

Open Data in practice

Part II: The practicalities of open data



#WWW2012

Hadley Beeman
Hadley@LinkedGov.org

@HadleyBeeman
@LinkedGov

WWW Tutorial Track
Lyon, France, 17 April 2012



The many sources of open data



UK Government Open Data

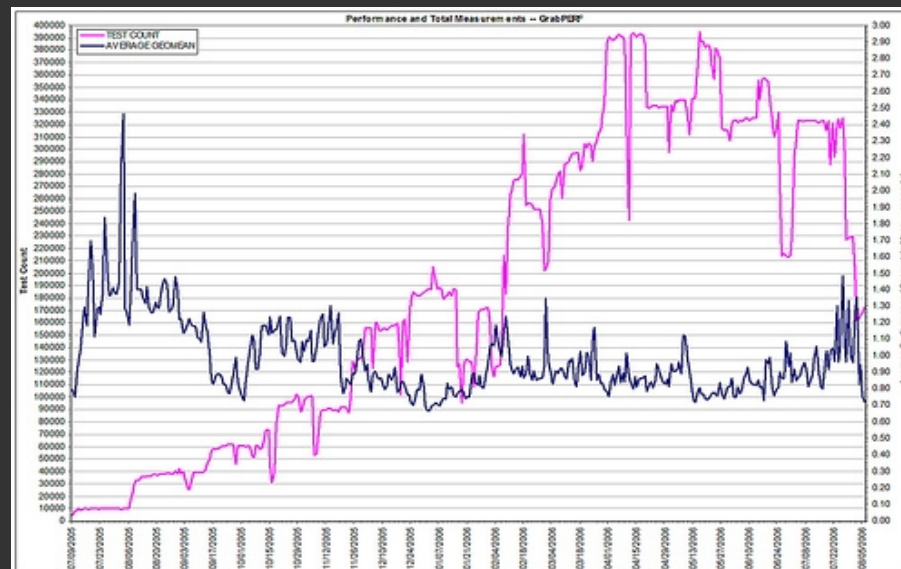


Purchase Register

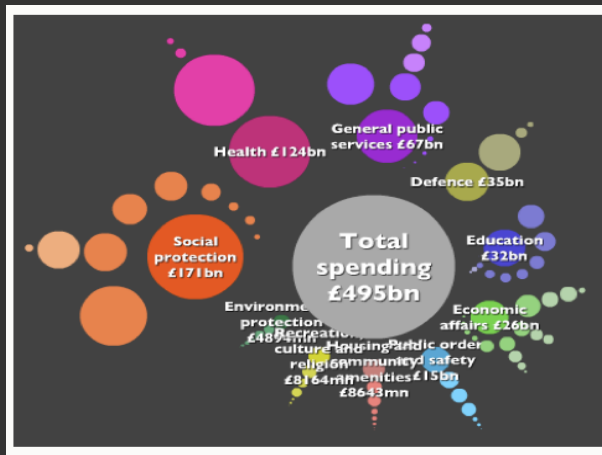
Voucher No.: 100
Account Number: 1000000000
Vendor Name: 1000000000
Address: 1000000000
Purchase Tax Type: 1000000000
Date: 10/10/2008
Store Name: 1000000000
Currency: 1000000000
Invoice Number: 1000000000

No.	Item Code	Item Name	Qty	Rate	Amount
1	1000000000	1000000000	1000	1000.00	1000000.00
2	1000000000	1000000000	1000	1000.00	1000000.00
3	1000000000	1000000000	1000	1000.00	1000000.00
4	1000000000	1000000000	1000	1000.00	1000000.00
5	1000000000	1000000000	1000	1000.00	1000000.00
6	1000000000	1000000000	1000	1000.00	1000000.00
7	1000000000	1000000000	1000	1000.00	1000000.00
8	1000000000	1000000000	1000	1000.00	1000000.00
9	1000000000	1000000000	1000	1000.00	1000000.00
10	1000000000	1000000000	1000	1000.00	1000000.00

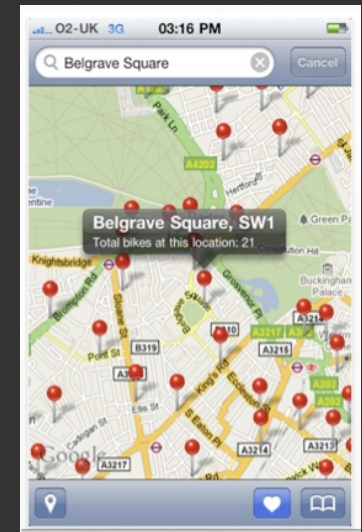
Grand Total: 10000000.00



What can we do with it?



UK	Countries	Institutions	About
UK Total 2008-09			
Totals			
Students	2,396,050	+3.9%	
Undergrads	1,859,240	+3.0%	
Postgrads	536,810	+7.1%	
Staff	382,760	+2.8%	
Academic	179,040	+2.3%	
Non-academic	203,720	+3.1%	
Income (£k)	25,274,298	+8.2%	
Expenditure (£k)	24,944,436	+9.0%	
Undergraduate Modes			
Full-time	1,272,030	+3.2%	
Part-time	587,205	+2.5%	
Undergraduate Domiciles			
UK	1,673,655	+2.6%	
Other EU	73,375	+5.0%	
Non EU	112,215	+7.4%	
Postgraduate Modes			



Transparency and accountability



Delivering services to/on behalf of government



Improving commercial products and services



Public sector efficiencies



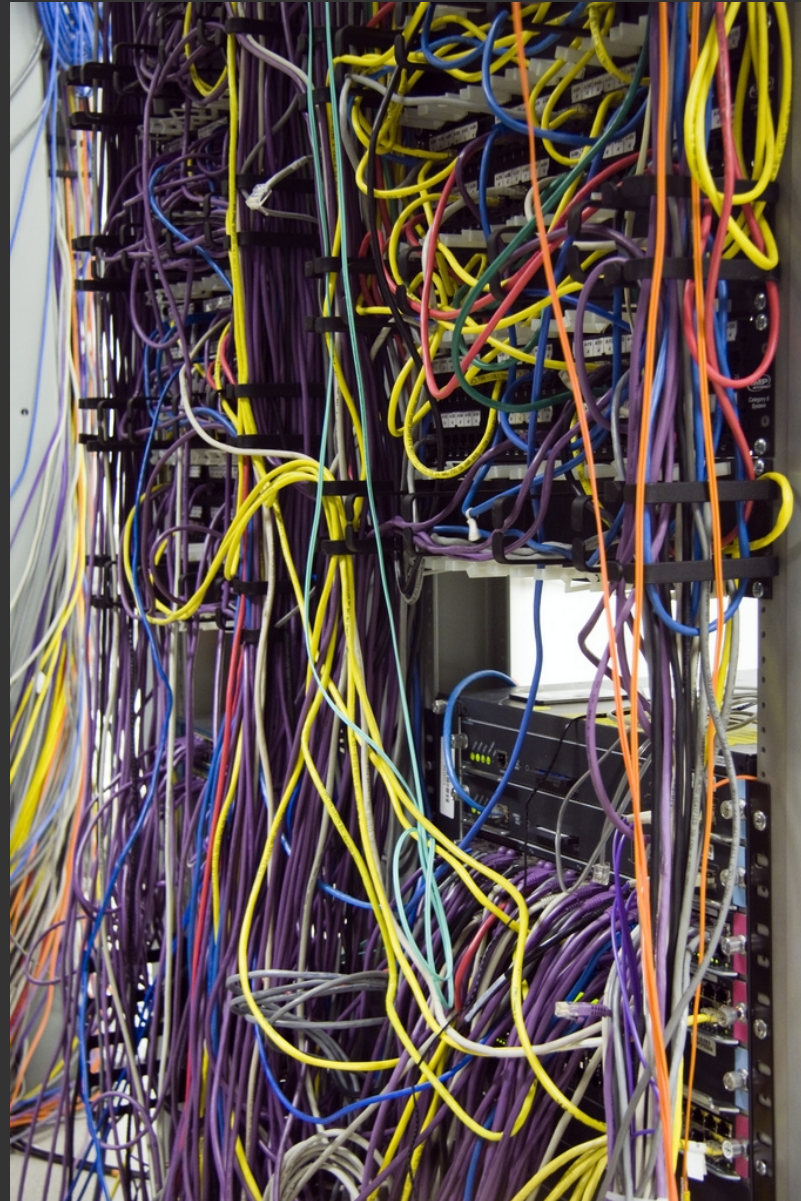
Economic impact



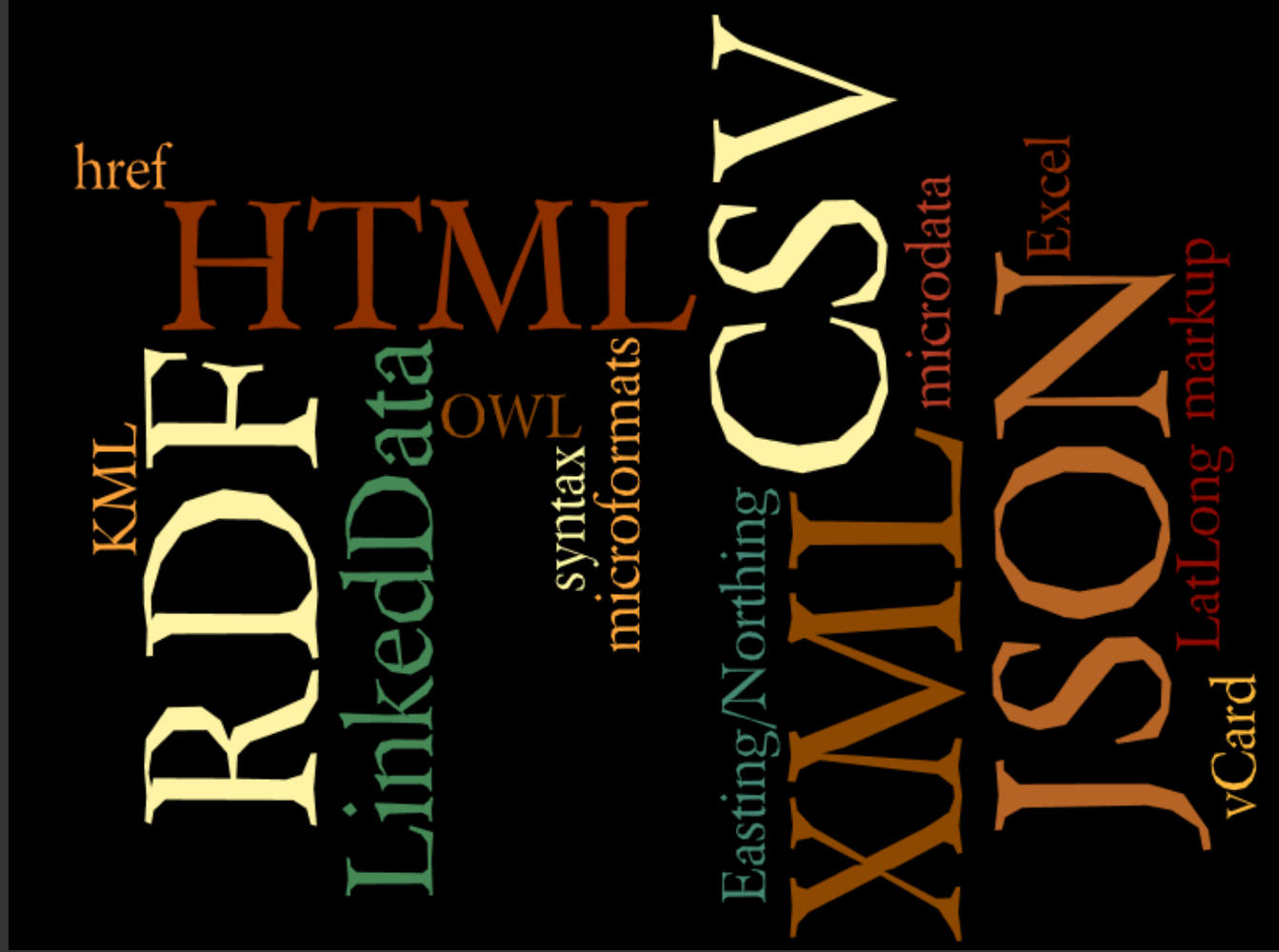
Why we're not there yet



Variety... of publishers



Variety... of formats



Variety... of licences



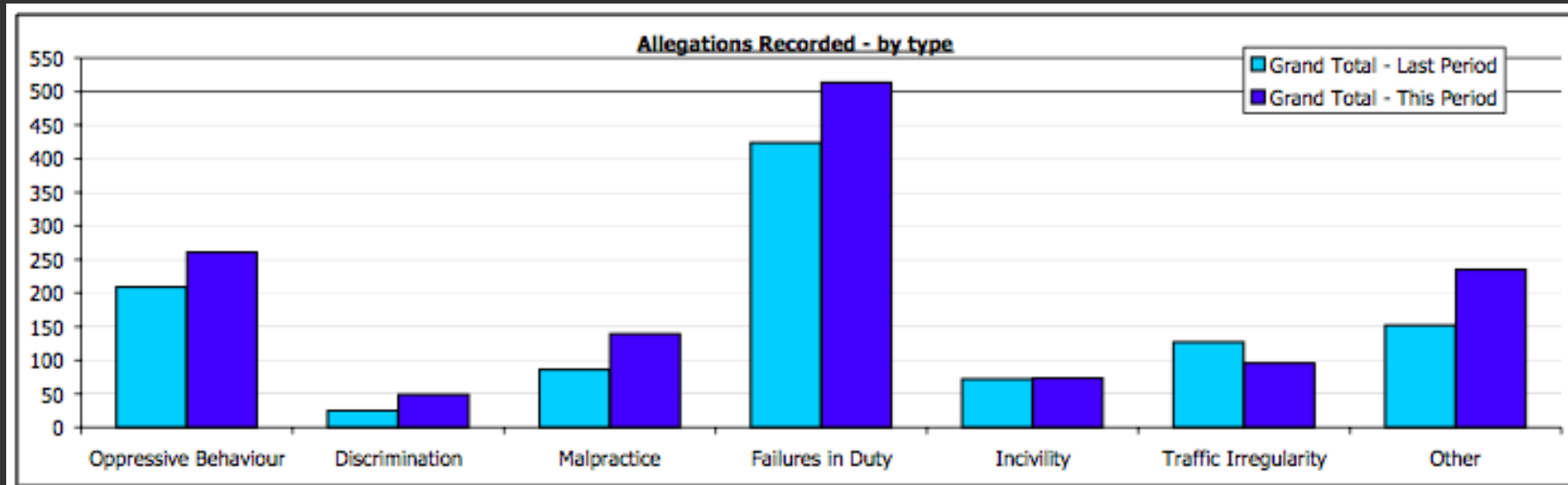
© All Rights Reserved



Open Government Licence
for public sector information



Machine readable? Raw data?



2							
3	SPENDING						
4							
5	Budget	£million		Common Areas of Spend			Q
6							
7		Q2 2011-12	Q2 2010-11	Estate Costs	Total office estate (m2)		2
8	Total Departmental Expenditure Limit (DEL)	4,220	5,475		Total cost of office estate (£million)		
9	of which Resource DEL (excl. Depreciation)	4,245	4,872		Cost per FTE (£)		
10	Upto top 5 contributory elements	A: HEFCE	1,860		Cost per m2 (£)		
11		B: Skills Funding Agency	805	Procurement	Total Procurement Spend (£million)		
12		C: Higher Education Student Support and Miscellaneous HE Grants	498		Price of standard box of A4 white copier paper (£/2500 sheets)		
13		D: Research Councils	593		Average price of energy (£/KWH)		

Variety... of access points and SLAs



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:

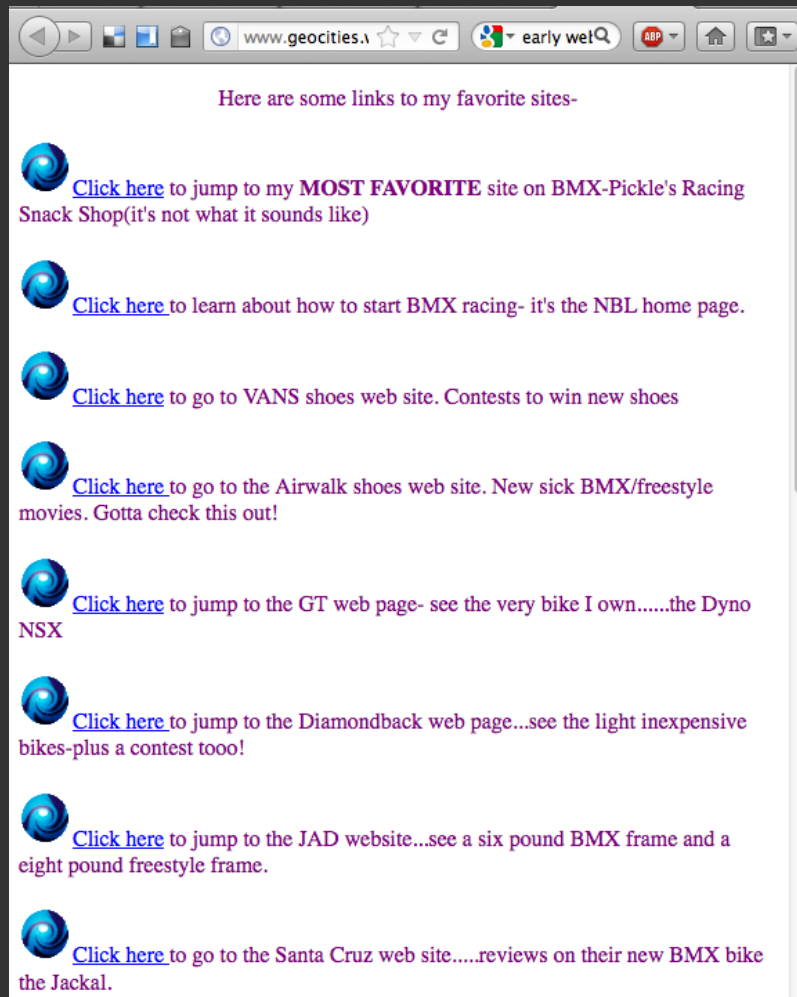
- if you typed the address in, please check the address and spelling and try again



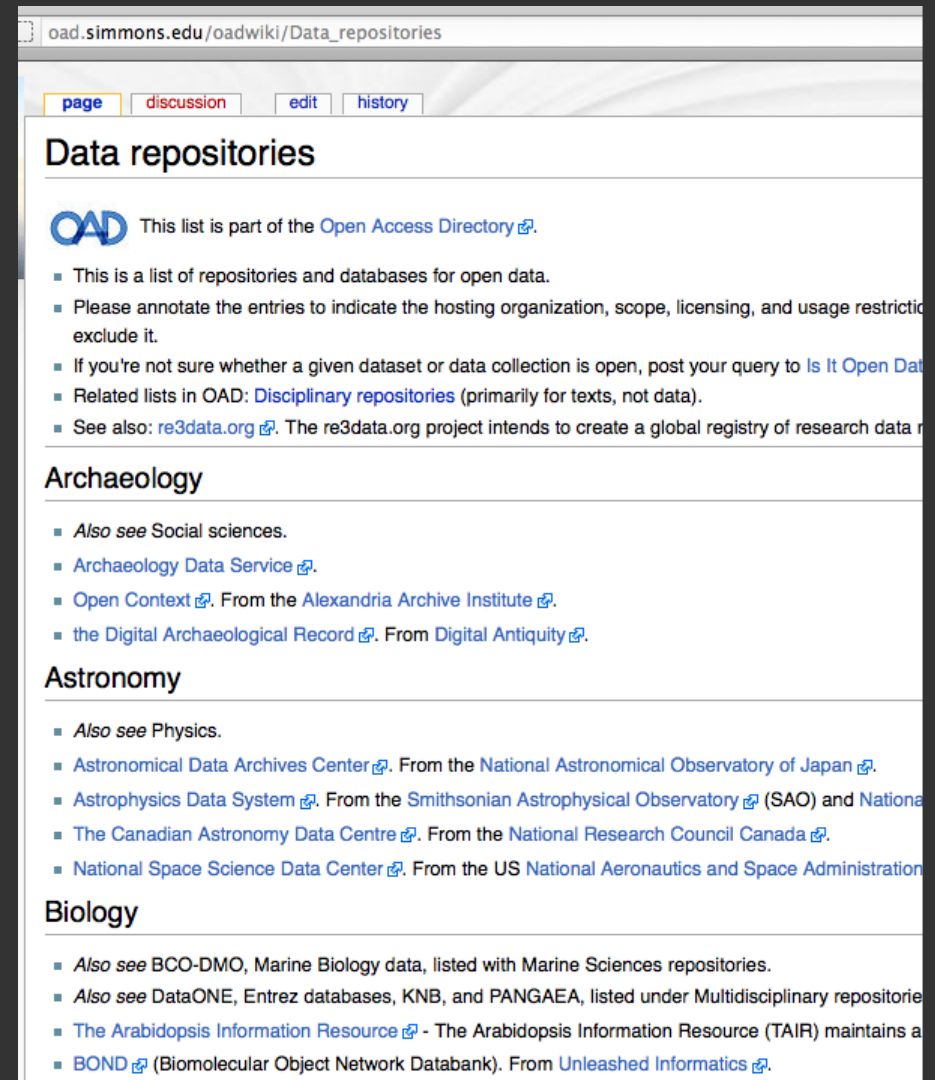
Data quality



Data discovery on the Web



Early lists of sites on the Web



Current (early) lists of data on the Web



How to make sense of it all?



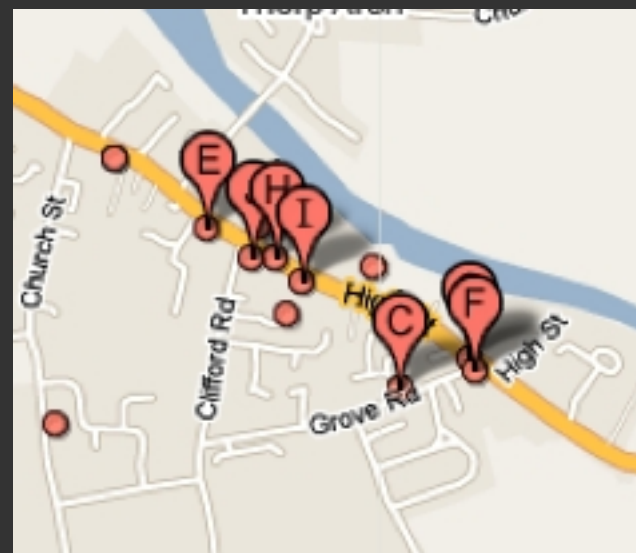
Adding context through linked data

	A	B	C
1	Potholes in our county council		
2			
3	Date reported	Street	
4	15/01/11	High Street	
5	16/01/11	Tree Lane	
6	16/01/11	High Street	
7	18/01/11	B3068	
8	19/01/11	B3068	
9	22/01/11	B2013	
10	23/01/11	High Street	
11	24/01/11	High Street	
12	24/01/11	London Road	
13	26/01/11	Birmingham Road	
14	27/01/11	High Street	
15	28/01/11	High Street	
16			
17			

County Council

Street	Post code
High Street	BM1 3
Tree Lane	BM1 2
B3068	BM1 1
London Road	BM1 2
Birmingham Road	BM 1 3

Royal Mail



	A	B	C	D	E	F	
1	BM1 1AE	10	291960	92581	E92000001	E19000002	E18
2	BM1 1AT	10	291778	92355	E92000001	E19000002	E18
3	BM1 1BA	10	291725	92265	E92000001	E19000002	E18
4	BM1 1BB	10	291786	92251	E92000001	E19000002	E18
5	BM1 1BD	10	291763	92290	E92000001	E19000002	E18

Ordnance Survey



Semantically enriching the data: linked data

1		
	Directorate	Description
2	Adult, Health and Community Services	Medium term financial plan allocation
3	Adult, Health and Community Services	Demographic pressure
4	Adult, Health and Community Services	In
5	Adult, Health and Community Services	C
6	Adult, Health and Community Services	A
7	Adult, Health and Community Services	S
8	Adult, Health and Community Services	Li
9	Children, Young People and Families	M
10	Children, Young People and Families	S
11	Children, Young People and Families	Le
12	Children, Young People and Families	S
13	Children, Young People and Families	Ti
14	Customers, Workforce and Governance	M
15	Customers, Workforce and Governance	In
16	Environment and Economy	M
17	Environment and Economy	H
18	Environment and Economy	G
19	Environment and Economy	P
20	Environment and Economy	P
21	Environment and Economy	D

	Directorate	Description
1	Adult, Health and Community Services	Medium term financial plan allocation
2	Adult, Health and Community Services	Demographic pressure
3	Adult, Health and Community Services	Increasing care needs in Learning Disability Services
4	Adult, Health and Community Services	Carers services
5	Adult, Health and Community Services	Appointeeship services for vulnerable adults
6	Adult, Health and Community Services	Safeguarding
7	Adult, Health and Community Services	Libraries planned four-year transformation programme
8	Adult, Health and Community Services	Medium term financial plan allocation
9	Children and family services	Social care placements for Looked After Children
10	Children and family services	Legal case work
11	Children and family services	Special Educational Needs transport
12	Children and family services	Transforming Education / Transforming Communities p
13	Customers, Workforce and Governance	Medium term financial plan allocation
14	Customers, Workforce and Governance	Intranet
15	Environment and Economy	Medium term financial plan allocation
16	Environment and Economy	Highways maintenance drainage
17	Environment and Economy	Grass cutting - contribution to District & Borough Cou
18	Environment and Economy	DEL and management costs for De

Semantically enriching the data: typing and mark-up

What a machine
understands before

(CSV, TSV, Excel)

	Column	Column	Column	Column	Column	Column	Column
Row	number	word	number	word	date	number	number
Row	number	word	number	word	date	number	number
Row	number	word	number	word	date	number	number
Row	number	word	number	word	date	number	number
Row	number	word	number	word	date	number	number

Semantically enriching the data: typing and mark-up

What a machine
understands
after

(machine-readable format)

	Temp	Name	Gas/hour	Postcode	Date	Water /hour	Height
Building	Celsius	string	kWh	Postcode	date	m3	metres
Building	Celsius	String	kWh	Postcode	date	m3	metres
Building	Celsius	string	kWh	Postcode	date	m3	metres
Building	Celsius	string	kWh	Postcode	date	m3	metres
Building	Celsius	string	kWh	Postcode	date	m3	metres

1



1

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

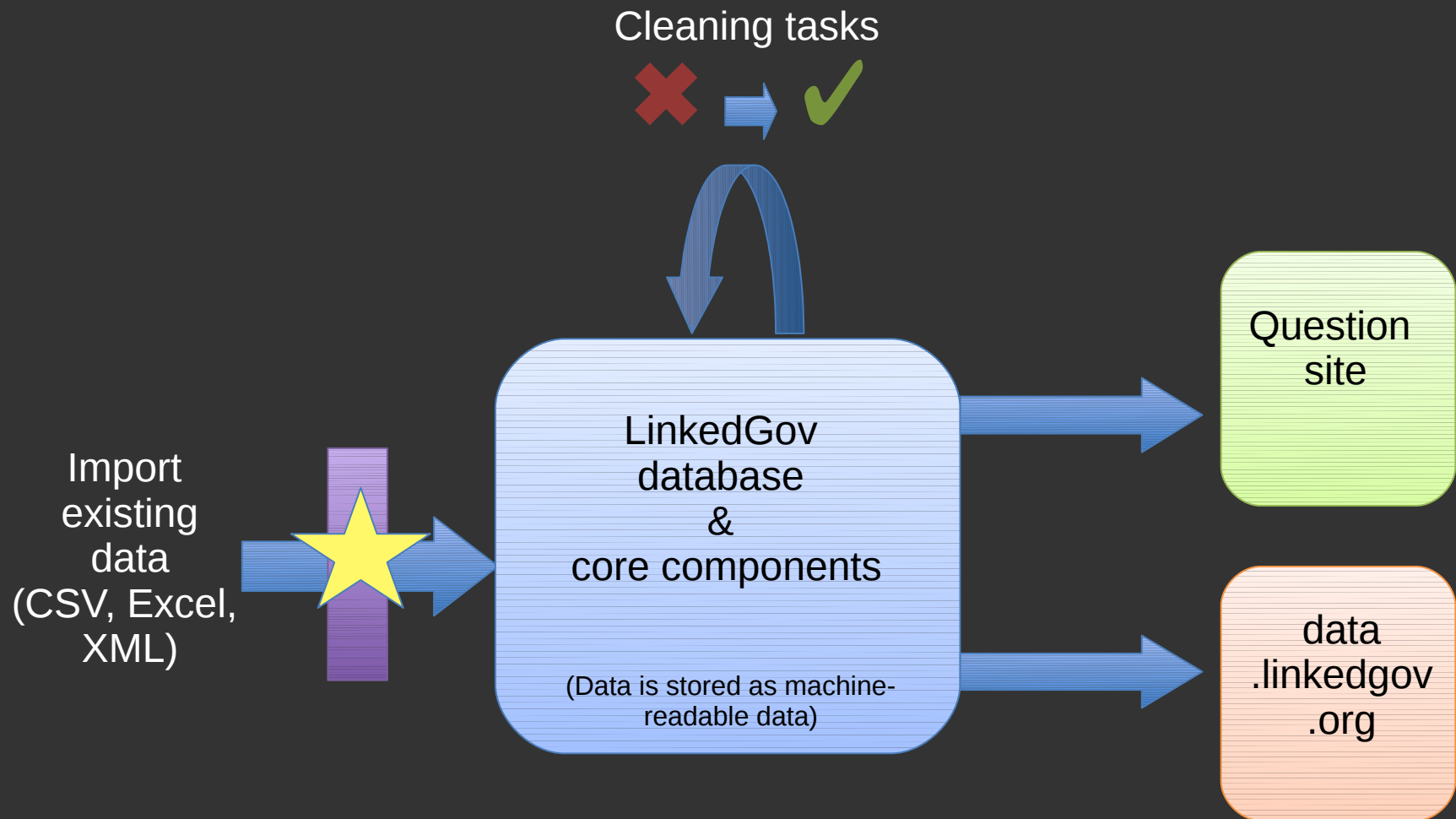
A little bit of help from
everyone

All in one place

Make the most of their skills
(technical or not, civil
service or citizen)



LinkedGov



Crowdsourced data cleaning

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



freeing Greater Manchester's public data

datagm.org.uk/package/atco-cif

Greater Manchester bus timetables d

Login · Register

What are you looking for?

search

Home Data News Uses of data Resources About DataGM

View Comments & Questions History

Bus stops and schedules

 [Subscribe »](#)

Source: <http://store.datagm.org.uk/sets/TfGM/>

Bus stop and bus time data for all services within the Greater Manchester boundary. Known technically as "ATCO-CIF".

Downloads & Resources

Description	Format	Hash
ATCO-CIF Data (downloaded 189 times)	ZIP	

This is a list of all known formats and datasets for *Bus stops and schedules*. If you know of another (CSV, SPARQL end-point etc.) please edit this page and add it to the list.

First time at DataGM - Data Greater Manchester?
DataGM - Data Greater Manchester is a catalogue for data. Click here to find out more ...

Tags

- bus-schedule
- bus
- bus-stops
- metrolink

Groups

- Transport for Greater Manchester

The data – bus stop locations



The screenshot shows a web browser window with the address bar displaying `data.gov.uk/dataset/naptan`. The page header includes the HM Government logo and the text "data.gov.uk^{BETA} Opening up government". A navigation bar contains links for Home, Data (highlighted), Apps, Consultation, Forum, Blogs, and Ideas. The main content area features the title "National Public Transport Access Nodes (NaPTAN)" in large, bold, dark blue text. Below the title, it states "Posted by Department for Transport on 03/04/2012" followed by "37 comments" and a five-star rating. A paragraph of text describes NaPTAN as a GB national system for uniquely identifying all the points of access to public transport in GB, used by other UK standards and information systems.

data.gov.uk^{BETA}
Opening up government

Home Data Apps Consultation Forum Blogs Ideas

National Public Transport Access Nodes (NaPTAN)

Posted by Department for Transport on 03/04/2012 | 37 comments | ★★★★★

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

Functional demo

Enter latitude/longitude values or OS grid references into the test boxes to try out the calculations:

WGS84 Lat/Lon	<input n"="" type="text" value="52°39'28.7158"/>	<input e"="" type="text" value="001°42'57.7383"/>
OSGB36 Lat/Lon	<input n"="" type="text" value="52°39'27.2459"/>	<input e"="" type="text" value="001°43'04.4690"/>
OS Grid Ref	TG 51409 13177	= 651409,313177

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



The full data journey

1 Location from the mobile phone in lat/long



2

Functional demo

Enter latitude/longitude values or OS grid references into the test boxes to try out the calculations:

WGS84 Lat/Lon

OSGB36 Lat/Lon

OS Grid Ref = 651409,313177

Conversion of the user's location to Easting/Northing

3

	Standard	Standard	Standard	Standard
1	LocalityCentre	GridType	Easting	Northing
2	0	U	364208	176288
3	0	U	358596	172813
4	0	U	358886	177828
5	0	U	358839	177839
6	0	U	358980	177882
7	0	U	359009	177863
8	0	U	358839	177529
9	0	U	358945	173526

Lookup of bus stops near the user

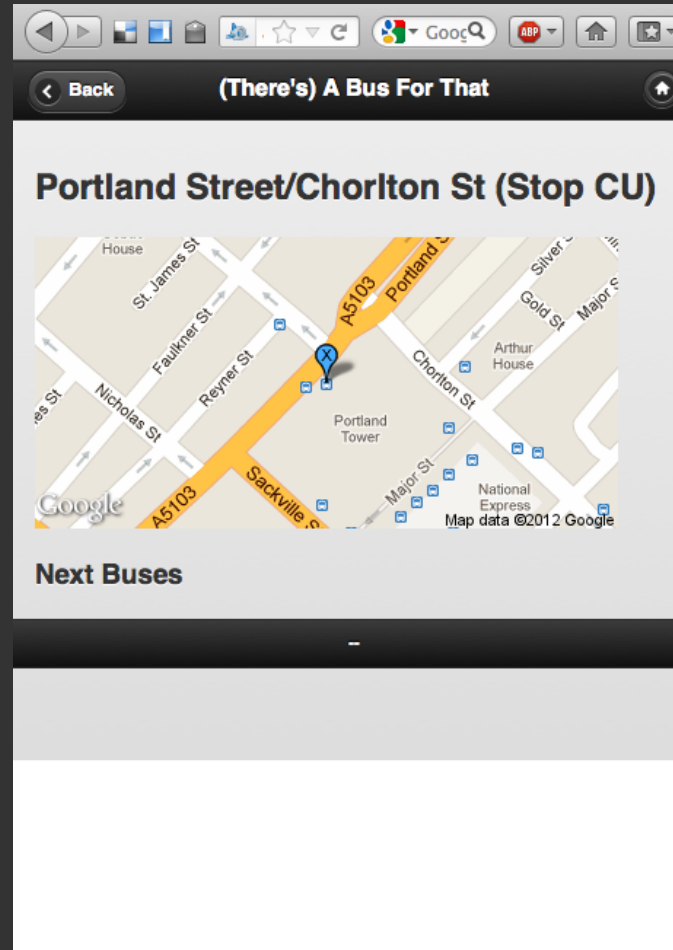
4

ATCO-CIF 500TFGM GM_AIM 20120410145510
ZLGM__8_S111
ZD2012010320120415Mondays to Fridays
ZSGM__8_8BURNLEY - BACUP - RAWTENSTALL - BURY
ZAN250020330 BURNLEY, Bus Station P1W0T1
ZAN2500IMG2776 TOWNELEY, Golf Course P1W0T1
ZAN2500DCL5776 WEIR, The Moorlands P1W0T0
ZAN2500DCL5780 WEIR P1W0T1
ZAN2500DCL704 BACUP Stand B P1W0T1
ZAN2500IMG1625 STACKSTEADS P1W0T0
ZAN25001433 WATERFOOT, Late Shop P1W0T1
ZAN2500IMG2517 RAWTENSTALL, Bus Station P1W0T1
ZAN25001406 RAWTENSTALL Fire Station P1W0T0
ZAN1800NC33551 Rochdale Road/Seven Stars P1W0T1
ZAN1800NC33541 Rochdale Road/Roach Hotel P1W0T0
ZAN1800NC01871 Rochdale Road/Pimhole Road P1W0T0
ZAN1800NC33531 Rochdale Road/Iceland P1W0T0
ZAN1800NC20431 Rochdale Road/Robin Hood P1W0T0

Producing the relevant bus routes, filtering by time



Info to the user: Mobile-friendly HMTL 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 - mihalysot

<http://www.flickr.com/photos/mihalysot/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/4113274705/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

