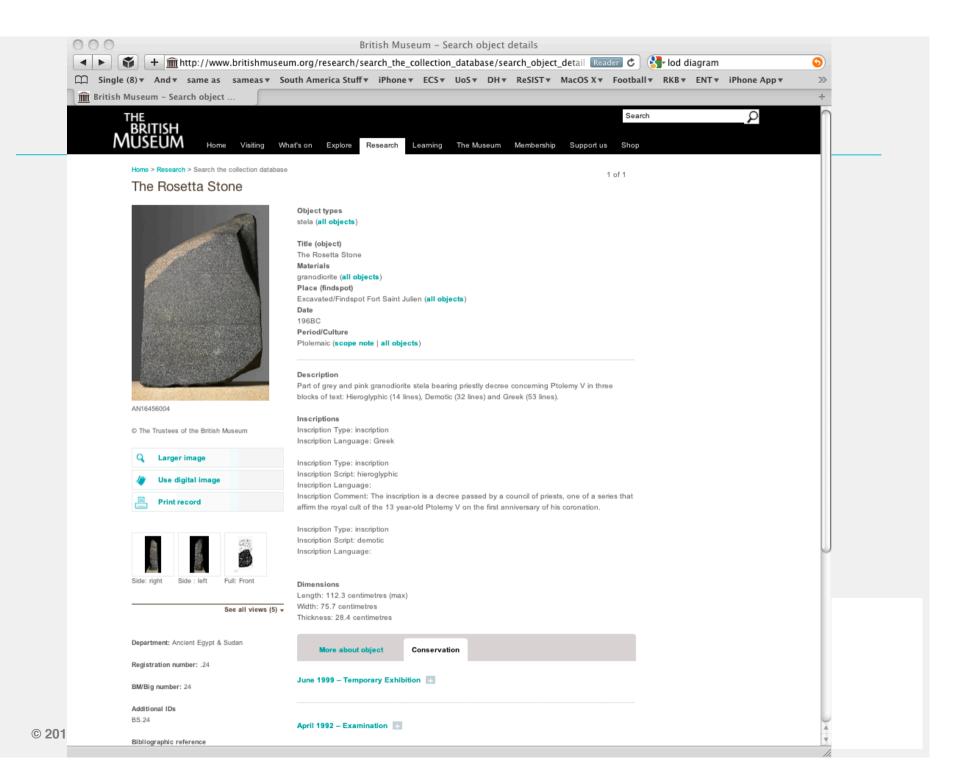




## **Product: British Museum Integration**

- http://www.britishmuseum.org/research/collection\_online/ collection\_object\_details.aspx?objectId=117631&partId=1
- Linked data used to connect three separate internal systems
- Enables research and conservation data to be viewed and published live into BM website

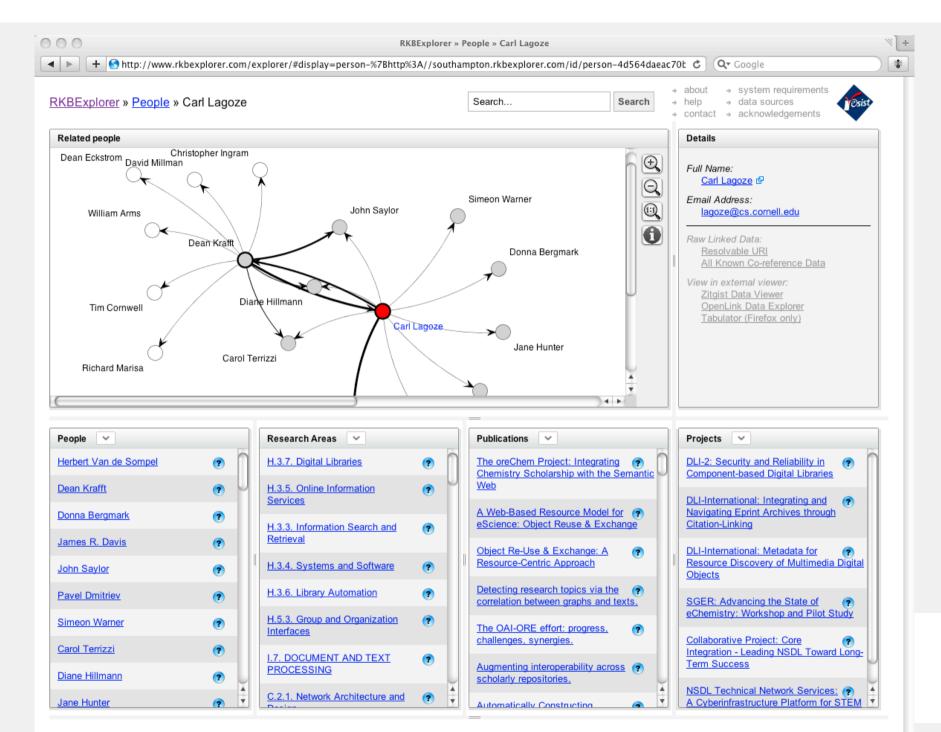


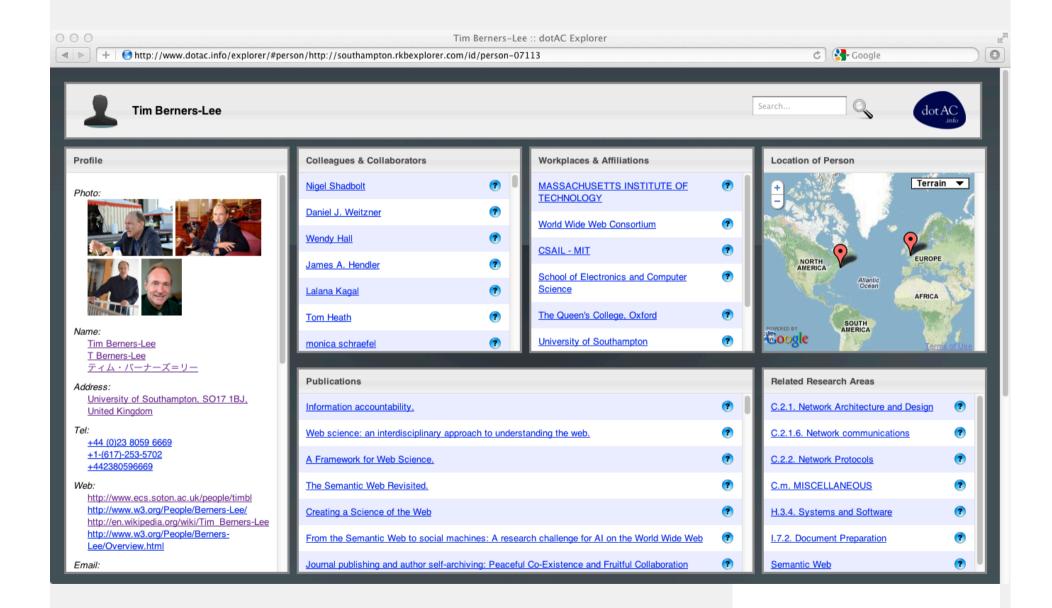


## **Products: RKBExplorer and dotAC**

- http://www.rkbexplorer.com/
- http://www.dotac.info/explorer/
- Members of the Seme4 team created many of the first Open Linked ked Data resources, detailing researchers, publications and projects
- The infrastructure hosts many individual Linked Data datasets, ts, helping define emerging best practices
- Addresses problems of equivalence mapping and distributed querying across multiple sources
- Applications give synergistic view over data, highlighting which resources are related, and (equally importantly) why they are related







# **Demo: Genealogy Viewer**

- http://genealogy.ragld.com/
- Rapid application to visualise family tree
- Client-side javascript / ajax web application
- Server-side uses only stock services s
- Location data included from Wikipedia



## **RAGLD Genealogy Browser**

### P Edith Nora Brandham

### **Edith Nora Brandham**



### **Edith Nora Brandham**

Born: 1912-09-02

Windmill Road, Flitwick, Beds

Died: 2007-08-06 Bedford Hospital

Life span: 95 years

Married: Reginald Kenneth Goodwin

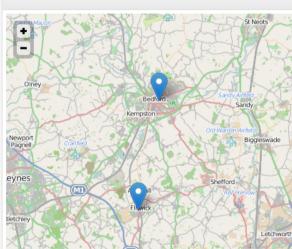


Parents



Lizzie Hill

### Locations





Arthur Ralph



### Grandparents



Ansell







# Brandham

Ernest Brandham

### Siblings



Brandham

### Children

## http://gothwin.data.seme4.com/id/person/I0276

**Edith Nora Brandham** 

Linked Data Browser

ancestors Amy Turney, Ann Osborn, Ann Tearle, Ann Unknown, Annie

Ansell, Elizabeth Unknown, Ernest Brandham, George Hinson, Henry Hill, Jeremiah Brandam, Jeremiah Brandham, Jeremiah Brandham, Job Turney, John Ansell, Lizzie Hill, Mary Unknown, Sarah Hinson, Thomas Tearle, William Brandham (and 3 others)

children David Roger Goodwin descendants Daniel Edward Goodson, David Roger Goodwin, Jane Elizabeth

Goodwin, John David Goodwin and Lucy Jane Goodson

event 372 and 413

family unit F085 and F093

Brandham

gender female

family name

Edith Nora given name

grandchildren Jane Elizabeth Goodwin and John David Goodwin

grandparents Amy Turney, Annie Ansell, Henry Hill **and** Jeremiah Brandham

label Edith Nora Brandham

Edith Nora Brandham name

# Demo: Ysgolion – tell me about schools

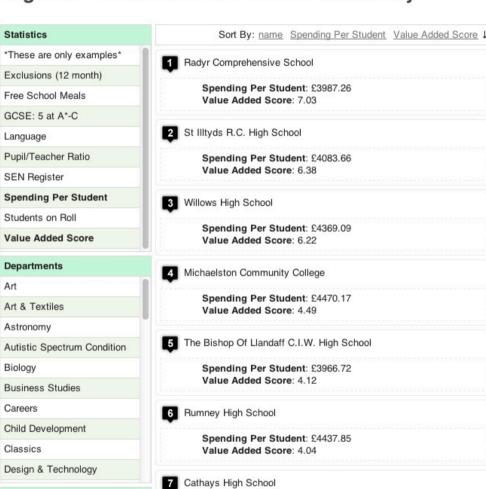
- http://ysgolion.com
- "Ysgolion" application (meaning "schools")
- Combines data from a number of disjoint government departments about secondary schools in the Cardiff area
- Application enables users to compare statistics, subjects offered and deperformance of different schools within the area
- Contrasts these with public sector indices of multiple deprivation etc. for each region of Cardiff







## Ysgolion - Welsh Schools: Cardiff Secondary



Spending Per Student: £4465.48

Spending Per Student: £4173.31 Value Added Score: 2.01

Value Added Score: 3.06

8 Cardiff High School



### News feed

## Mary Immaculate High School

14 Jan 2012, 8:38 am

Thank you to Yr7 Families feedback from Parents Evening on Wednesday. 97% believe the school provides an 'Excellent' quality of education.

### Mary Immaculate High School

14 Jan 2012, 8:36 am

Congratulations to the Yr11 students who sat linear Maths in November. 56% of those entered achieved their target and now passed! Well done

### Awards

Active Marc

ASDAN

CAB

Cardiff Arts Mark

Careers Wales

Cicco Notworking Academy

# **Demo: Hampshire Open Data**

http://hampshire.data.seme4.com/services/visualisation/

- Rapidly Assembled (one person day)
- Combination of many different datasets
- Catchment areas data from multiple sources



