Open Data: Powering the Information Age

Professor Nigel Shadbolt FREng
Electronics and Computer Science
University of Southampton
UK Open Data Advisor and Transparency Board
Twitter @Nigel_Shadbolt



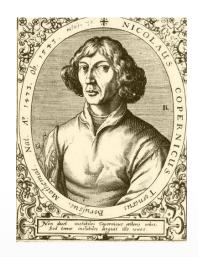
Open Data: Powering the Information Age

Professor Nigel Shadbolt FREng
UK Open Data Advisor and Transparency Board
Twitter @Nigel Shadbolt

28th March 2012



Data and the rise of science...



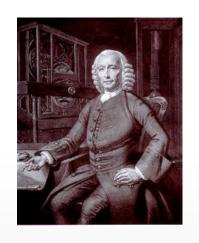
- Copernicus researching from 1497 - 1543
- De revolutionibus orbium coelestium
- Theoretical and empirical reformulation of planetary motion





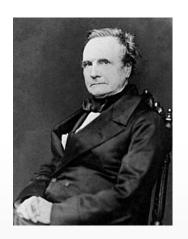


Data and the ascent of engineering...







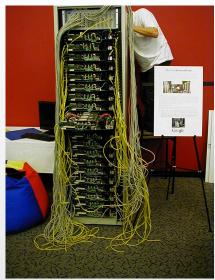






The Science and Engineering of Data...





$$I^0 = c_1 v_1 + c_2 v_2 + \ldots + c_n v_n$$

$$I^{1} = \mathbf{S}I^{0} = c_{1}v_{1} + c_{2}\lambda_{2}v_{2} + \ldots + c_{n}\lambda_{n}v_{n}$$

$$I^{2} = \mathbf{S}I^{1} = c_{1}v_{1} + c_{2}\lambda_{2}^{2}v_{2} + \ldots + c_{n}\lambda_{n}^{2}v_{n}$$

$$\vdots \qquad \vdots$$

$$I^{k} = \mathbf{S}I^{k-1} = c_{1}v_{1} + c_{2}\lambda_{2}^{k}v_{2} + \ldots + c_{n}\lambda_{n}^{k}v_{n}$$

$$\mathbf{G}I^k = lpha \mathbf{H}I^k + lpha \mathbf{A}I^k + rac{1-lpha}{n} \mathbf{1}I^k$$

The Power of Data...

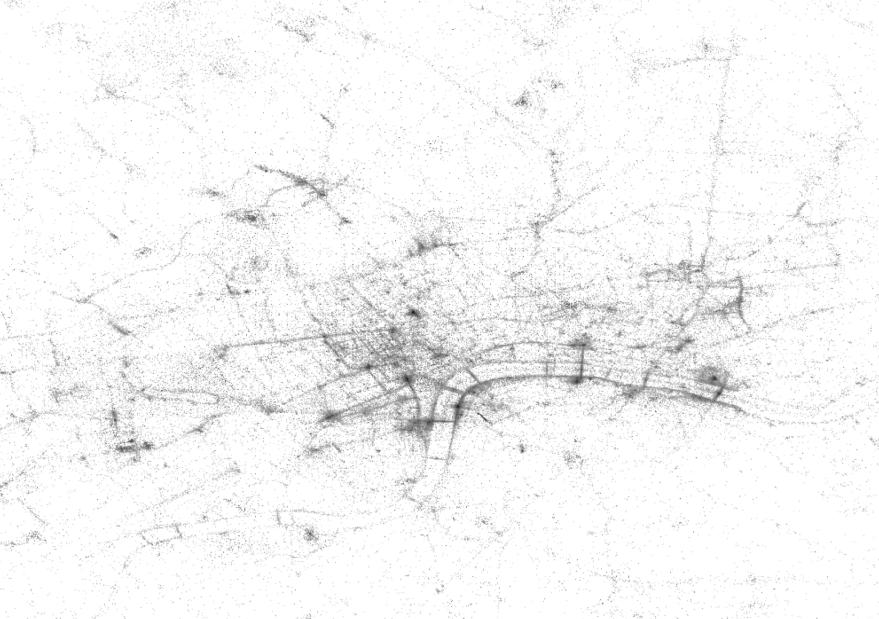
- Halevy et al Intelligent SystemsMarch/April 2009 (vol. 24 no. 2) pp. 8-12
- Large scale, comprehensive data
- New forms of research alliance
- New forms of direct data access



Search query topic	Top 45 queries		Next 55 queries	
	n	Weighted	n	Weighted
Influenza complication	11	18.15	5	3.40
Cold/flu remedy	8	5.05	6	5.03
General influenza symptoms	5	2.60	1	0.07
Term for influenza	4	3.74	6	0.30
Specific influenza symptom	4	2.54	6	3.74
Symptoms of an influenza complication	4	2.21	2	0.92
Antibiotic medication	3	6.23	3	3.17
General influenza remedies	2	0.18	1	0.32
Symptoms of a related disease	2	1.66	2	0.77
Antiviral medication	1	0.39	1	0.74
Related disease	1	6.66	3	3.77
Unrelated to influenza	0	0.00	19	28.37
Total	45	49.40	55	50.60

The top 45 queries were used in our final model; the next 55 queries are presented for comparison purposes. The number of queries in each topic is indicated, as well as query-volume-weighted counts, reflecting the relative frequency of queries in each topic.

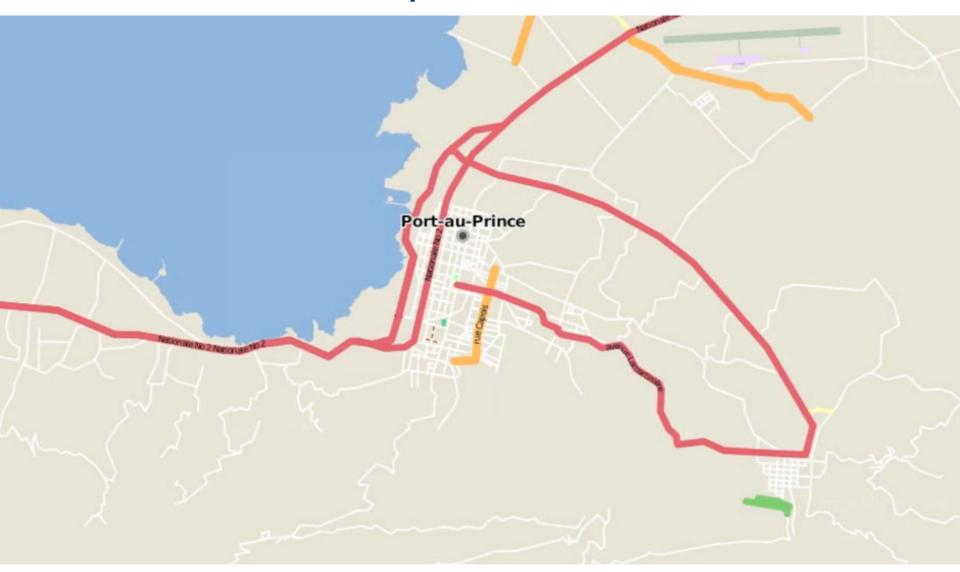




London, via Flickr (Crandall et al 2009)



The Power of People



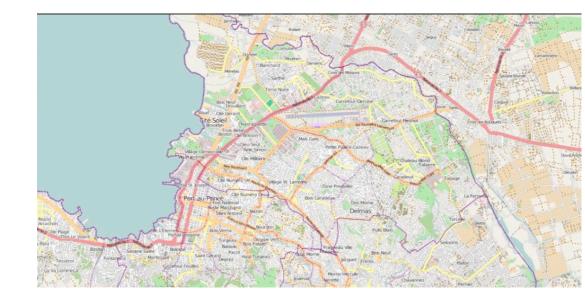
The Power of Open

- Open Standards
- Open Licences
- Open Data
- Open Source

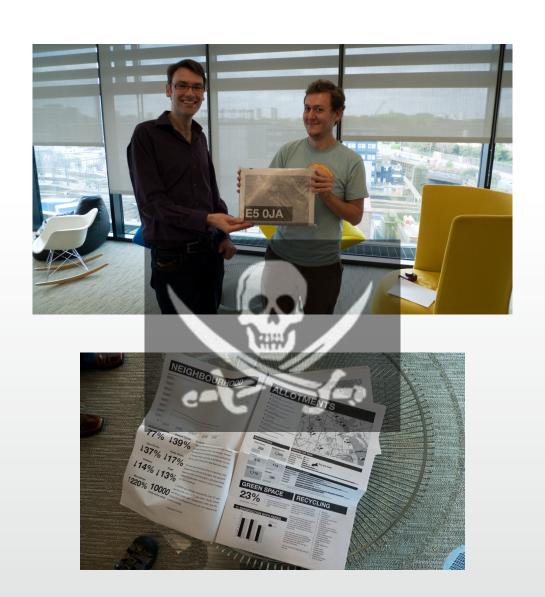


The Power of Open

- Open Standards
- Open Licences
- Open Data
- Open Source



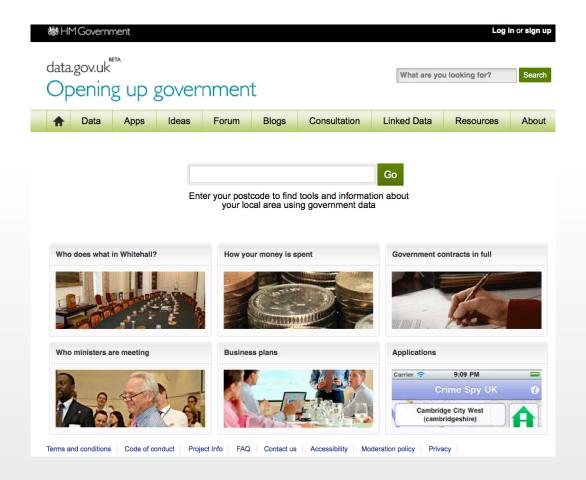
From this....



to this in three months



to this in 24 months...



Beginnings in UK



University of Southampton project to show how open data could deliver services across Government (2005-06)





























Reported to UK Parliament





The United Kingdom Implementation of the European Directive on the Re-use of Public Sector Information

the first two years

AKTive PSI: Leading by Example

- 4.20 The UK public sector is a source of rich, high quality and sought after data. While much of this information is published and available for re-use by others, it is often trapped by poor data structures, locked up in legacy data formats or in fragmented databases.
- 4.21 To explore the issues more fully, in 2005-6, OPSI worked with Advanced Knowledge Technologies (AKT),³⁶ an inter-disciplinary research project led by the University of Southampton. OPSI's work with AKT, in a research project called AKTive PSI, had two aims:
 - to raise awareness about and disseminate the capabilities of semantic web technologies amongst government departments, agencies and local authorities;
 - to show what is possible using this technology.
- 4.22 OPSI brought together a diverse collection of public sector information assets to experiment with. A number of public sector organisations were involved in the project, including Ordnance Survey, the Met Office, the Department for Communities and Local Government, the Office for National Statistics, the Department for Environment, Food and Rural Affairs, the Environment Agency and the London Boroughs of Camden and Lewisham. The project underlined the potential for the use of semantic web technology in large scale integration of public sector information and the benefits such aggregation would bring. Semantic Web technology provides the best model for a range of interoperability issues. If widely adopted it would do much to harness the re-use of public sector information.
- 4.23 AKTive PSI has spawned further work in government using Semantic Web technology³⁷. OPSI is using this technology in the following ways:

A Report by the Office of Public Sector Information

July 2007

The Power of People



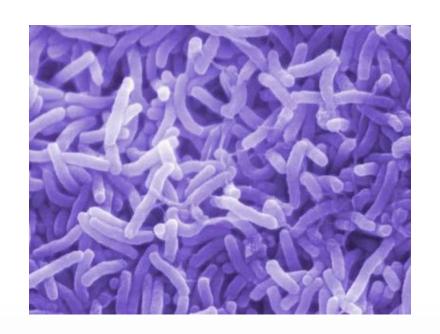
















theguardian

News | Sport | Comment | Culture | Business | Money | Life & style

News Society MRSA and superbugs

DATABLOG

Facts are sacred

Previous

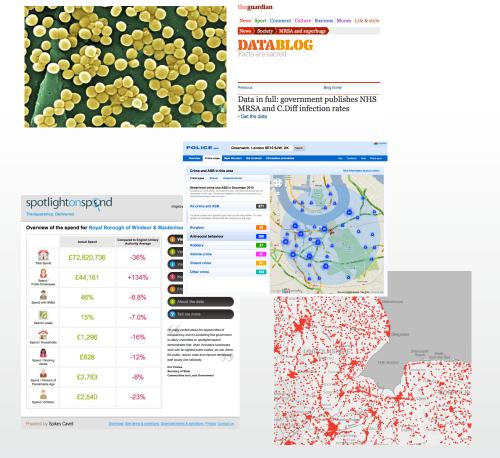
Blog home

Data in full: government publishes NHS MRSA and C.Diff infection rates

Get the data

Open Government Data – reasons

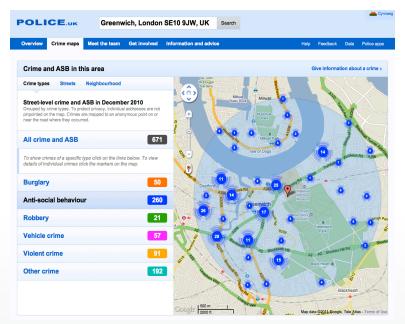
- Transparency
- Accountability
- Efficiency
- Public Service Delivery
- Engagement
- Data Improvement
- Society value
- Economic value



open the data...

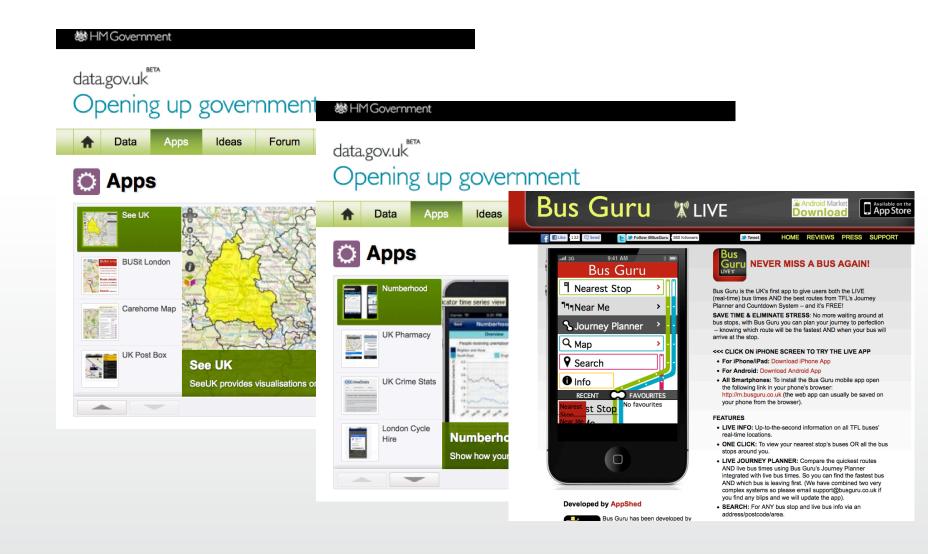




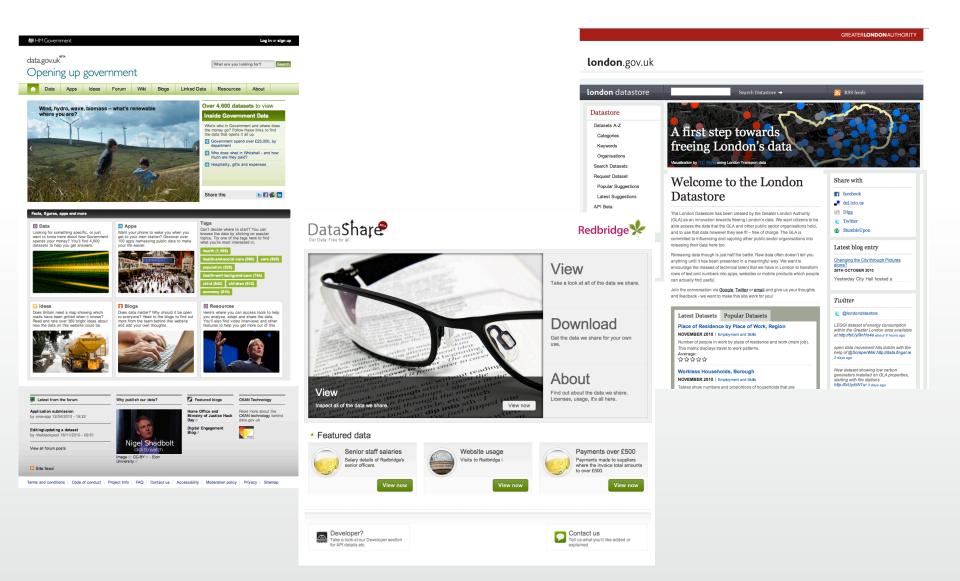




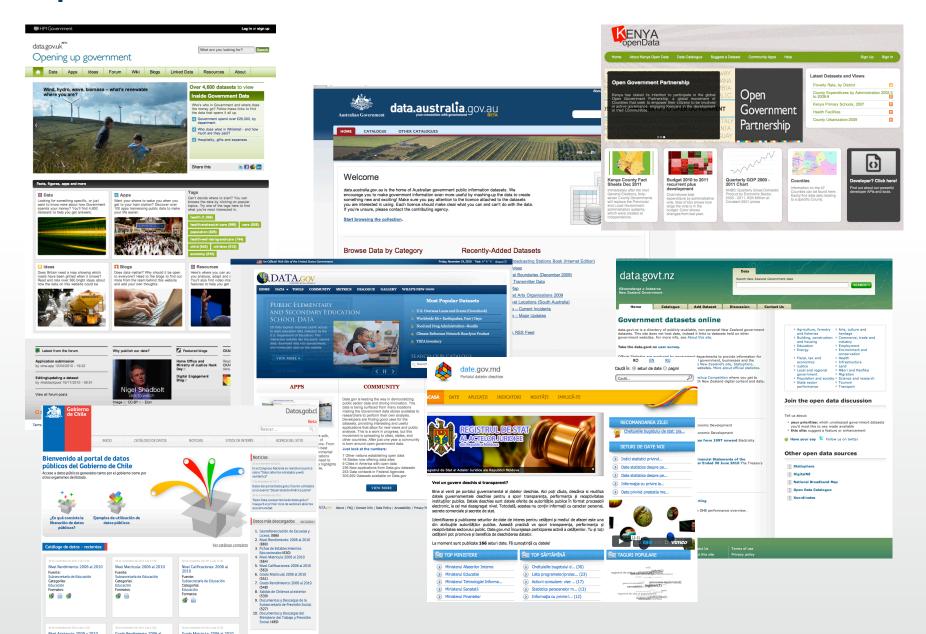
...and the applications follow



Open Government Data: at all scales

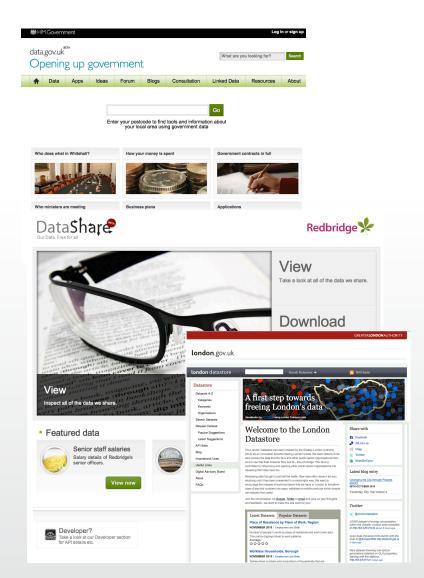


Open Government Data: around the world...

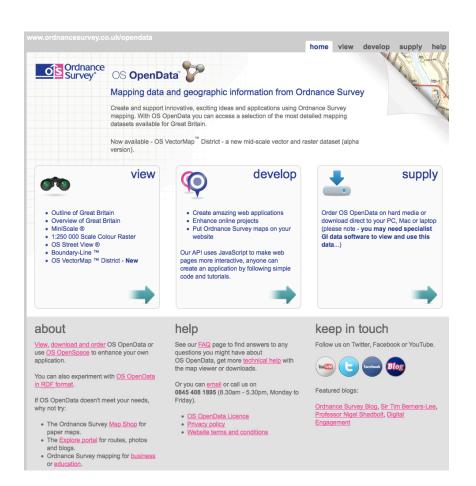


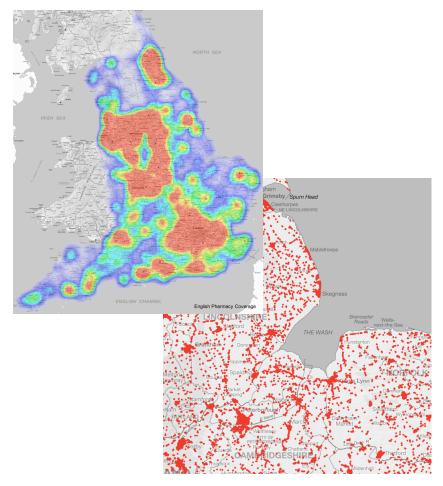
Open Data achievements

- National, Regional and City portals launched
- Significant data sets released
- Public Data Principles
- Open Licences
- Developer communities
- Reviews and consultation
- International collaboration



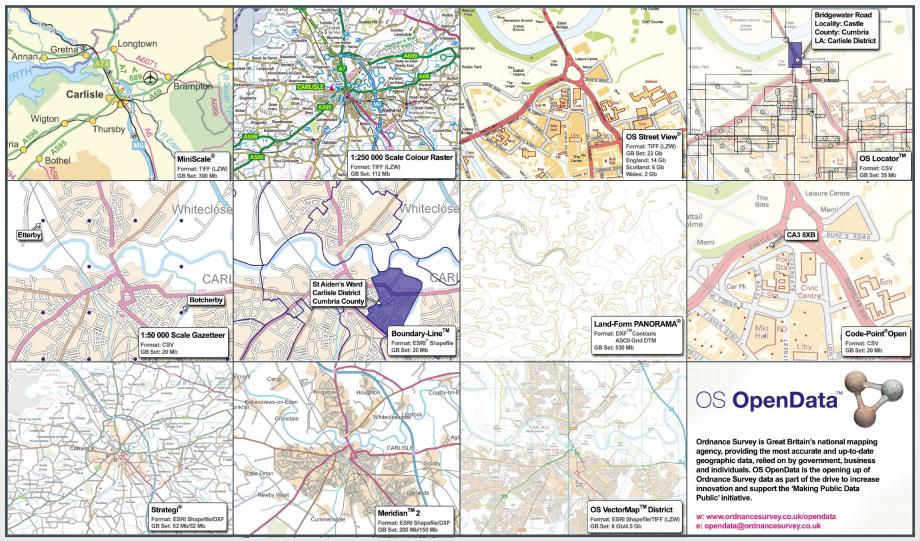
Location, location, location...



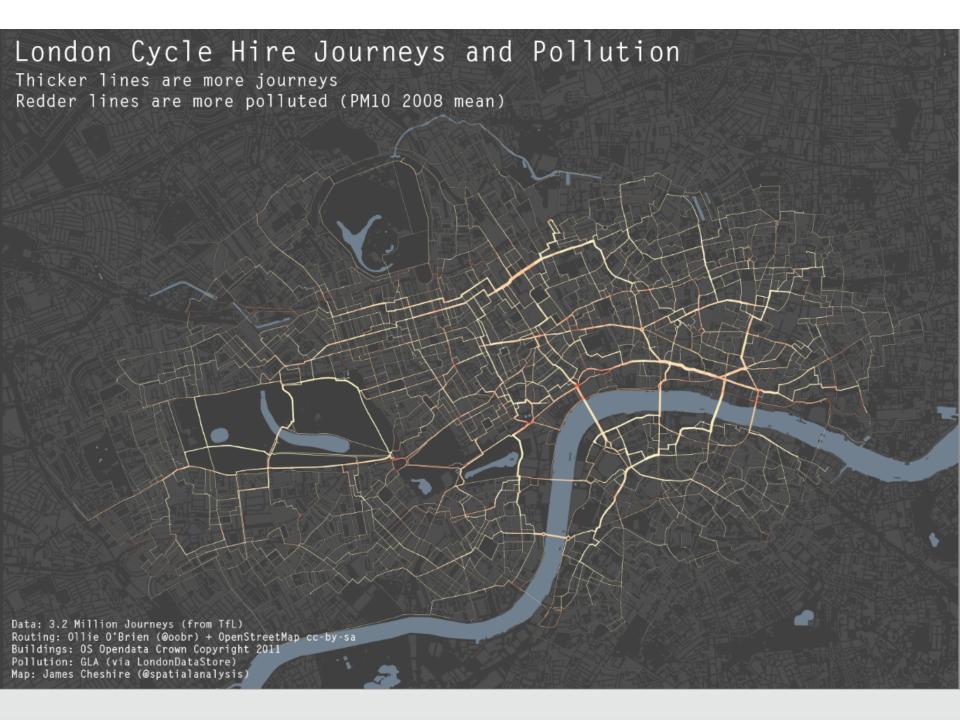




www.ordnancesurvey.co.uk/opendata



See also: Linked data http://www.ordnancesurvey.co.uk/opendata/linkeddata.html





UK Crime Heatmap for SO17 1BJ Waythli Copperfield Road Ethelburt Avenue Willis Road Swaythii Flowers Estate Wesse Hampton Park Hamson Ro lighfield Southampton Common Θ Θ 0 UNE DI ROSE

UK Crime Heatmap for SO17 1BJ Waythl Coppertield Road Θ Ethelburt Avenue Willis Road Swaythii Flowers Estate Θ Θ Hampton Park lighfield Southampton Common Θ Θ 0 THE THOOP

UK Crime Heatmap for SO17 1BJ Wayth Copperfield Road Ethelburt Avenue Θ Flowers Estate θ Hampton Park Hamson Ro lighfield Southampton Common Θ Θ 0 LINE OF ROOM

UK Crime Heatmap for SO17 1BJ Leaside Way Bassett Green Roa Swaythling Copperfield Road Θ Flowers Estate 0 Hampton Park Hamson Highfield Southampton Common Θ 0 WOOD A ROAD

UK Crime Heatmap for SO17 1BJ Vayth (Copperfield Road Ethelburt Avenue Θ Flowers Estate 0 Θ 0 lighfield Southampton Common Θ 0 LINE OF ROOM

UK Crime Heatmap for SO17 1BJ Wayth Copperfield Road Ethelburt Avenue Willis Road Swaythii 0 Flowers Estate Θ Θ Hampton Park O P lighfield Southampton Common Θ 0 Θ Paged A

UK Crime Heatmap for SO17 1BJ 0 Wayth Copperfield Road Ethelburt Avenue Flowers Estate Θ Hampton Park 0 lighfield Southampton Common Θ 0 NOOD A. WOOD

UK Crime Heatmap for SO17 1BJ Wayth Copperfield Road Ethelburt Avenue Willis Road Swaythii Flowers Estate Θ Θ Hampton Park Hamson O P lighfield Southampton Common Θ 0 Θ THE THE ROBERT

UK Crime Heatmap for SO17 1BJ Leaside Way Bassett Green Roa A Swaythling Θ Coppertield Road Willis Road Swaythii 0 Flowers Estate Θ Θ Hampton Park Hamson 0 Highfield Southampton Common Θ 0 LINE OF BOOM

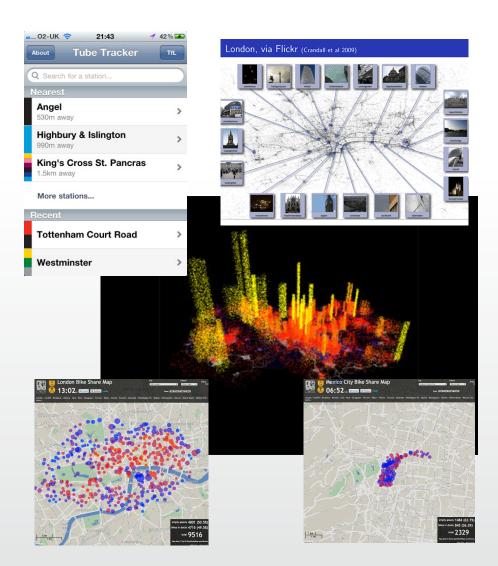
UK Crime Heatmap for SO17 1BJ Wayth Copperfield Road Ethelburt Avenue Θ Θ Flowers Estate 0 Θ Hampton Park 0 Highfield Southampton Common Θ 0 UNE DI ROSE

UK Crime Heatmap for SO17 1BJ Wayth Ethelburt Avenue Flowers Estate Θ Hampton Park 0 Highfield Southampton Common Θ Θ 0 .wood he.

UK Crime Heatmap for SO17 1BJ vayth Copperfield Road Ethelburt Avenue Flowers Estate Θ Hampton Park Hamson Re 0 Highfield Southampton Common Θ Θ • WOOD A. WOOD

Open Data and Smart Cities

- Smart Cities get Open Data
- Many of the best Open Data apps are urban
- Authority over their data
- Network effect of urban data

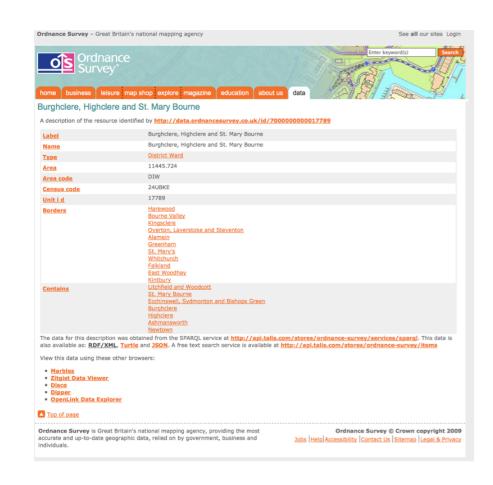


Linked Open Data – Aiming for the Stars

★ Put your data on the Web with an Open Licence
★★
★★
Use open, standard formats
★★★
Use URIs to identify things
★★★
★★★
Link your data to other people's data

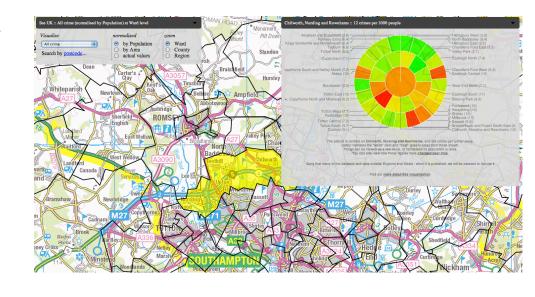
5★ Linked Open Data

- With data.gov.uk a national digital infrastructure being built
- URIs for schools, roads, bus stops, post codes, admin boundaries...
- Some of the data links across and connects other data together
- Key data link points exist

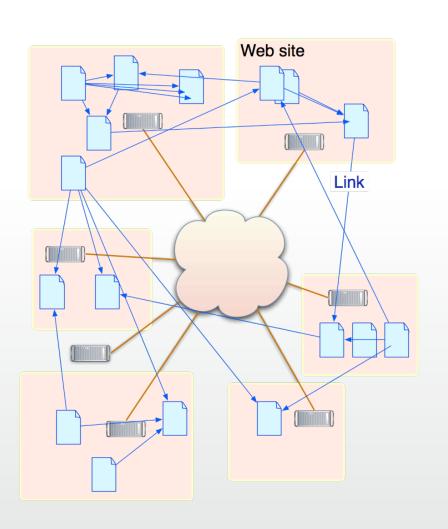


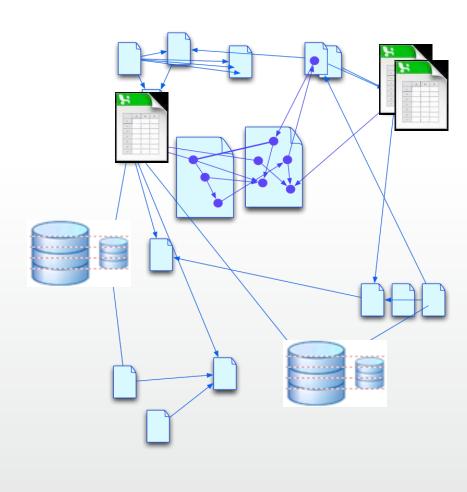
5★ Linked Open Data

- With data.gov.uk a national digital infrastructure being built
- URIs for schools, roads, bus stops, post codes, admin boundaries...
- Some of the data links across and connects other data together
- Key data link points exist



The Web ... of documents and data





Linked Data Principles

The four micro principles of the Semantic Web

- 1. All entities of interest, such as information resources, real-world objects, and vocabulary terms should be identified by URI references.
- 2. URI references should be dereferenceable, meaning that an application can look up a URI over the HTTP protocol and retrieve RDF data about the identified resource.
- 3. Data should be provided using the RDF/XML syntax
- 4. Data should be interlinked with other data.

The Semantic Web Revisited

Nigel Shadbolt and Wendy Hall, University of Southampton Tim Berners-Lee, Massachusetts Institute of Technology



Nigel Shadbolt is a professor of artificial intelligence in the School of Electronics and Computer Science at Southampton University, Contact him at nrs@ecs. soton.ac.uk.



Tim Berners-Lee is the director of the World Wide Web Consortium, a senior researcher at the Massachusetts Institute of Technology's Computer Science

gence Laboratory, and a professor of computer science in the Department of Electronics and Computer Science at Southampton University, Contact him at timbl@w3.org.



Wendy Hall is a professor of computer science in the School of Electronics and Computer Science at Southampton University, Contact her at wh@ecs.soton. ac.uk.

Principle 1 URIs for Everything

- Uniform Resource Identifiers for all items of interest
- You can dereference them
 use http protocol to get
 data back
- Using a simple Web
 Knowledge Representation
 Language RDF

http://rdf.ecs.soton.ac.uk/person/2686

http://rdf.ecs.soton.ac.uk/project/464

http://rdf.ecs.soton.ac.uk/publication/11065

http://education.data.gov.uk/doc/school/120805

http://southampton.rkbexplorer.com/id/person-02686

http://dbpedia.org/resource/Nigel_Shadbolt

Link URIs together

URIs will dereference Principle 2

- Uniform Resource Identifiers for all items of interest
- You can dereference them use http protocol to get data back
- Language RDF
- http://id.ecs.soton.ac.uk/role/2686 member Of http://id.ecs.soton.ac.uk/project/221 http://id.ecs.soton.ac.uk/project/293 http://id.ecs.soton.ac.uk/project/312 http://id.ecs.soton.ac.uk/project/353 http://id.ecs.soton.ac.uk/project/361 http://id.ecs.soton.ac.uk/project/381 http://id.ecs.soton.ac.uk/project/395 http://id.ecs.soton.ac.uk/project/443 http://id.ecs.soton.ac.uk/project/44 http://id.ecs.soton.ac.uk/project/45 http://id.ecs.soton.ac.uk/project/463 http://id.ecs.soton.ac.uk/project/464 http://id.ecs.soton.ac.uk/project/465 http://id.ecs.soton.ac.uk/project/466 http://id.ecs.soton.ac.uk/project/467 http://id.ecs.soton.ac.uk/project/508 http://id.ecs.soton.ac.uk/project/623 http://id.ecs.soton.ac.uk/project/629 http://id.ecs.soton.ac.uk/project/630 Person Using a simple Web Shadbolt family name Nigel R givenname http://users.ecs.soton.ac.uk/nrs/ homepage http://www.ecs.soton.ac.uk/image.php?id=person_268 Knowledge Representation http://www.ecs.soton.ac.uk/image.php?id=person_268 amp;checksum=6c2158b33a0c10e328ec62f9ed1cc73 mbox mailto:nrs@ecs.soton.ac.uk Professor Nigel R Shadbolt

Shadbolt

Nigel Shadbolt

Southampton ECS People: Professor Nigel R Shadbolt

Description

Date Created

Appellation Family Name

Full Name

Given Name

Source

label

This rdf document contains information about a perso

Southampton ECS People: Professor Nigel R Shadbo

Southampton ECS People: Professor Nigel R Shadbo

http://www.ecs.soton.ac.uk/people/nrs

Southampton.

Ontology

Professor

Shadbolt

Nigel R

2009-11-26T10:53:20Z

Professor Nigel R Shadbolt

Southampton ECS People: Professor Nigel R

Link URIs together

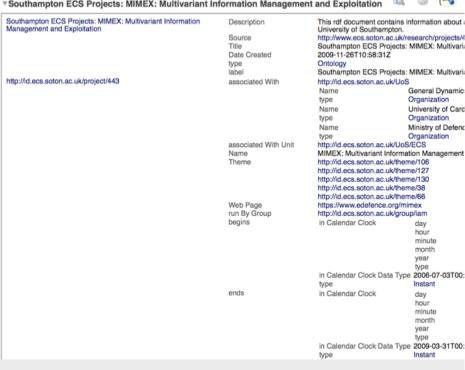
Get RDF Back Principle 3

- Uniform Resource Identifiers for all items of interest
- You can dereference them use http protocol to get data back
- Using a simple Web **Knowledge Representation** Language - RDF
- Link URIs together







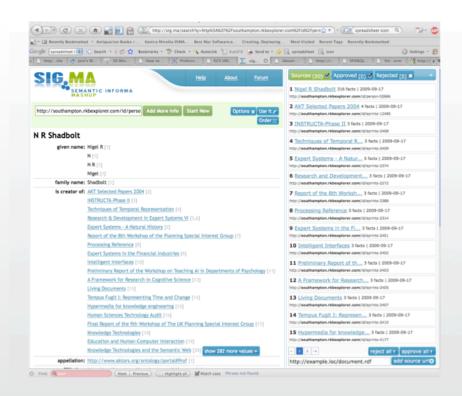


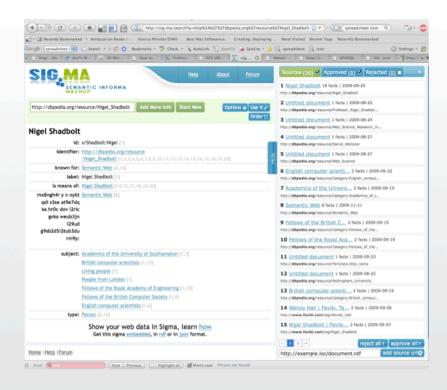
Principle 4 Link Data URIs

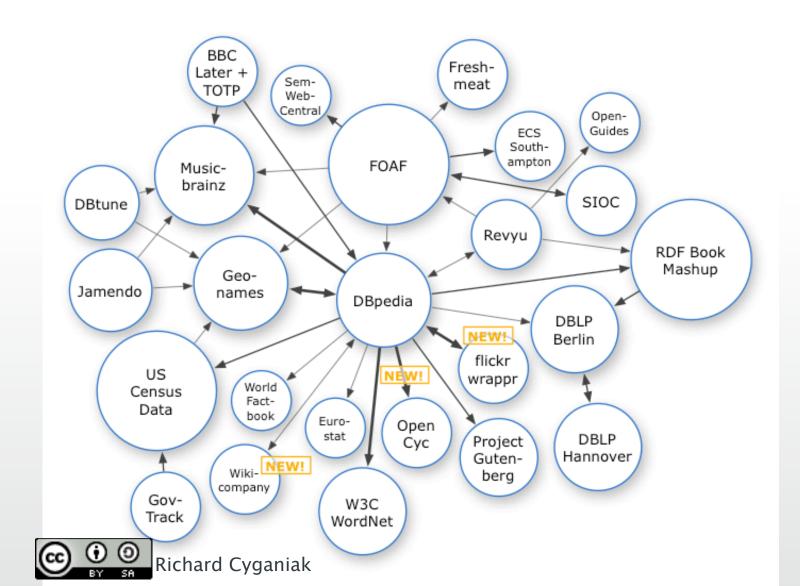
http://rdf.ecs.soton.ac.uk/person/2686 SameAs

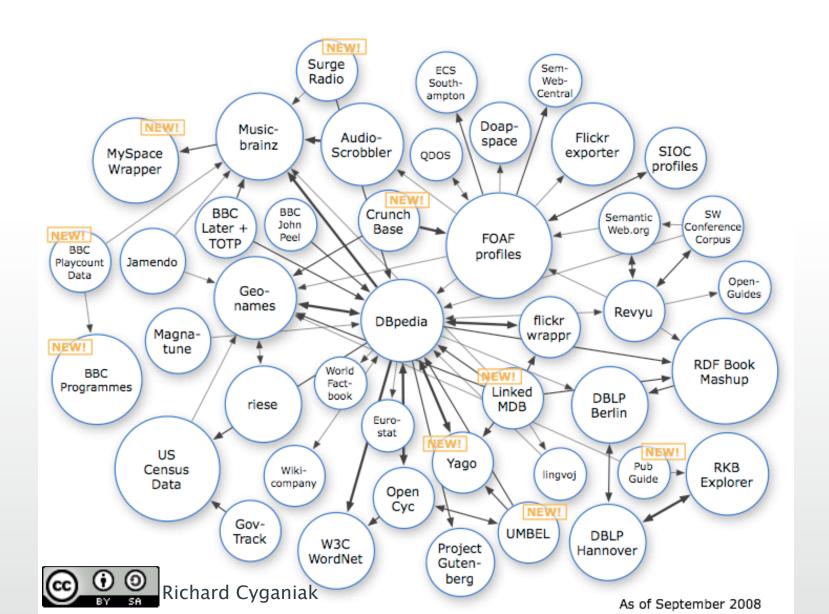
http://southampton.rkbexplorer.com/id/person-02686 SameAs

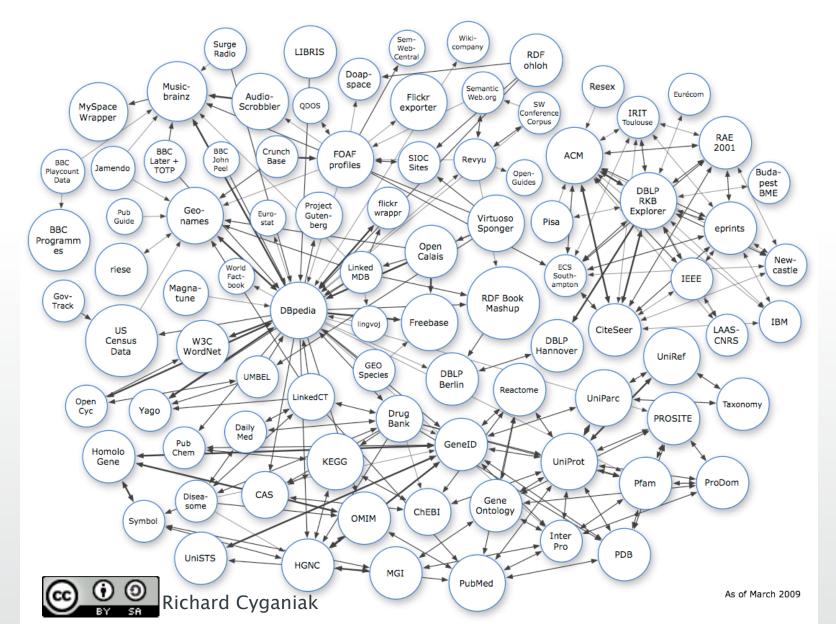
http://dbpedia.org/resource/Nigel_Shadbolt

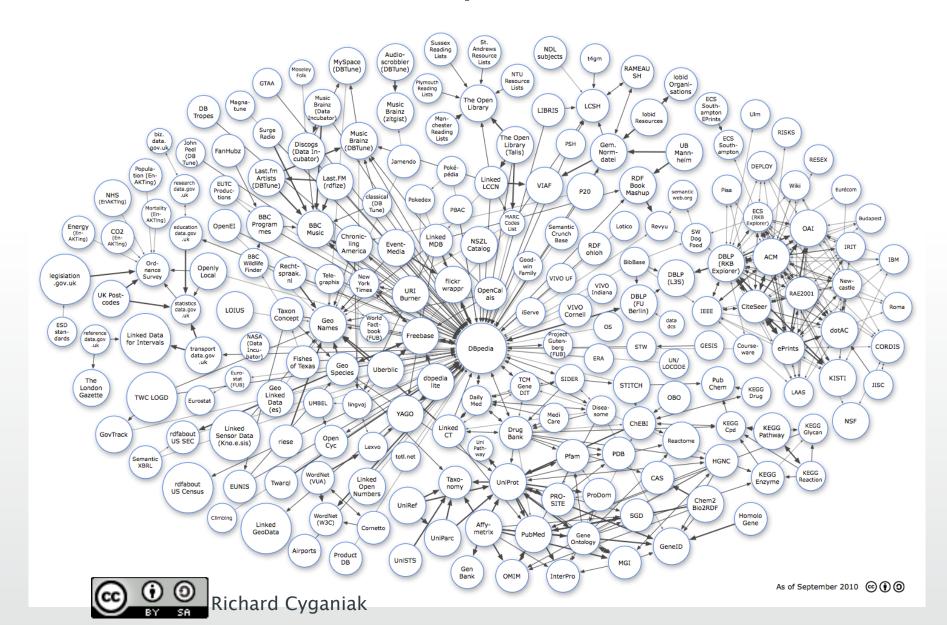


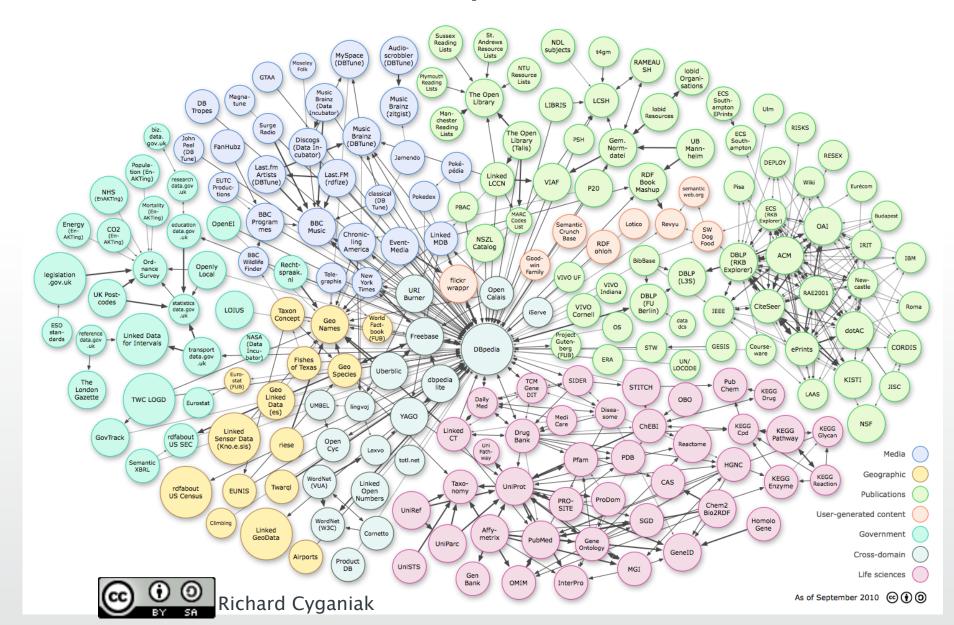




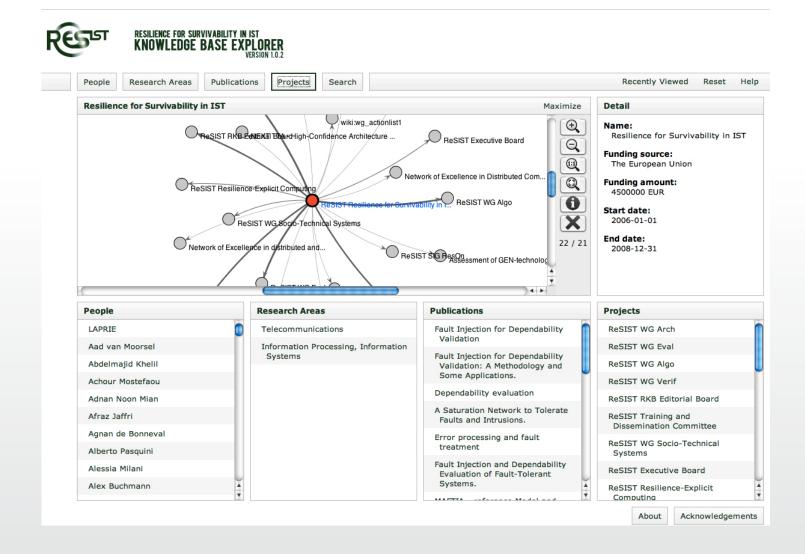






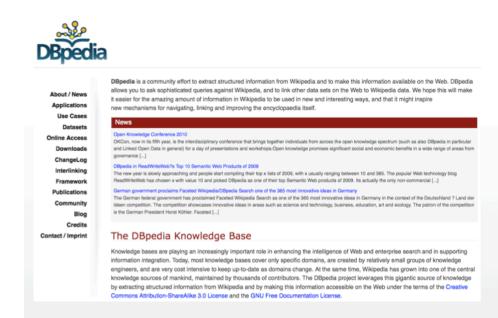


Linked Data Application – RKB Explorer



Wikipedia as Linked Data - DBpedia

- 2,900,000 things
- >282,000 persons
- 339,000 places
- 130,000 species
- 4,400 diseases
- 88,000 music albums



•

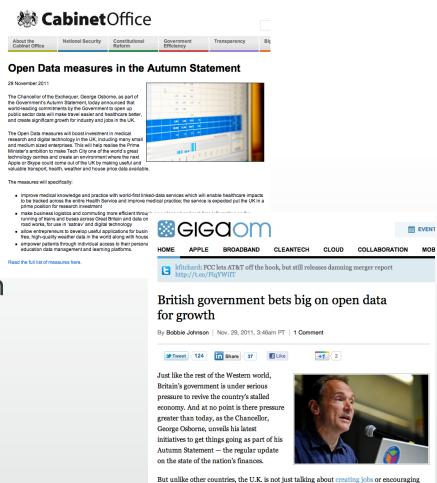
Adding a little SPARQL to the Web

- A data access language for the Web of Linked Data
- Can query across diverse data sources
- SPARQL can query required and optional patterns

country_name	population
Ethiopia	82825000
Uganda	32710000
Nepal	29331000
Afghanistan	28150000
Uzbekistan	27606007
Burkina Faso	15757000
Niger	15290000
Malawi	15263000

Latest UK HMG Developments

- Significant new data releases
 - Weather
 - Transport
 - Health
 - **—** :
- Open Data and its role in growth and innovation
- Open Data Institute
- Located Tech City, Shoreditch, London



infrastructural projects: it's also taking a gamble on data.

Open Data Challenges

- Infrastructure
- Quality
- Interpretation
- Security and Privacy
- Open Data 2.0 ...





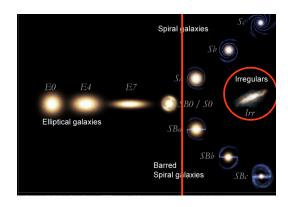




OGD 2.0 – Crowd Sourcing Open Data

- Closing the feedback loops
- Publish out
- Acquire & Enhance
- Write Back

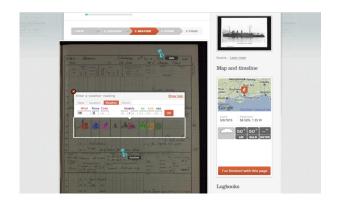




OGD 2.0 – Crowd Sourcing Open Data

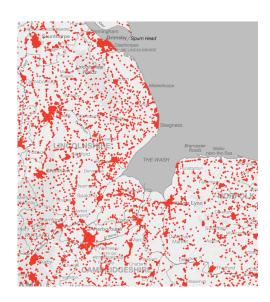
- Closing the feedback loops
- Publish out
- Acquire & Enhance
- Write Back





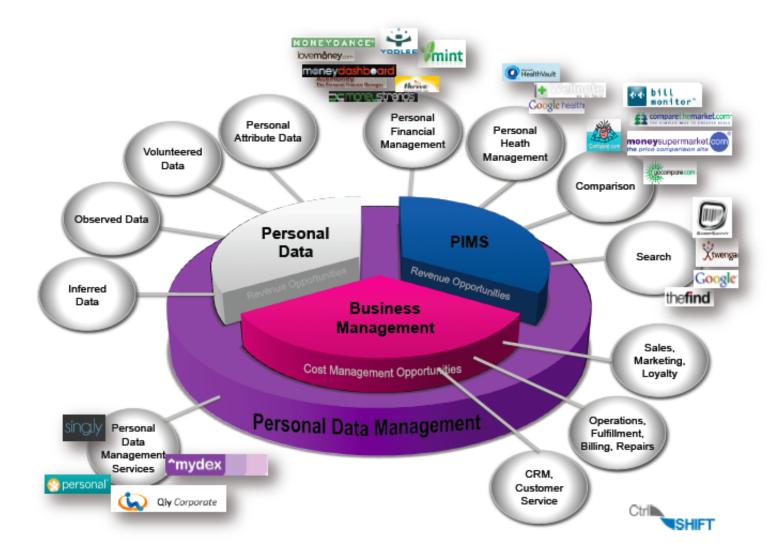
OGD 2.0 – Crowd Sourcing Open Data

- Closing the feedback loops
- Publish out
- Acquire & Enhance
- Write Back





Consumers' own data open to themselves...

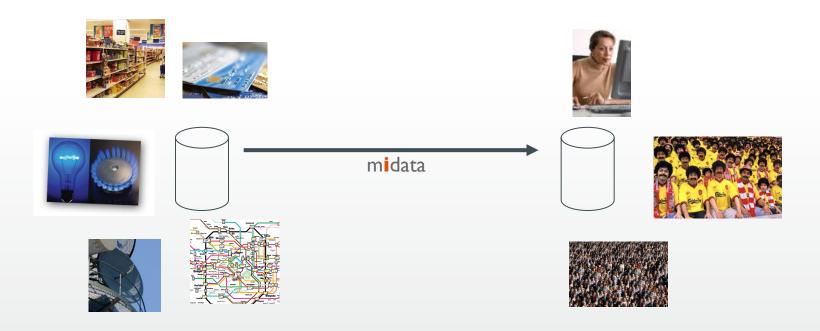


B S Department for Business Innovation & Skills

midata



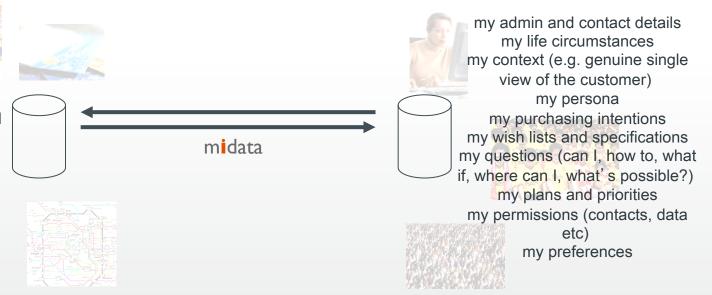
Step 1: A precondition for...



Step 2: New interactions...

Technical product attributes
Health & Safety
Supply chain responsibility
service, repair, problem solving
warranties, insurances
'Other people like you' data and
advice (purchases, behaviours,
problems)
Basic admin data
Contact details
Transaction history
Interaction history
Online behavioural data

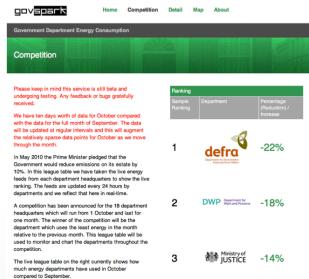
New products and services
Offers and incentives



Linked Open Data OTW – a platform for modern government

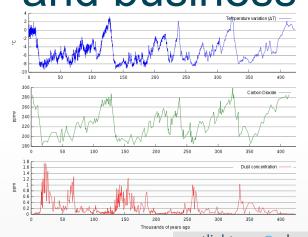


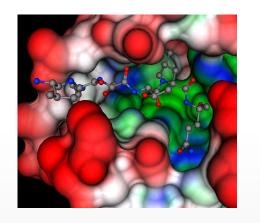






Linked Open data OTW – a platform for modern science, engineering, education and business







Download Site terms & conditions Download terms & conditions Privacy Contact us

Powered by Spikes Cavell

