Open Data in practice
Part II: The practicalities of open data

#WWW2012
Hadley Beeman
Hadley@LinkedGov.org

WWW Tutorial Track
Lyon, France, 17 April 2012
The many sources of open data
UK Government Open Data
What can we do with it?
Transparency and accountability
Delivering services to/on behalf of government
Improving commercial products and services
Public sector efficiencies
Economic impact
Why we're not there yet
Variety... of publishers
Variety... of formats
Variety... of licences
### Allegations Recorded - by type

- **Oppressive Behaviour**
- **Discrimination**
- **Malpractice**
- **Failures in Duty**
- **Incivility**
- **Traffic Irregularity**
- **Other**

**Grand Total - Last Period**
**Grand Total - This Period**

---

### SPENDING

**Budget**

<table>
<thead>
<tr>
<th>Total Departmental Expenditure Limit (DEL)</th>
<th>Q2 2011-12</th>
<th>Q2 2010-11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>of which Resource DEL</strong> (excl. Depreciation)</td>
<td>4,220</td>
<td>5,475</td>
</tr>
<tr>
<td><strong>A: HEFCE</strong></td>
<td>1,860</td>
<td>1,899</td>
</tr>
<tr>
<td><strong>B: Skills Funding Agency</strong></td>
<td>805</td>
<td>962</td>
</tr>
<tr>
<td><strong>C: Higher Education Student Support and Miscellaneous HE Grants</strong></td>
<td>498</td>
<td>491</td>
</tr>
<tr>
<td><strong>D: Research Councils</strong></td>
<td>593</td>
<td>963</td>
</tr>
</tbody>
</table>

**Common Areas of Spend**

<table>
<thead>
<tr>
<th>Estate Costs</th>
<th>Total office estate (m2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost per FTE (£)</strong></td>
<td>Total cost of office estate (£million)</td>
</tr>
<tr>
<td><strong>Cost per m2 (£)</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Procurement</th>
<th>Total Procurement Spend (£million)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Price of standard box of A4 white copier paper (£/2500 sheets)</strong></td>
<td><strong>Average price of energy (£/KWH)</strong></td>
</tr>
</tbody>
</table>
Variety... of access points and SLAs

Business Link

www.businesslink.gov.uk

0845 600 9 006

Business Link services are available online and through a network of local operators.

Page not found
We're sorry - the businesslink.gov.uk page that you requested could not be found. It may have been moved, updated or deleted.

As part of ongoing changes to the site, improvements are being implemented to enhance our services. If you are a registered user, your bookmarks may be affected by this.

Please try the following options:
• if you typed the address in, please check the address and spelling and try again
Data quality

TOTAL BODY MODIFICATION
A FULLY LICENSED PROFESSIONAL STUDIO!
- STERILIZED AND ONE TIME USE EQUIPMENT
- HYGENIC CONDITION
- HIGH QUALITY JEWELRY
- FULL AFTER CARE ADVISORY
Data discovery on the Web

Early lists of sites on the Web

- Click here to jump to my MOST FAVORITE site on BMX-Pickle's Racing Snack Shop (it's not what it sounds like)
- Click here to learn about how to start BMX racing - it's the NBL home page
- Click here to go to VANS shoes web site. Contests to win new shoes
- Click here to go to the Airwalk shoes web site. New sick BMX/freestyle movies. Gotta check this out!
- Click here to jump to the GT web page- see the very bike I own.......the Dyno NSX
- Click here to jump to the Diamondback web page...see the light inexpensive bikes-plus a contest too!
- Click here to jump to the JAD website...see a six pound BMX frame and an eight pound freestyle frame.
- Click here to go to the Santa Cruz web site......reviews on their new BMX bike the Jackal.

Current (early) lists of data on the Web

- Data repositories
  - Archaeology
    - Also see Social sciences.
    - Archaeology Data Service
  - Astronomy
    - Also see Physics.
    - Astronomical Data Archives Center
  - Biology
    - Also see BCO-DMO, Marine Biology data, listed with Marine Sciences repositories.
    - Also see DataONE, Ontex databases, KB, and PANGAEA, listed under Multidisciplinary repositories.
  - The Anthropocene Information Resource - The Anthropocene Information Resource (TAIR) maintains a
    - BOND (Biomolecular Object Network Databank). From Unleashed Informatics
How to make sense of it all?
## Adding context through linked data

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Potholes in our county council</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Date reported</td>
<td>Street</td>
</tr>
<tr>
<td>3</td>
<td>15/01/11</td>
<td>High Street</td>
</tr>
<tr>
<td>4</td>
<td>16/01/11</td>
<td>Tree Lane</td>
</tr>
<tr>
<td>5</td>
<td>16/01/11</td>
<td>High Street</td>
</tr>
<tr>
<td>6</td>
<td>18/01/11</td>
<td>B3068</td>
</tr>
<tr>
<td>7</td>
<td>19/01/11</td>
<td>B3068</td>
</tr>
<tr>
<td>8</td>
<td>22/01/11</td>
<td>B2013</td>
</tr>
<tr>
<td>9</td>
<td>23/01/11</td>
<td>High Street</td>
</tr>
<tr>
<td>10</td>
<td>24/01/11</td>
<td>High Street</td>
</tr>
<tr>
<td>11</td>
<td>24/01/11</td>
<td>London Road</td>
</tr>
<tr>
<td>12</td>
<td>26/01/11</td>
<td>Birmingham Road</td>
</tr>
<tr>
<td>13</td>
<td>27/01/11</td>
<td>High Street</td>
</tr>
<tr>
<td>14</td>
<td>28/01/11</td>
<td>High Street</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Street</th>
<th>Post code</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Street</td>
<td>BM1 3</td>
</tr>
<tr>
<td>Tree Lane</td>
<td>BM1 2</td>
</tr>
<tr>
<td>B3068</td>
<td>BM1 1</td>
</tr>
<tr>
<td>London Road</td>
<td>BM1 2</td>
</tr>
<tr>
<td>Birmingham Road</td>
<td>BM1 3</td>
</tr>
</tbody>
</table>

Royal Mail

County Council

Ordnance Survey
Semantically enriching the data: linked data

<table>
<thead>
<tr>
<th>Directorate</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult, Health and Community Services</td>
<td>Medium term financial plan allocation</td>
</tr>
<tr>
<td>Adult, Health and Community Services</td>
<td>Demographic pressure</td>
</tr>
<tr>
<td>Adult, Health and Community Services</td>
<td>Increasing care needs in Learning Disability Services</td>
</tr>
<tr>
<td>Adult, Health and Community Services</td>
<td>Carers services</td>
</tr>
<tr>
<td>Adult, Health and Community Services</td>
<td>Appointeeship services for vulnerable adults</td>
</tr>
<tr>
<td>Adult, Health and Community Services</td>
<td>Safeguarding</td>
</tr>
<tr>
<td>Adult, Health and Community Services</td>
<td>Libraries planned four-year transformation programme</td>
</tr>
<tr>
<td>Children, Young People and Families</td>
<td>Medium term financial plan allocation</td>
</tr>
<tr>
<td>Children, Young People and Families</td>
<td>Social care placements for Looked After Children</td>
</tr>
<tr>
<td>Children, Young People and Families</td>
<td>Legal case work</td>
</tr>
<tr>
<td>Children, Young People and Families</td>
<td>Special Educational Needs transport</td>
</tr>
<tr>
<td>Children, Young People and Families</td>
<td>Transforming Education / Transforming Communities programme</td>
</tr>
<tr>
<td>Customers, Workforce and Governance</td>
<td>Medium term financial plan allocation</td>
</tr>
<tr>
<td>Customers, Workforce and Governance</td>
<td>Intranet</td>
</tr>
<tr>
<td>Environment and Economy</td>
<td>Medium term financial plan allocation</td>
</tr>
<tr>
<td>Environment and Economy</td>
<td>Highways maintenance drainage</td>
</tr>
<tr>
<td>Environment and Economy</td>
<td>Grass cutting - contribution to District &amp; Borough Council</td>
</tr>
<tr>
<td>Environment and Economy</td>
<td>PFI management costs for Pedestrian Transport</td>
</tr>
</tbody>
</table>
Semantically enriching the data: typing and mark-up

What a machine understands **before**

(CSV, TSV, Excel)
Semantically enriching the data: typing and mark-up

What a machine understands

_after_

(machine-readable format)

<table>
<thead>
<tr>
<th>Building</th>
<th>Temp</th>
<th>Name</th>
<th>Gas/hour</th>
<th>Postcode</th>
<th>Date</th>
<th>Water /hour</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Celsius</td>
<td>string</td>
<td>kWh</td>
<td></td>
<td></td>
<td>m3</td>
<td>metres</td>
</tr>
<tr>
<td>Building</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Celsius</td>
<td>String</td>
<td>kWh</td>
<td></td>
<td>date</td>
<td>m3</td>
<td>metres</td>
</tr>
<tr>
<td>Building</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Celsius</td>
<td>string</td>
<td>kWh</td>
<td></td>
<td>date</td>
<td>m3</td>
<td>metres</td>
</tr>
<tr>
<td>Building</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Celsius</td>
<td>string</td>
<td>kWh</td>
<td></td>
<td>date</td>
<td>m3</td>
<td>metres</td>
</tr>
<tr>
<td>Building</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The power of linking

NHS geo data
GP Surgery address data
NHS events data
GP Surgery energy use data

LinkedGov
Coordinated effort: LinkedGov

Create a clean, usable, understandable body of UK government data for any purpose.

What that means:
- Machine readable
- Human readable
- Typed
- Internally linked

How we do that:
- An open source project
- A little bit of help from everyone
- All in one place
- Make the most of their skills (technical or not, civil service or citizen)
LinkedGov

Import existing data (CSV, Excel, XML)

Cleaning tasks

✔ ➔ ✗

LinkedGov database & core components
(Data is stored as machine-readable data)

Question site

data.linkedgov.org
Hi! You are from BIS?
Can you tell us what
072529-1092
means in the dataset
Budget forecasts Q3 2012-13 BIS
Programme for Innovation in the South West

Are these two terms related?
Environmental services
Environ svcs
Access to data to encourage the data publishers

What would you like to know?

How many government employees work in Victoria St, SW1?

Result:

1,570

Data sources:
Westminster City Council Venues
BIS employee distribution
DTI site census

Reliability score: 8
Bit by bit, we're getting there.

Let's try an example of fitting the pieces together.
The goal: to build this

A BUS FOR THAT

by Elliot Long (@elroid)

Mobile-web based application using the Greater Manchester CIF bus timetables. System imports CIF files, and parses out route and stop info, geocodes this data and makes it available for queries. Available as JSON or KML placemarks of the route.

User can select a certain bus stop and see all services due to arrive in the near future.

Links

- Demo URL: http://abusforthat.appspot.com/
- Source Code: http://elroid.com/abus-src.zip

With thanks to Elliot Long (@elroid), from his work at linkedgov.hackcamp.org.uk
The data - timetables
The data – bus stop locations

National Public Transport Access Nodes (NaPTAN)

NaPTAN is a GB national system for uniquely identifying all the points of access to public transport in GB. It is a core component of the GB national transport information infrastructure and is used by a number of other UK standards and information systems. Every GB station, coach terminus, airport, ferry terminal, bus stop, etc., is allocated at least one identifier.
Converting the bus stops to lat/long

Example from, and for more info, try
http://www.movable-type.co.uk/scripts/latlong-gridref.html
Not always necessary, depending on the data source and formats used

<table>
<thead>
<tr>
<th>Functional demo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enter latitude/longitude values or OS grid references into the test boxes to try out the calculations:</td>
</tr>
<tr>
<td>WGS84 Lat/Lon</td>
</tr>
<tr>
<td>OSGB36 Lat/Lon</td>
</tr>
<tr>
<td>OS Grid Ref</td>
</tr>
</tbody>
</table>
The full data journey

1. Location from the mobile phone in lat/long

2. Conversion of the user's location to Easting/Northing

3. Lookup of bus stops near the user

4. Producing the relevant bus routes, filtering by time
Info to the user: Mobile-friendly HTML page

http://abusforthat.appspot.com/#/stop?id=800SB04791
Image credits

Bus - Pimlico Badger
http://www.flickr.com/photos/salford_ian/2178180169/sizes/o/in/photostream/

Financials - mihalysoft
http://www.flickr.com/photos/mihalysoft/5378101443/sizes/s/

Line graph - spierzchala
http://www.flickr.com/photos/spierzchala/209113241/sizes/m/

NHS - comedy_nose
http://www.flickr.com/photos/comedynose/3797160719/sizes/s/

Senate House - secpil
http://www.flickr.com/photos/umdrums/413274705/sizes/l/in/photostream/

Newspaper - carmichaellibrary
http://www.flickr.com/photos/carmichaellibrary/3820938629/sizes/s/in/photostream/

Head banger - sybrenstuvel
http://www.flickr.com/photos/sybrenstuvel/2468506922

Receipt - Dan4th
http://www.flickr.com/photos/dan4th/2213552778/sizes/s/

Market – zoetnet
http://www.flickr.com/photos/zoetnet/4529384602/

Sunlit Whitehall – garryknight
http://www.flickr.com/photos/garryknight/3717643414/

1910 World map – pbarry
http://www.flickr.com/photos/pbarry/3156691873/

Geocities cached page:
http://www.geocities.ws/Colosseum/Track/1186/bmxlink.html

At Napier, for real and on screen – wmacphail
http://www.flickr.com/photos/wmacphail/230874881/sizes/m/in/photostream/

IMG_5078 - fccdotgov
http://www.flickr.com/photos/fccdotgov/4808160161/
Questions or suggestions?

#WWW2012
Hadley Beeman
Hadley@LinkedGov.org

@HadleyBeeman
@LinkedGov

WWW Tutorial Track
Lyon, France, 17 April 2012