Transparency and Open Data: Why Bother?

Being also a discourse on some lessons learnt

Professor Nigel Shadbolt FREng

7th June 2012

European Data Forum, Copenhagen
open the data…
and the applications follow…

Bus Guru is the UK's first app to give users both the LIVE (real-time) bus times AND the best routes from TFL's Journey Planner and Countown System – and it's FREE!

SAVE TIME & ELIMINATE STRESS: No more waiting around at bus stops, with Bus Guru you can plan your journey to perfection – knowing which route will be the fastest AND when your bus will arrive at the stop.

<<< CLICK ON PHONE SCREEN TO TRY THE LIVE APP

- For iPhone/iPod: Download iPhone App
- For Android: Download Android App

All Smartphones: To install the Bus Guru mobile app open the following link in your phone's browser: http://bit.ly/BusGuru (the web app can usually be saved on your phone from the browser).

FEATURES:

- LIVE INFO: Up-to-the-second information on all TFL buses' real-time locations.
- ONE CLICK: To view your nearest stop's buses OR all the buses around you.
- LIVE JOURNEY PLANNER: Compare the quickest routes AND live bus times using Bus Guru's Journey Planner integrated with live bus times. So you can find the fastest bus AND which bus is leaving first. (We have combined two very complex systems so please email support@busguru.co.uk if you find any bugs and we will update the app).
- SEARCH: For ANY bus stop and live bus info via an address/postcode/area.

Developed by AppShed

Bus Guru has been developed by AppShed.
at all scales...
around the world…
leading to…

- Improved transparency & accountability
- Better Public Services
- Participation… people enhance, improve, capture the data
- Improved efficiency
- Creation of social value
- Economic value, innovation and growth
Lesson: Open Data... should be non-partisan
“Our ambition is to become the world leader in open data, and accelerate the accountability revolution that the internet age has made possible.”

“We are determined to go further in breaking down the walled garden of government, using technology and information to provide greater transparency…”

“My aim is clear: to make the provision of data (with reasonable formats and licenses) a routine task of public administrations.”

“Greater transparency across Government is at the heart of our shared commitment to enable the public to hold politicians and public bodies to account.”
Lesson: Open Data... needs a steady stream of success
Data in full: government publishes NHS MRSA and C.Diff infection rates

- Get the data
Lesson: Open Data... is not a technical problem
It is about changing behaviour

UK 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 – and in addition to the legal Right To Data itself this overriding principle should apply to the implementation of all the other principles.

- Public data will be published in reusable, machine-readable form – publication alone is only part of transparency – the data needs to be reusable, and to make it reusable it needs to be machine-readable. At the moment a lot of Government information is locked into PDFs or other unprocessable formats.

- Public data will be released under the same open licence which enables free reuse, including commercial reuse – all data should be under the same easy to understand licence. Data released under the Freedom of Information Act or the new Right to Data should be automatically released under that licence.

- Public data will be available and easy to find through a single easy to use online access point (data.gov.uk) – the public sector has a myriad of different websites, and search does not work well across them. It’s important to have a well-known single point where people can find the data.

- Public data will be published using open standards, and following relevant recommendations of the World Wide Web Consortium. Open, standardised formats are essential. However to increase reusability and the ability to compare data it also means openness and standardisation of the content as well as the format.

- Public data underlying the Government’s own websites will be published in reusable form for others to use – anything published on Government websites should be available as data for others to reuse. Public bodies should not require people to come to their websites to obtain information.

- Public data will be timely and fine grained – Data will be released as quickly as possible after its collection and in as fine a detail as is possible. Speed may mean that the first release may have inaccuracies; more accurate versions will be released when available.

- Release data quickly, and then re-publish it in linked data form – Linked data standards allow the most powerful and easiest re-use of data. However most existing internal public sector data is not in linked data form. Rather than delay any release of the data, our recommendation is to release it ‘as is’ as soon as possible, and then work to convert it to a better format.

- Public data will be freely available to use in any lawful way – raw public data should be available without registration, although for API-based services a developer key may be needed. Applications should be able to use the data in any lawful way without having to inform or obtain the permission of the public body concerned.

- Public bodies should actively encourage the re-use of their public data – in addition to publishing the data itself, public bodies should provide information and support to enable it to be reused easily and effectively. The Government should also encourage and assist those using public data to share knowledge and applications, and should work with business to help grow new, innovative uses of data and to generate economic benefit.

- Public bodies should maintain and publish inventories of their data holdings – accurate and up-to-date records of data collected and held, including their format, accuracy and availability.

They are asking everyone to help shape and define these important principles and have set up
However…

★ 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
Create, Use and Share URIs

- 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
Create, Use and Share URIs

• 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
Create, Use and Share URIs

- Schools
- Post-codes
- Government posts
- Ministers
- Bus stops

This is the SPARQL query that was generated to pull together the data that is provided about the selected item. You can modify it here and re-run the...
Lesson: Open Data…
if it doesn’t scale it will fail
Datasets on data.gov.uk over time
Open Data – the network effect

- Network effects
  - Accessibiltly
  - Popularisation
  - Centreless
  - Timeliness
  - Convergence

[Diagram of network effects]

Courtesy Mathew Hurst
Lesson: Open Data… needs location, location, location
See also: Linked data  http://www.ordnancesurvey.co.uk/opendata/linkeddata.html
the connective tissue of open data
London Cycle Hire Journeys and Pollution

Thicker lines are more journeys
Redder lines are more polluted (PM10 2008 mean)
**Geodemographics of Housing in Great Britain**

A new visualisation in the style of Charles Booth’s map

Created by Oliver O’Brien (@oobr), UCL CASA – more information

**Showing the Index of Multiple Deprivation (2010)**

**Switch to the Output Area Classification**

Link to the Current View (Bookmark This Place)

*Place names & geographical features: Contains Ordnance Survey data © Crown copyright and database right 2011. The colours shown are only applicable to residential buildings, but all types of buildings, regardless of type, are shown on the map. IMPORTANT NOTE: Classifications are an average across the local Output Area or LSOA, rather than distinguishing between single houses (or even streets), therefore the colour coding on an individual house is not necessarily accurate.*
Lesson: Open Data…
release data that matters
UK Reported Crime Data

• A tough data set to release

• All reported crime each month

• By crime type

• At the street level

• Police now use the apps themselves
From Schools to Arterial Surgery

Priestlands School, Lymington

48% of pupils achieved 5 A*-C grades at GCSE, inc maths and English - compared to the England state school average of 58%
Community School, Mixed
GCSE by subject
English: 73% of pupils achieved expected progress at GCSE level
Maths: 76% of pupils achieved expected progress at GCSE level
GCSEs: reaching the expected level

73
76

English
Maths

The English baccalaureate: This covers five subjects. At this school, 16% of pupils achieve an English bac. 20% of children enter all areas
A levels (if applicable)
[NA] % of students achieve three of more A-levels, grade A*-E. The England average is 60.9%
About this school
Religious denomination: Does not apply
This school spends £4745 per pupil, compared to an England average of £5,712 per pupil

Download the fusion table behind this map as a CSV
LOCAL AUTHORITY: Hampshire, SCHOOL POSTCODE: SO41 8FZ, DFE NUMBER: 880418

What's the mortality rate for elective major artery surgery at your hospital?
For the first time the Vascular Society have reported the mortality rates from major artery surgery for hospitals in the UK. Here’s the data from the report for you to explore

Get the data
Sarah Boseley and Lisa Evans
guardian.co.uk, Wednesday 29 February 2012 12.18 GMT
Article history

Your chances of surviving major artery surgery really do partly depend on which hospital you choose, a groundbreaking audit from the Vascular Society shows.
The first published audit from the Vascular Society has found that hospitals where surgeons carry out many AAA (abdominal aortic
Lesson: Open Data... must have a balanced ecosystem
The open data ‘ecosystem’

Open government data
Data produced, collected or paid for by the public sector, subject to restrictions relating to *sub judice*, national security, commercial sensitivity and privacy.

Special commercial arrangements also being made for certain trading funds, including Companies House, the Ordnance Survey, the Meteorological Office and HM Land Registry, which together form the newly created Public Data Group.

Open business data
Data produced or collected by the private sector and published freely and openly, subject to restrictions that individual businesses decide to put in place.

Open citizen data
The personal and non-personal data of individual citizens published into the open domain.

Source Deloitte LLP
• The Network Effect
  • for data, people and organisations
• A Focal Point
  • coordinate, convene, link and integrate
• The Power of Open
  • Standards, Data, Licences, SW
• Open Data as a Platform for
  • Government
  • Business
  • Research
“The vision is to establish the Open Data Institute as a world leading centre to innovate, exploit and research the opportunities for the UK created by the Government’s Open Data policy.”

- Business Innovation
- Public Sector Innovation
- Training the Open Data Generation
- Developing and Researching Open Data
- Developing Open Data Standards and Policies
- Advisor to Government & International Collaboration
OGD 2.0

• We haven’t worked out how to close the feedback loops

• Publish out

• Acquire or Enhance

• Write Back
Lesson: Open Data... needs Business
Economic Value, Innovation and Growth
## Innovation using open data

Using the growing estate of open government data

<table>
<thead>
<tr>
<th>Company</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red Spotted Hanky</td>
<td>• an online ticket retailer which aims to offer customers an easier way to book without any administration or payment fees</td>
</tr>
<tr>
<td>Duedil</td>
<td>• a business information provider that gives free access to governance and financial information for every company in the UK and Ireland, and combines this with data from online sources, APIs, social networks, and more</td>
</tr>
<tr>
<td>Parkopedia</td>
<td>• a small UK-based business, which uses live data from local authorities to help drivers identify free car parking spaces</td>
</tr>
<tr>
<td>Spotlightonspend</td>
<td>• a managed service delivering everything necessary to facilitate cost-effective publication of the spend and related information public bodies elect to make available to the general public</td>
</tr>
</tbody>
</table>
## Innovation using open data

### Opening up business data

<table>
<thead>
<tr>
<th>Company</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enel</td>
<td>Enel, Italy’s largest power company, launched its open data site in August 2011 to increase transparency and involvement by all stakeholders, and to encourage innovation and stimulate the development of new applications.</td>
</tr>
<tr>
<td>Nike</td>
<td>Nike publishes an annual Sustainable Business Performance Summary, which describes key data and metrics relating to the company’s sustainability efforts. Nike also provides tools where users can check the environmental impact of various material selections and explore the company’s global contracted manufacturing base.</td>
</tr>
<tr>
<td>Asos</td>
<td>In December 2011, the online fashion giant Asos opened up its product and basket services data to external web developers in a bid to extend its wider online presence by allowing others to build Asos-related services.</td>
</tr>
<tr>
<td>MusicMetric</td>
<td>MusicMetric provides insights into online consumer behaviour for the entertainment industry. MusicMetric currently tracks and indexes data for 600,000 artists and over 10 million individual releases.</td>
</tr>
</tbody>
</table>
Lesson: Open Data… isn’t everything
Varieties of Data

- Big - Small
- Public – Private
- Open - Closed
- Personal - Non-Personal
- Anonymous - Identified
- Aggregate - Individual
Personal Data Asset Revolution...
Lesson: Open Data... will need defending
Challenges

- Infrastructure
- Quality
- Interpretation
- Security and Privacy
- Evidence of impact
- Sir Humphrey 2.0
A sample of “data hugging” excuses

- It’s held separately by \( n \) different organisations, and we can’t join it up
- It will make people angry and scared without helping them
- It is technically impossible
- We do not own the data
- The data is just too large to be published and used
- Our website cannot hold files this large
- We know the data is wrong
- We know the data is wrong, and people will tell us where it is wrong
- We know the data is wrong, and we will waste valuable resources inputting the corrections people send us
- People will draw superficial conclusions from the data without understanding the wider picture
- People will construct league tables from it
- It will generate more Freedom of Information requests
- It might be combined with other data to identify individuals/sensitive information
- It will cost too much to put it into a standard format
- Our IT suppliers will charge us a fortune to do an ad hoc extract
A sample of “data hugging” excuses

- It’s held separately by $n$ different organisations, and we can’t join it up
- It will make people angry and scared without helping them
- It is technically impossible
- We do not own the data

- All these sometimes have some truth in them
- Often they rationalise the official fear of the unknown
- Data owners need to be helped through
- Example & precedent are your friends
- For first phase, compromise on data not licence

- People will draw superficial conclusions from the data without understanding the wider picture
- People will construct league tables from it
- It will generate more Freedom of Information requests
- It might be combined with other data to identify individuals/sensitive information
- It will cost too much to put it into a standard format
- Our IT suppliers will charge us a fortune to do an ad hoc extract

Often they rationalise the official fear of the unknown

Example & precedent are your friends

For first phase, compromise on data not licence
Lesson: Open Data... keep reminding them
Open Data and the Power of Open
never one driver nor one reason

• Open Licences
• Open Standards
• Open Data
• Open Data Principles
• Open Participation
• Open Minds