Why Open Government Data Lessons from data.gov.uk

Professor Nigel Shadbolt FREng

UK Public Sector Transparency Board ChairUK CLG Local Public Data Panel

Twitter Nigel_Shadbolt

ISSWC, Shanghai, China
November 2010

Nigel Shadbolt
The Power of Open Data

Cholera cases

Bicycling traffic accidents
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

A Report by the Office of Public Sector Information
July 2007

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:
OGD takes hold

- Tim Berners-Lee and myself appointed 9th June 2009 and set about creating data.gov.uk
- Reappointed to UK Public Sector Transparency Board June 2010
- Open Government Data (OGD) is taking hold
- Governments, local authorities, cities releasing data
OGD takes hold
data.gov.uk

- Tim Berners-Lee and myself appointed 9th June 2009 and set about creating data.gov.uk
- Reappointed to UK Public Sector Transparency Board June 2010
- Open Government Data (OGD) is taking hold
- Governments, local authorities, cities releasing data
OGD takes hold
data.gov

- Tim Berners-Lee and myself appointed 9th June 2009 and set about creating data.gov.uk
- Reappointed to UK Public Sector Transparency Board June 2010
- Open Government Data (OGD) is taking hold
- Governments, local authorities, cities releasing data
OGD takes hold
London data store

- Tim Berners-Lee and myself appointed 9th June 2009 and set about creating data.gov.uk
- Reappointed to UK Public Sector Transparency Board June 2010
- Open Government Data (OGD) is taking hold
- Governments, local authorities, cities releasing data
OGD takes hold
Lichfield Local Authority

- Tim Berners-Lee and myself appointed 9th June 2009 and set about creating data.gov.uk

- Reappointed to UK Public Sector Transparency Board June 2010

- Open Government Data (OGD) is taking hold

- Governments, local authorities, cities releasing data
OGD takes hold
Windsor Local Authority

- Tim Berners-Lee and myself appointed 9th June 2009 and set about creating data.gov.uk

- Reappointed to UK Public Sector Transparency Board June 2010

- Open Government Data (OGD) is taking hold

- Governments, local authorities, cities releasing data
OGD takes hold
Redbridge Local Authority

- Tim Berners-Lee and myself appointed 9th June 2009 and set about creating data.gov.uk
- Reappointed to UK Public Sector Transparency Board June 2010
- Open Government Data (OGD) is taking hold
- Governments, local authorities, cities releasing data
• data.gov.uk itself (now with over 4000 datasets)
  – open source, open standards, open licence
  – key data sets released inc OS OpenData, COINS etc.
  – applications from outside HMG
• overcoming the many objections to transparency of data
• community of data users and developers
• political and technical leadership
• establishing public data principles
"Public Data" is the objective, factual, non-personal data on which public services run and are assessed, and on which policy decisions are based, or which is collected or generated in the course of public service delivery.
Public Data Principles

• Public data policy and practice will be clearly driven by the public and businesses who want and use the data, including what data is released when and in what form

• Public data will be published in reusable, machine-readable form

• Public data will be released under the same open licence which enables free reuse, including commercial reuse

• Public data will be available and easy to find through a single easy to use online access point (data.gov.uk)

• Public data will be published using open standards, and following relevant recommendations of the World Wide Web Consortium
Public Data Principles

- Public data underlying the Government’s own websites will be published in reusable form for others to use
- Public data will be timely and fine grained
- Release data quickly, and then re-publish it in linked data form
- Public data will be freely available to use in any lawful way
- Public bodies should actively encourage the re-use of their public data
- Public bodies should maintain and publish inventories of their data holdings
UK Coalition Government’s Commitments

- Prime Minister’s letter to Ministers on Transparency and Open Data - established Public Sector Transparency Board
- Creating a powerful new right to government data, enabling the public to request and receive government datasets
- Publishing data in open and standardised formats
- Transforming the culture of the public sector to one that presumes datasets should be open and shared with the public on an ongoing basis.
- Bringing in new measures to enable to public to scrutinise the government's accounts
- Requiring public bodies to publish online job titles, salaries and expenses of senior staff
- Publishing in full government contracts for good and services worth over £25,000 …
Making the argument
Early wins and eye-catching

ASBOrometer is a mobile application that measures levels of anti-social behaviour at your current location (within England and Wales) and gives you access to key ASB statistics.

ASBOrometer is available for iPhone and Android phones. Get it FREE from the iTunes App Store or Android Market now!

This app was created by Jeff Gifford and made possible by the data.gov.uk initiative, which is opening up UK government data for public reuse.
Making the argument
Early wins and eye-catching
Making the argument
Economic and Social Efficiency – Location…
Making the argument
Transparency and Accountablity – Spending…

A few years ago I helped the Department for International Development run a small international conference in Bangladesh, on the subject of migration. Its general conclusion was that migration is a good thing — remittances from migrant workers to their home countries far outweigh official aid, and, arguably, do more good.

I’m guessing that this is exactly the sort of jamboree the new government had in mind when, in opposition, it promised to open up the detailed ledgers of public spending. While I, and many real experts, might defend spending tens of thousands of taxpayer pounds promoting migration. It’s not hard to see what the popular press would have made of it.

In the new transparency age, ushered in today by the release of the first tranche of public data from the Treasury’s zealously guarded Coins (Combined Online Information System), the migration would have a very much ride instead.

In fact, in the current climate, the chances of funding it would be zero. Likewise, whole swathes of public spending that doesn’t immediately fall into the categories politicians like to boast about — hospitals on the beat, hospital beds, etc.

Despite this danger, I’m hearty in favour of opening Coins along with trimmed government spending data. It’s the latest stage in a revolution that was beyond our wildest dreams when Cherelle Arthur and I launched the Guardian’s Free Our Data campaign, four years ago.
Making the argument
Publish it…and the apps flow
Innovation
Making the argument Localism...

- New items of Local government spending over £500 – council by council from Jan 2011
Making the argument
Data to decisions - behaviour...

Real time energy data is changing behaviours – the decisions people make
Individual Gov Depts can take a lead

"Now it's our turn. I don't believe in one rule for councils and one for this Department. Central Government spends billions of tax pounds every year and transparency at this level of spending is just as important."

"This Department, like the rest of Whitehall, needs to look at where every penny is going and getting this data out in the open will help that process." Eric Pickles 12-8-10
And more to come…

- New local government contracts and tender documents for expenditure over £500 to be published in full from Jan 2011

- ...information on senior salaries, names and job descriptions, councillor allowances and expenses, minutes and papers of council meetings, job vacancies that will enable people to see why council wage bills are so high and how many of the positions are for key local services frontline service data - including rubbish and recycling rates, council tax collection rates and detail of major planned projects data such as food hygiene reports for food outlets - information which is routinely collected and of interest to residents, but not currently shared in an easily accessible format

- Crime data to be published at a level that allows the public to see what is happening on their streets from Jan 2011
Don’t Forget the Local Perspective
UK Local Data Panel: Current Activities

Working with new Transparency division on 3 key tasks

1. How to standardise local authority data and which data should be standardised as a priority

2. Assisting local authorities/public bodies to work through technical/ policy issues relating to making data openly available – in particular the Jan 2011 data release;

3. Encouraging citizens, developers, public and private sector organisations to access and exploit data provided by CLG and LAs
Don’t Forget the Local Perspective
UK Local Data Panel: Current Activities

Working with new Transparency division on 3 key tasks

1. How to standardise local authority data and which data should be standardised as a priority

2. Assisting local authorities/public bodies to work through technical/ policy issues relating to making data openly available – in particular the Jan 2011 data release;

3. Encouraging citizens, developers, public and private sector organisations to access and exploit data provided by CLG and LAs
Don’t Forget the Local Perspective
UK Local Data Panel: Current Activities

Working with new Transparency division on 3 key tasks

1. How to standardise local authority data and which data should be standardised as a priority

2. Assisting local authorities/public bodies to work through technical/ policy issues relating to making data openly available – in particular the Jan 2011 data release;

3. Encouraging citizens, developers, public and private sector organisations to access and exploit data provided by CLG and LAs

Appendix i: CSV file format

<table>
<thead>
<tr>
<th>Column</th>
<th>Field Name</th>
<th>What is Required</th>
<th>Reason For Inclusion</th>
<th>Additional Information</th>
<th>Inclusion Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Body</td>
<td>The statistical area UK that represents the authority, e.g. London Borough Council is 30001 or 30003 for an EPC rating or an electoral authority, e.g. E080081341.</td>
<td>To allow the file to be self-describing</td>
<td>Using a URI to identify the body, where that is uniquely identifiable to data users. The CLG should use the URI <a href="http://data.uk.gov/open">http://data.uk.gov/open</a> surve.com/</td>
<td>Mandatory</td>
</tr>
<tr>
<td>2</td>
<td>Body name</td>
<td>Name of the authority</td>
<td>Acts availability for casual reading</td>
<td>If it is not possible to source data using the payment data, then data may be used for other purposes. If it is possible to use data to validate the transaction, then data may be used, otherwise it must be used where necessary to the filing of a criminal complaint or to the claim of any statutory right or obligation.</td>
<td>Mandatory</td>
</tr>
<tr>
<td>3</td>
<td>Date</td>
<td>The payment date as recorded in departments’ purchase or general ledger</td>
<td>To identify the activity that cost the transaction took place</td>
<td></td>
<td>Mandatory</td>
</tr>
<tr>
<td>4</td>
<td>Transaction number</td>
<td>A unique reference number for each individual expenditure transaction</td>
<td>To act as a reference number when dealing with queries or FOI requests.</td>
<td>The transaction number used in departments’ own systems may be used.</td>
<td>Mandatory</td>
</tr>
<tr>
<td>5</td>
<td>Invoice Number</td>
<td>The reference number for the invoice if it is recorded on the accounts system</td>
<td>To allow matching to invoices. If payment may be for several invoices, and one invoice may be settled by several payments.</td>
<td>Some sources have been discussing the utility of publishing invoices, and the lack of data allows for an easy transition to this, should they do this.</td>
<td>Discouraged</td>
</tr>
<tr>
<td>6</td>
<td>Amount</td>
<td>The actual value of the transaction</td>
<td>To identify the full cost of the transaction</td>
<td>Amounts should be in sterling and inclusive of all purchase tax (VAT) - figures should be in pounds and pence. Each</td>
<td>Mandatory</td>
</tr>
</tbody>
</table>
Don’t Forget the Local Perspective
UK Local Data Panel: Current Activities

Working with new Transparency division on 3 key tasks

1. How to standardise local authority data and which data should be standardised as a priority

2. Assisting local authorities/public bodies to work through technical/policy issues relating to making data openly available – in particular the Jan 2011 data release

3. Encouraging citizens, developers, public and private sector organisations to access and exploit data provided by CLG and LAs
Don’t Forget the External Perspective
Integrate other Open Data
Twittering London

Enfield
Colindale
Barking
Heathrow
Kingston
Croydon
Purley

(Legend)

Number of Tweets

Time by 24 Hour Clock

Centre for Advanced Spatial Analysis
UCL
Data Publishing – Star Quality

★ Put your data on the Web (any format)

★★ Make it available as structured data (e.g. Excel, CSV, instead of PDF)

★★★ Use open, standard formats (e.g. XML, RDF)

★★★★ Use URLs to identify things (so people and machines can point at your data)

★★★★★ Link your data to other people’s data
Making the case for 5★ Linked Data?

- National digital infrastructure being built
- URIs for schools, roads, bus stops, postcodes, admin boundaries...
- Some of the data links across and connects other data together
- Key data link points exist
Making the case for 5★ Linked Data?

- 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
Joining up the dots…
http://map.psi.enakting.org/how
Joining up the dots…
http://map.psi.backlinks.org/

http://dbpedia.org/resource/Portsmouth

(sameAs) http://statistics.data.gov.uk/id/local-authority-district/00MR
- Further Education (2)
- Foundation School (2)
- LA Nursery School (1)
- Higher Education Institutions (1)
- Miscellaneous (3)
- Voluntary Controlled School (2)
- Other Independent School (9)
- Pupil Referral Unit (7)
- Community Special School (11)
- EY Setting (144)
- Voluntary Aided School (6)
- Community School (57)
- School (514)
- Playing for Success Centres (2)

(sameAs) http://data.ordnancesurvey.co.uk/id/7000000000037254
- CO2 Emission Statistic (85)
- Mortality Data Item (3)
- Road Energy Consumption (30)
EnAKTinG EPSRC UoS Linked Data Project http://myarea.psi.enakting.org/

How good is my area?

feedback ... please email us

Sources:
statistics.data.gov.uk, data.ordnancesurvey.co.uk, IMD Database, Boundary data from OS Open Data

This is a BETA app developed in an EnAKTing hack day. It is built on top of Semantic Web technologies and Linked Data.

It displays a region and its surroundings showing how good are those areas based on the IMD database ranking.

Ward Indicators (higher is better)
Source: data.gov.uk, openspi.org and data.ordnancesurvey.co.uk

Area: Bevois

group by area
Global Open Government Linked Data

This application presents a mashup of foreign aid data (represented in US Dollars) from the United States Agency for International Development (USAID) and UK Department for International Development (DFID) for the 2007 US Fiscal Year.

Users may retrieve foreign aid data for specific countries by clicking on a provided world map (shaded based on total combined contributions for USAID and DFID). Upon clicking on a desired country, three kinds of information are presented: Aid Figures, New York Times news, and Wikipedia description.

Global Open Government Linked Data

This demo uses data from [Dataset 311 on crime in the United States](census.gov) and [UK recorded crime](census.gov). It shows available information on crimes in cities grouped by US countries and UK Territorial Police area.
Competition and Collaboration

• Friendly competition is energising

• Collaboration enables learning and sharing best practice

• Provides external validation

• Offers even great opportunities for data exploitation
Culture Shifts
What needs to happen

Public bodies – need clear, published commitments on releasing data, and there needs to be consistency between similar bodies.

OGD needs to extend through central Government into local government and the other public services.

National agencies need always to publish for re-use the underlying data for web services that they supply e.g. in UK Directgov, NHS Choices.

Bodies should proactively release data, but there should also be the ability to challenge bodies to release more.

Nurture the Open Data Community and learn from the successes of efforts like data.gov.uk and data.gov.

Spend resources on action not reportage and support activities which would add new value to data.

Localism - support community activism and "hyper locality" in using this information and help integration with community networks - e.g. civic society services.

Give support, encouragement and targeted funding to people developing innovative software and services which use public data.
Types of Challenge

• Cultural & Organisational
  – Data hugging
  – We have released it - sort of…
  – License impediments
  – Legacy commitments

• Worries about
  – Confidentiality
  – How it will be Interpreted
  – Quality of Information
  – Lost value
  – Disrupted workflow
  – New IT procurement

• Capability
  – What assets have we got
  – Who can do this work
  – How to share best practice

• Policy
  – Extending the principle to all agencies and public bodies
  – Consistency in what is published and its format
  – Range of exceptions do need to be appreciated
  – Confidentiality, accuracy, liability and reputation
Why Open Government Data matters

More transparency
what is going on?

More accountability
is this acceptable?

More localism
it matters to me, my family, my locale

More economic/social capital
generates opportunities

More engagement
supports participation to exploit and improve data

More debate
data for evidenced based policy