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
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...

\[ I^0 = c_1v_1 + c_2v_2 + \ldots + c_nv_n \]

\[ I^1 = SI^0 = c_1v_1 + c_2\lambda_2v_2 + \ldots + c_nv_nv_n \]

\[ I^2 = SI^1 = c_1v_1 + c_2\lambda_2^2v_2 + \ldots + c_nv_n^2v_n \]

\[ \vdots \]

\[ I^k = SI^{k-1} = c_1v_1 + c_2\lambda_2^kv_2 + \ldots + c_nv_n^kv_n \]

\[ GI^k = \alpha H^k + \alpha A I^k + \frac{1-\alpha}{n} \frac{1}{n} I^k \]
The Power of Data...

- Halevy et al. Intelligent Systems March/April 2009 (vol. 24 no. 2) pp. 8-12
- Large scale, comprehensive data
- New forms of research alliance
- New forms of direct data access
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)
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), 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 technology. OPSI is using this technology in the following ways:
The Power of People
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 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…
London Cycle Hire Journeys and Pollution

Thicker lines are more journeys
Redder lines are more polluted (PM10 2008 mean)
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
★★ Make it available as structured data
★★★ 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
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
- Link URIs together

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
Principle 2  URIs will dereference

• 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 3  Get RDF Back

- 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
The Web of Linked Open Data

Richard Cyganiak
The Web of Linked Open Data

Richard Cyganiak

As of September 2008
The Web of Linked Open Data
The Web of Linked Open Data
The Web of Linked Open Data

Richard Cyganiak
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

```sparql
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX type: <http://dbpedia.org/class/yago/>
PREFIX prop: <http://dbpedia.org/property/>
SELECT ?country_name ?population
WHERE {
  ?country a type:LandlockedCountries ;
  rdfs:label ?country_name ;
  prop:populationEstimate ?population .
  FILTER (?population > 15000000) .
}
```

<table>
<thead>
<tr>
<th>country_name</th>
<th>population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethiopia</td>
<td>82825000</td>
</tr>
<tr>
<td>Uganda</td>
<td>32710000</td>
</tr>
<tr>
<td>Nepal</td>
<td>29331000</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>28150000</td>
</tr>
<tr>
<td>Uzbekistan</td>
<td>27606007</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>15757000</td>
</tr>
<tr>
<td>Niger</td>
<td>15290000</td>
</tr>
<tr>
<td>Malawi</td>
<td>15263000</td>
</tr>
</tbody>
</table>
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
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…
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

midata

my admin and contact details
my life circumstances
my context (e.g. genuine single view of the customer)
my persona
my purchasing intentions
my wish lists and specifications
my questions (can I, how to, what if, where can I, what’s possible?)
my plans and priorities
my permissions (contacts, data etc)
my preferences
Linked Open Data OTW – a platform for modern government
Linked Open data OTW – a platform for modern science, engineering, education and business