HM Government

UNIVERSITY OF Southampton
School of Electronics and Computer Science

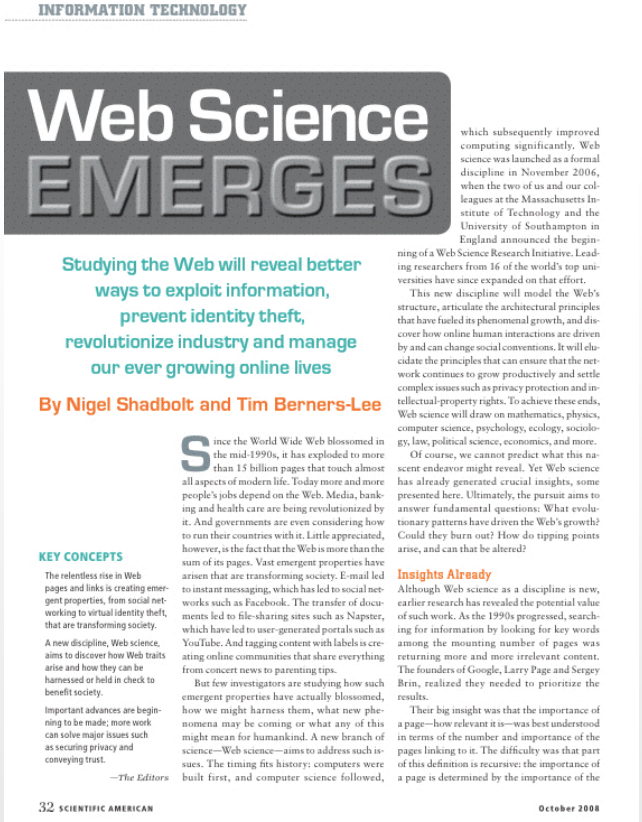
Open Government Data: A Case Study in Web Science

Professor Nigel Shadbolt

Twitter Nigel_Shadbolt

8th July 2010
OII Summer School
Oxford

The Emergence of Web Science



INFORMATION TECHNOLOGY

Web Science EMERGES

Studying the Web will reveal better ways to exploit information, prevent identity theft, revolutionize industry and manage our ever growing online lives

By Nigel Shadbolt and Tim Berners-Lee

KEY CONCEPTS

The relentless rise in Web pages and links is creating emergent properties, from social networking to virtual identity theft, that are transforming society.

A new discipline, Web science, aims to discover how Web traits arise and how they can be harnessed or held in check to benefit society.

Important advances are beginning to be made; more work can solve major issues such as securing privacy and conveying trust.

—*The Editors*

Since the World Wide Web blossomed in the mid-1990s, it has exploded to more than 15 billion pages that touch almost all aspects of modern life. Today more and more people's jobs depend on the Web. Media, banking and health care are being revolutionized by it. And governments are even considering how to run their countries with it. Little appreciated, however, is the fact that the Web is more than the sum of its pages. Vast emergent properties have arisen that are transforming society. E-mail led to instant messaging, which has led to social networks such as Facebook. The transfer of documents led to file-sharing sites such as Napster, which have led to user-generated portals such as YouTube. And tagging content with labels is creating online communities that share everything from concert news to parenting tips.

But few investigators are studying how such emergent properties have actually blossomed, how we might harness them, what new phenomena may be coming or what any of this might mean for humankind. A new branch of science—Web science—aims to address such issues. The timing fits history: computers were built first, and computer science followed, which subsequently improved computing significantly. Web science was launched as a formal discipline in November 2006, when the two of us and our colleagues at the Massachusetts Institute of Technology and the University of Southampton in England announced the beginning of a Web Science Research Initiative. Leading researchers from 16 of the world's top universities have since expanded on that effort.

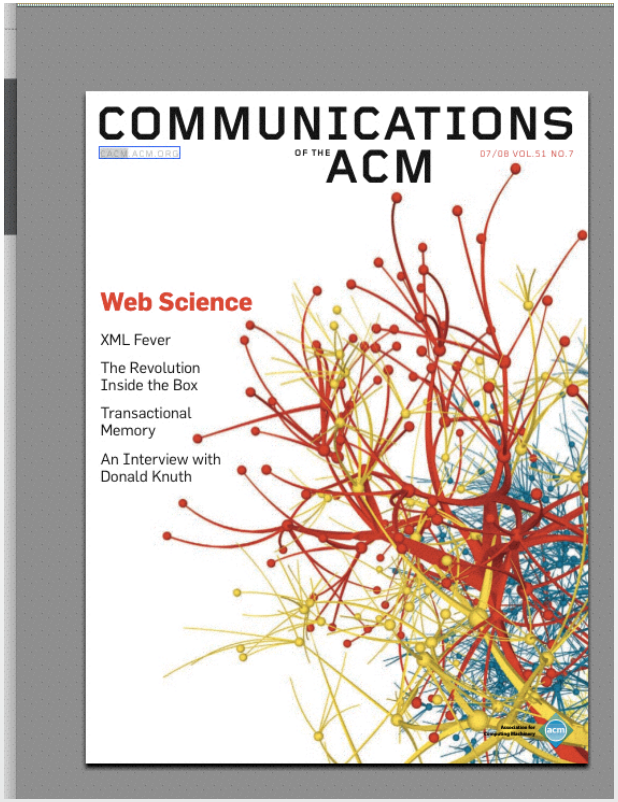
This new discipline will model the Web's structure, articulate the architectural principles that have fueled its phenomenal growth, and discover how online human interactions are driven by and can change social conventions. It will elucidate the principles that can ensure that the network continues to grow productively and settle complex issues such as privacy protection and intellectual-property rights. To achieve these ends, Web science will draw on mathematics, physics, computer science, psychology, ecology, sociology, law, political science, economics, and more.

Of course, we cannot predict what this nascent endeavor might reveal. Yet Web science has already generated crucial insights, some presented here. Ultimately, the pursuit aims to answer fundamental questions: What evolutionary patterns have driven the Web's growth? Could they burn out? How do tipping points arise, and can that be altered?

Insights Already

Although Web science as a discipline is new, earlier research has revealed the potential value of such work. As the 1990s progressed, searching for information by looking for key words among the mounting number of pages was returning more and more irrelevant content. The founders of Google, Larry Page and Sergey Brin, realized they needed to prioritize the results.

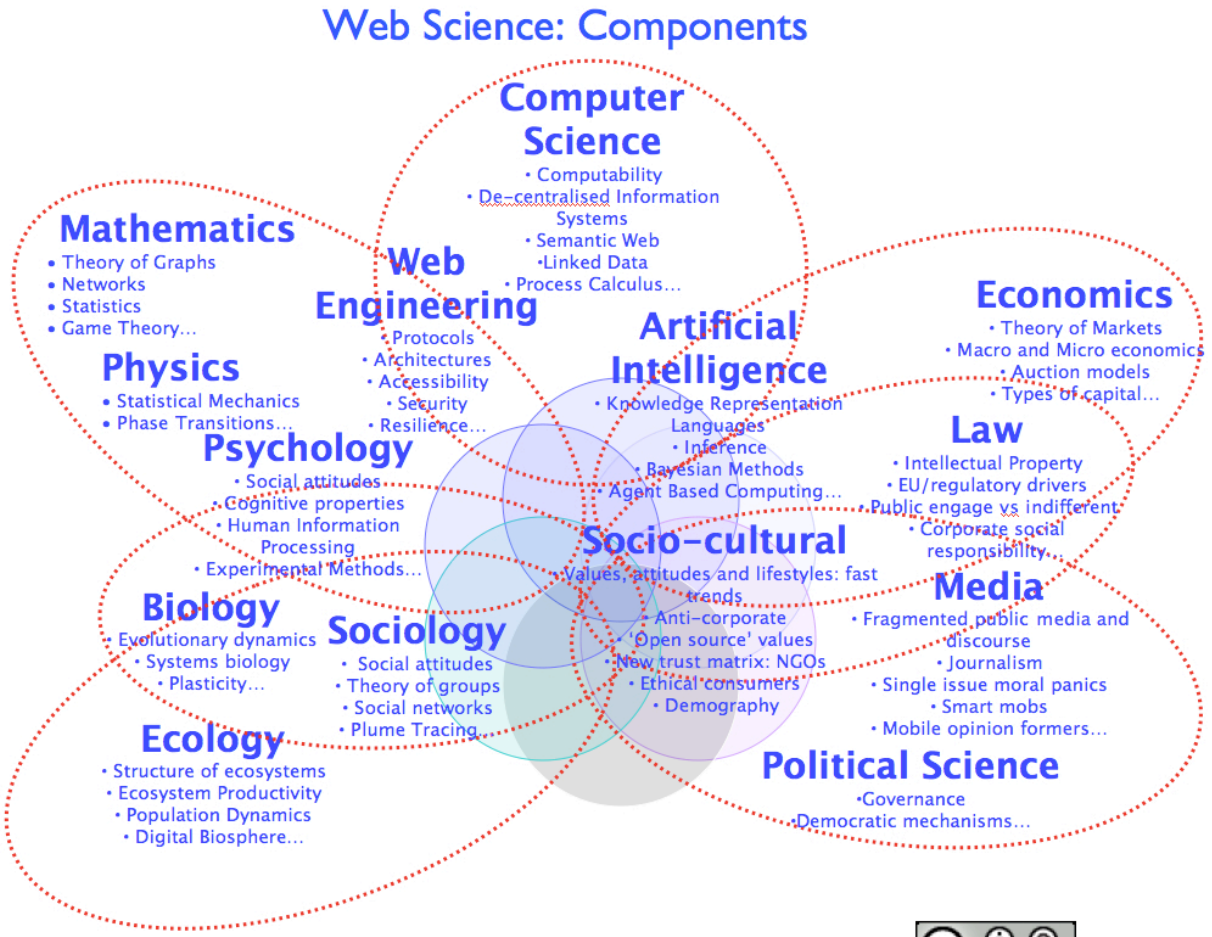
Their big insight was that the importance of a page—how relevant it is—was best understood in terms of the number and importance of the pages linking to it. The difficulty was that part of this definition is recursive: the importance of a page is determined by the importance of the



Web Science
is about additionality

Not the union of the disciplines

But more than their intersection

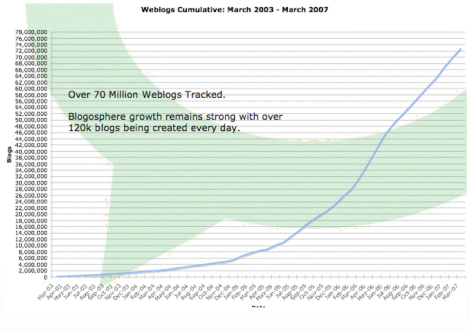


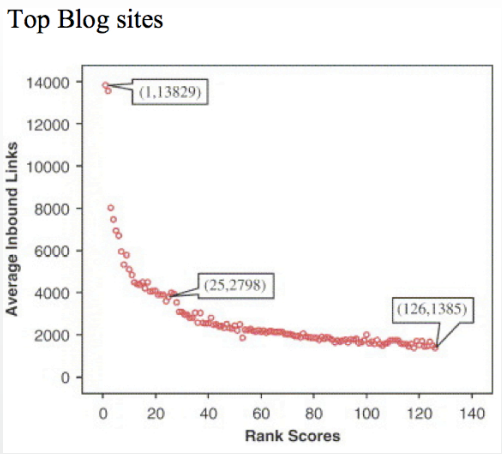
Web Science
essentials

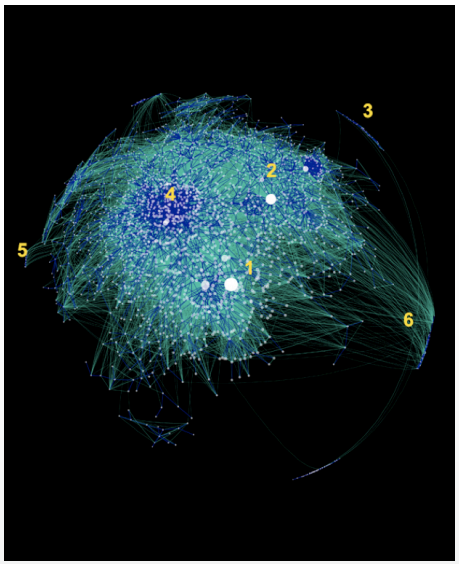
- Web architecture is a set of simple principles that give rise to complexity
- Simple micro rules give rise to complex macro phenomena
- Many systems in nature are like this
- A need for a systems oriented view of the Web and its ecosystem
- A process that accounts for design/ engineering and analysis/science
- An approach that is focused on the social and technical

Web Science
an example

- When and why did the blogosphere take off?
- Who blogs and how much?
- Why do they blog?
- Is it the same everywhere?

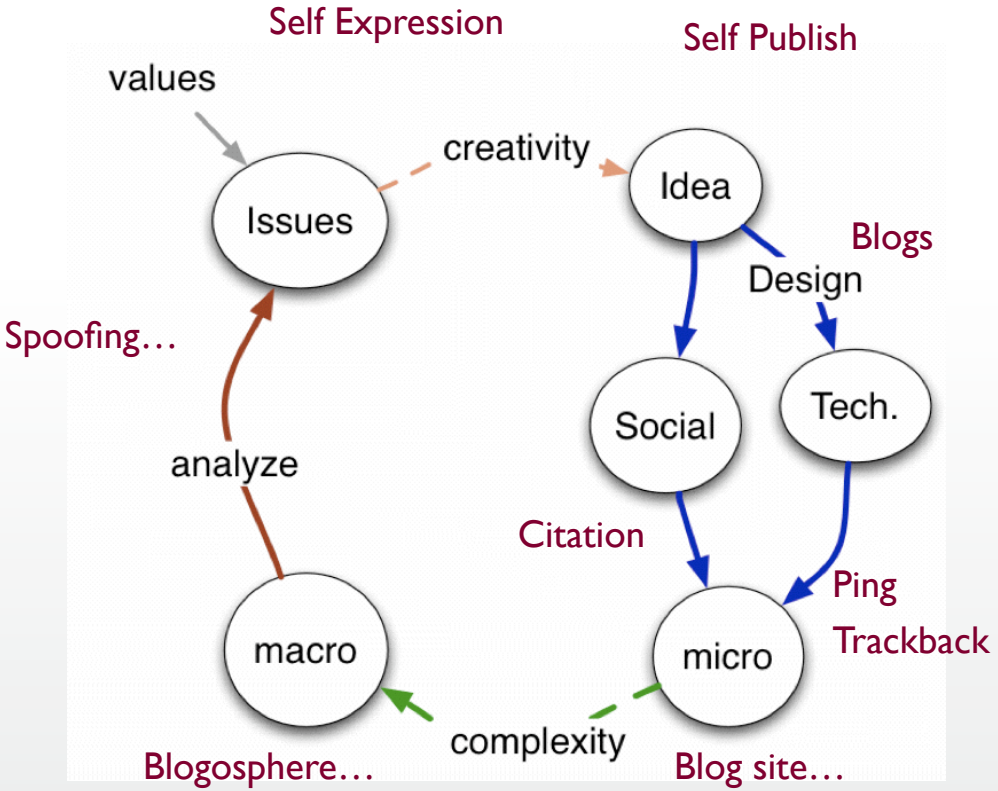






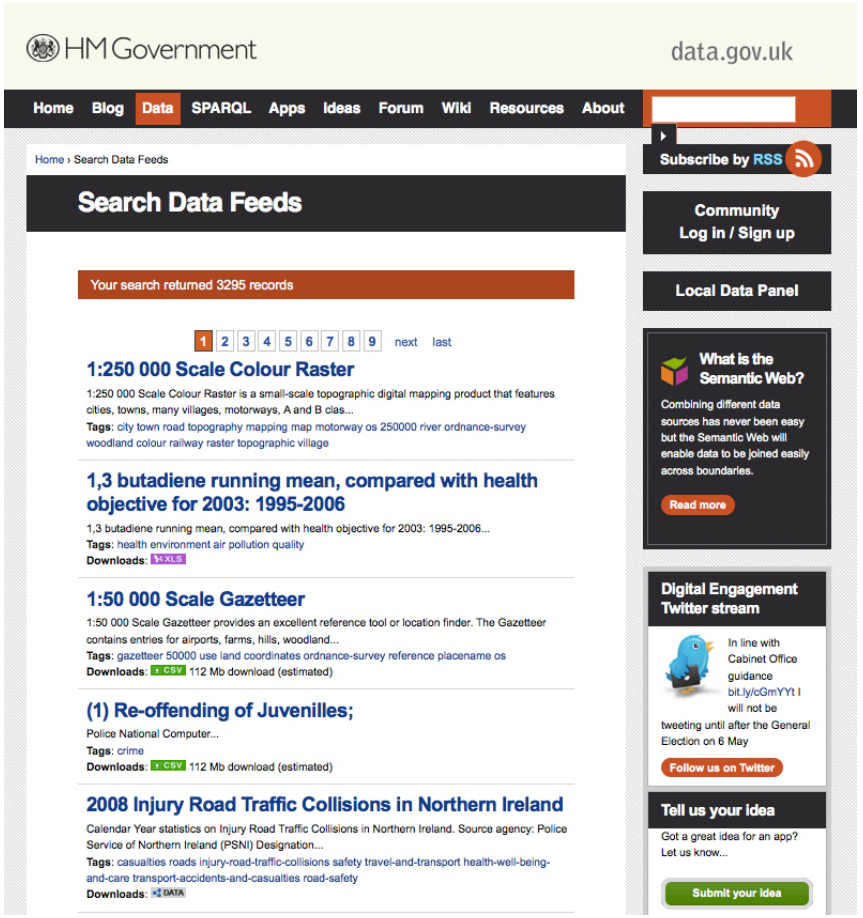
Web Science
what we need to understand

- creative innovation
- design and engineering
- the social and the technical
- interpretation and analysis



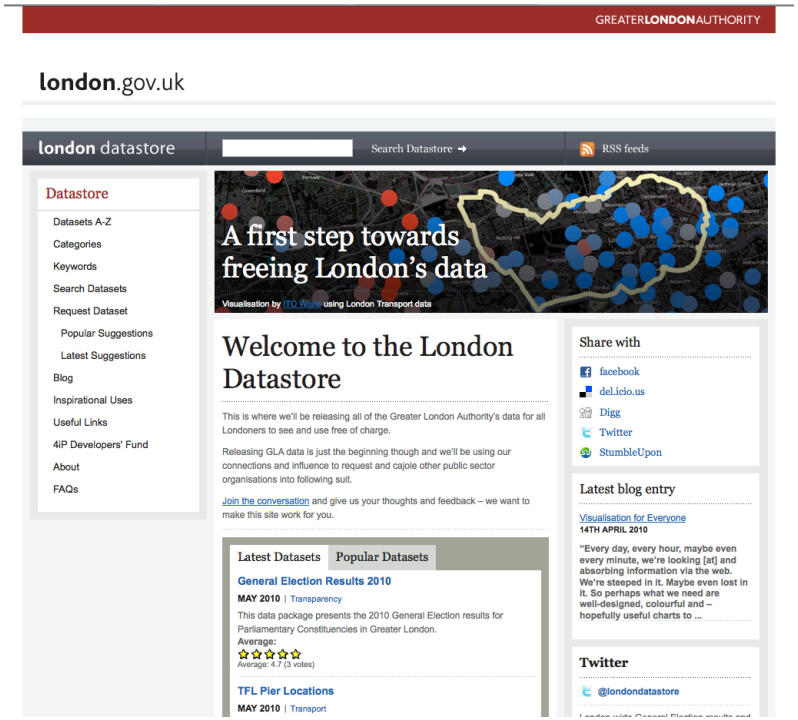
Context

- Sometimes you can catch a wave
- Governments, cities, organizations, individuals - all releasing data
- Our emphasis was non-personal public data
- Wouldn't think that this would set your heart racing – but it does!

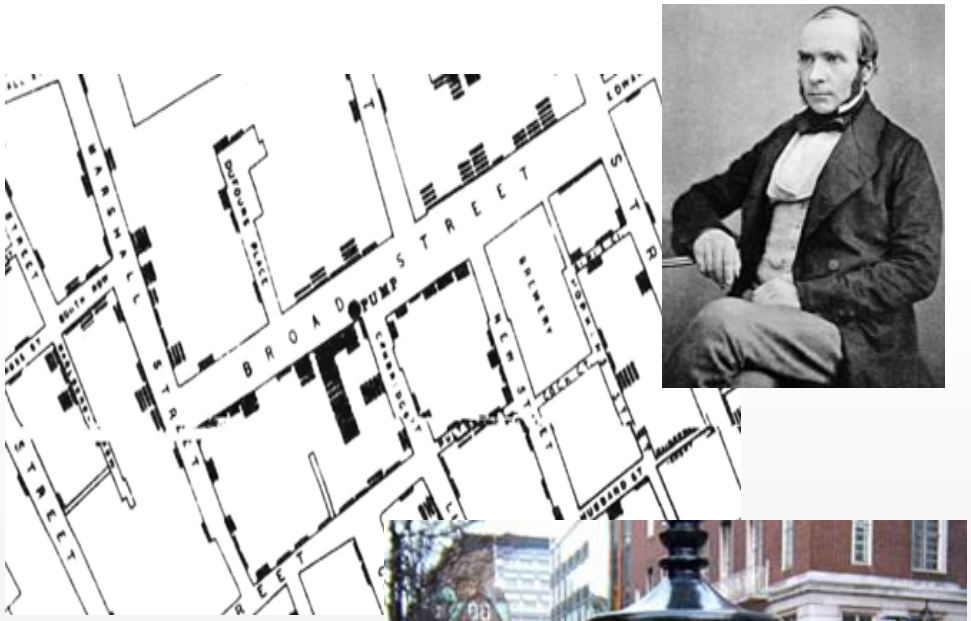


Context

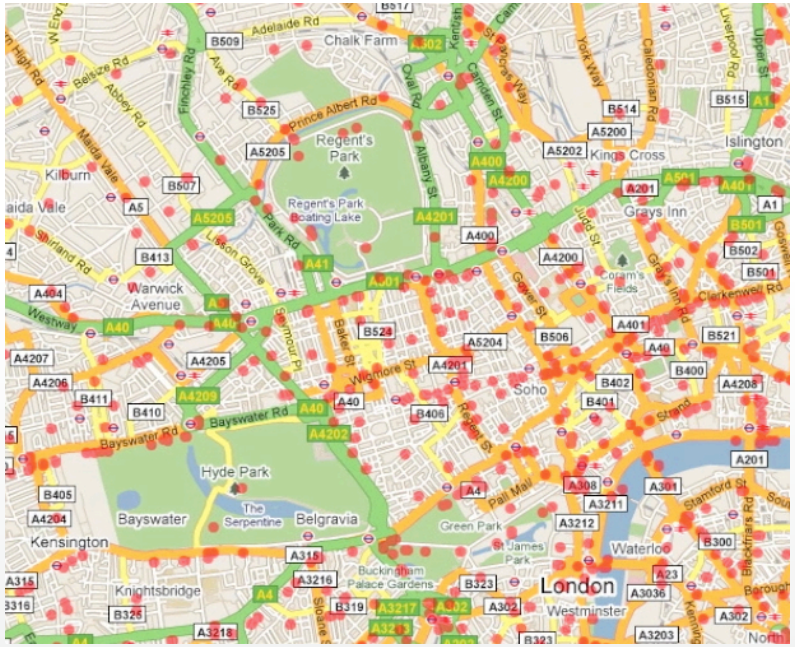
- Sometimes you can catch a wave
- Governments, cities, organizations, individuals - all releasing data
- Our emphasis was non-personal public data
- Wouldn't think that this would set your heart racing – but it does



The Power of Open Public Data



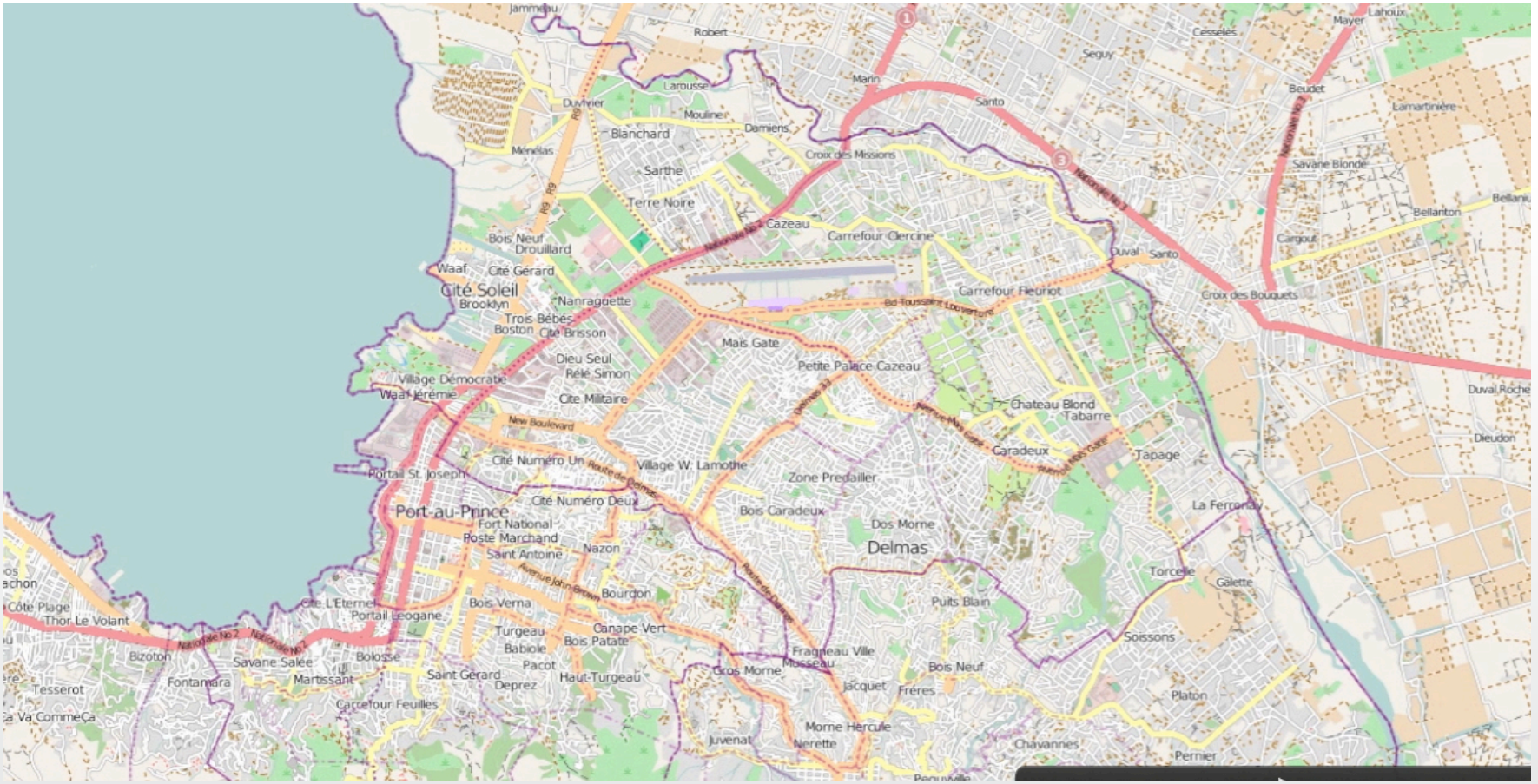
Cholera cases



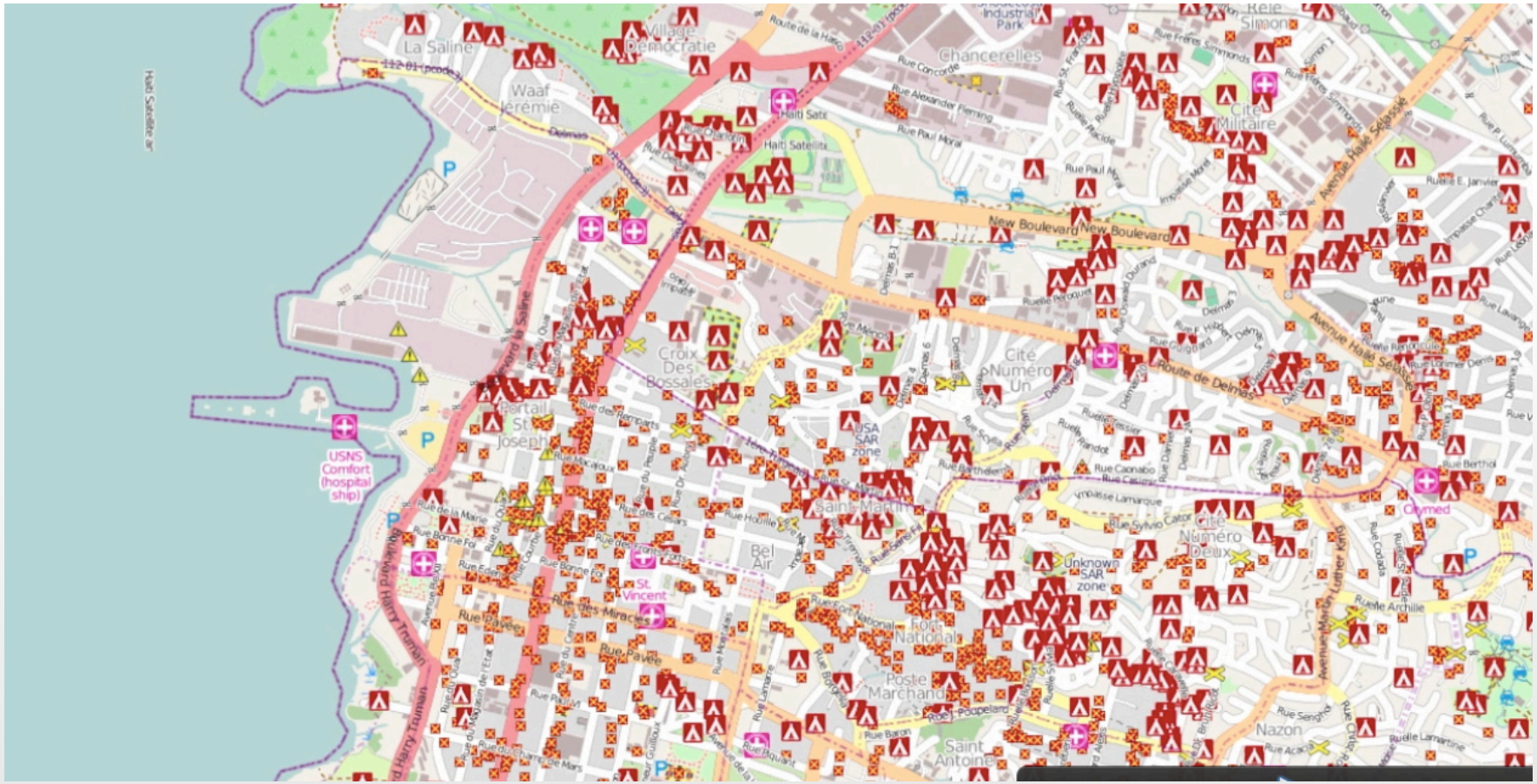
Bicycling traffic accidents

The Power of Open Public Data
Port-au-Prince < 12 Jan 2010

The Power of Open Public Data
Port-au-Prince ≅ 25 Jan 2010



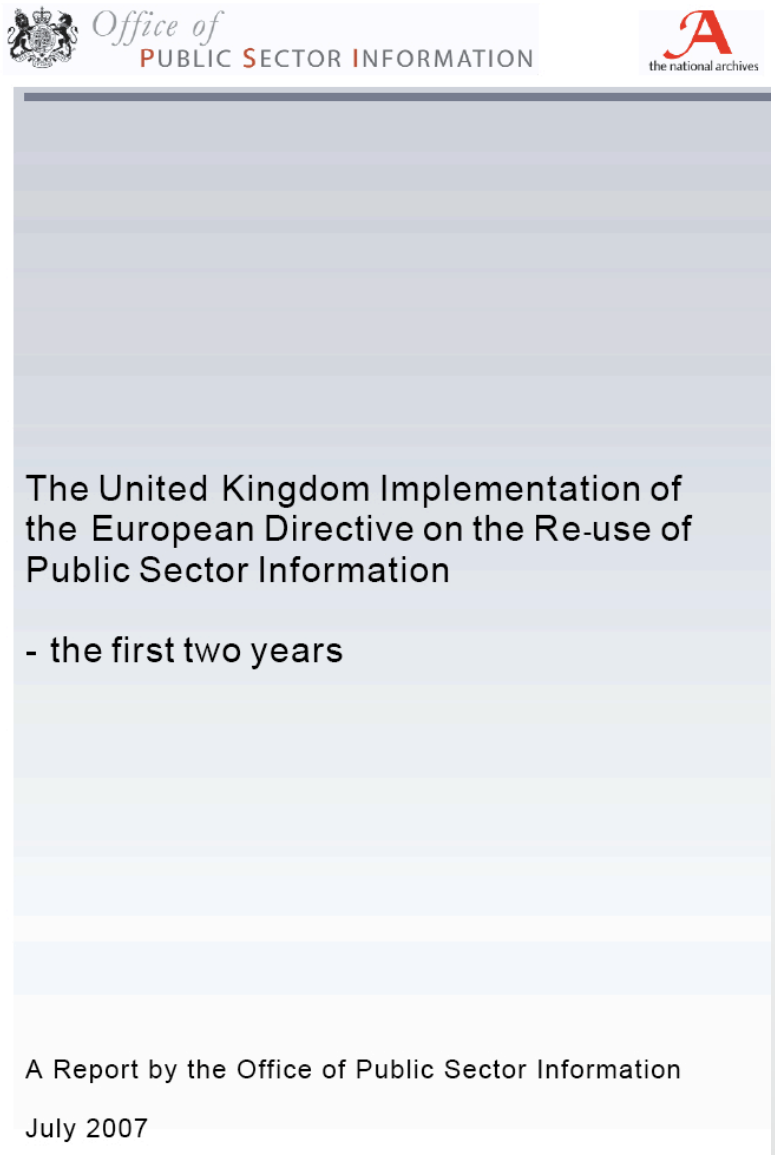
The Power of Open Public Data
Port-au-Prince ≅ 25 Jan 2010



AKTive PSI

Using next generation Web technology to improve the delivery of policy and public services across Government (2005-06)

The approach became embedded



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),[36] 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[37]. OPSI is using this technology in the following ways:

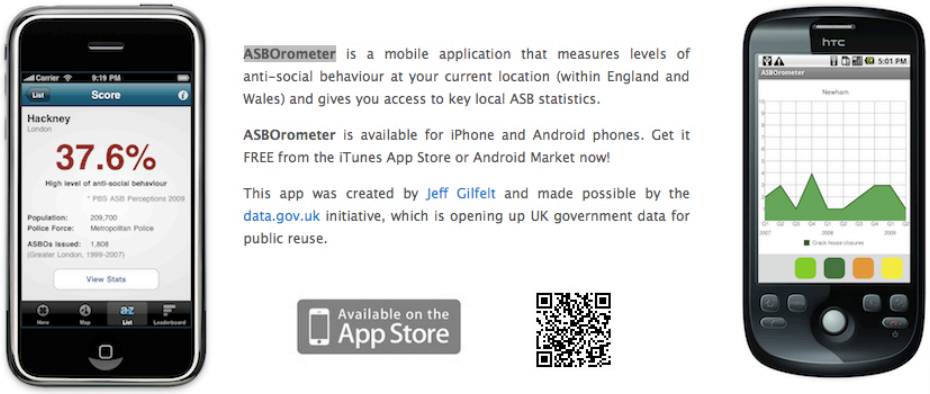
MPDP: Making Public Data Public

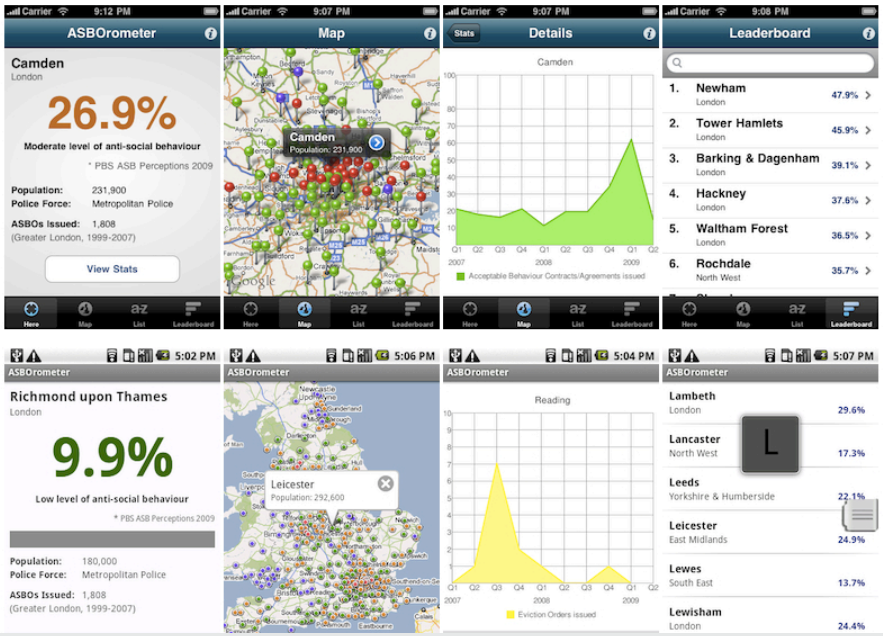
- Berners-Lee and Shadbolt – June 2009 Appointed HMG Information Advisors, 2010 Appointed to HMG Transparency Board
- Context – Power of Information Taskforce
- Previous work with OPSI
- We didn't know what we couldn't do!
- Strategic Drivers
  - Transparency and accountability
  - Economic and Social Value
  - Public Service Improvement
  - New Industries New Jobs

What data?

"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.

The Power of OGD - ASBOrometer





The Power of OGD – Where's the dentist

Home
News
Apps
- Find GPs
- Find Pharmacies
- Find Postboxes
- Find Pubs
- Find Toilets
- SF Trees
- UK Dentists

Support
Contact

follow us on twitter

Available on the App Store

UK Dentists

Find your nearest NHS dentist quickly and easily!

The National Health Service has over 7,500 dental surgeries across England but finding one can be a difficult and time consuming task. This app is based on data provided by the UK's Health and Social Care Information Centre (HSCIC)* and covers 99.4% of all registered NHS dental surgeries in England.

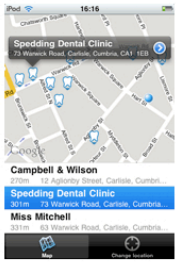
Search for an NHS dentist around your current location, or look for one in another area (where you may be moving to, or for a friend) via a simple place name or postcode search. Distances can be displayed in metric or imperial units (configured in the application preferences).

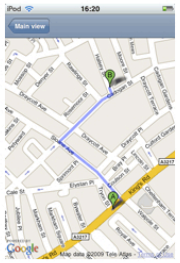
UK Dentists uses the built in GPS on the iPhone to get your current location, but can also locate an iPod Touch in built-up WiFi areas. If your location cannot be determined automatically then you can always enter your post code to get started. Please note that an internet connection is required to get location and map data.

* Data reproduced under the terms of the Office of Public Sector Information (OPSI) Click-Use Licence.

Application Screenshots

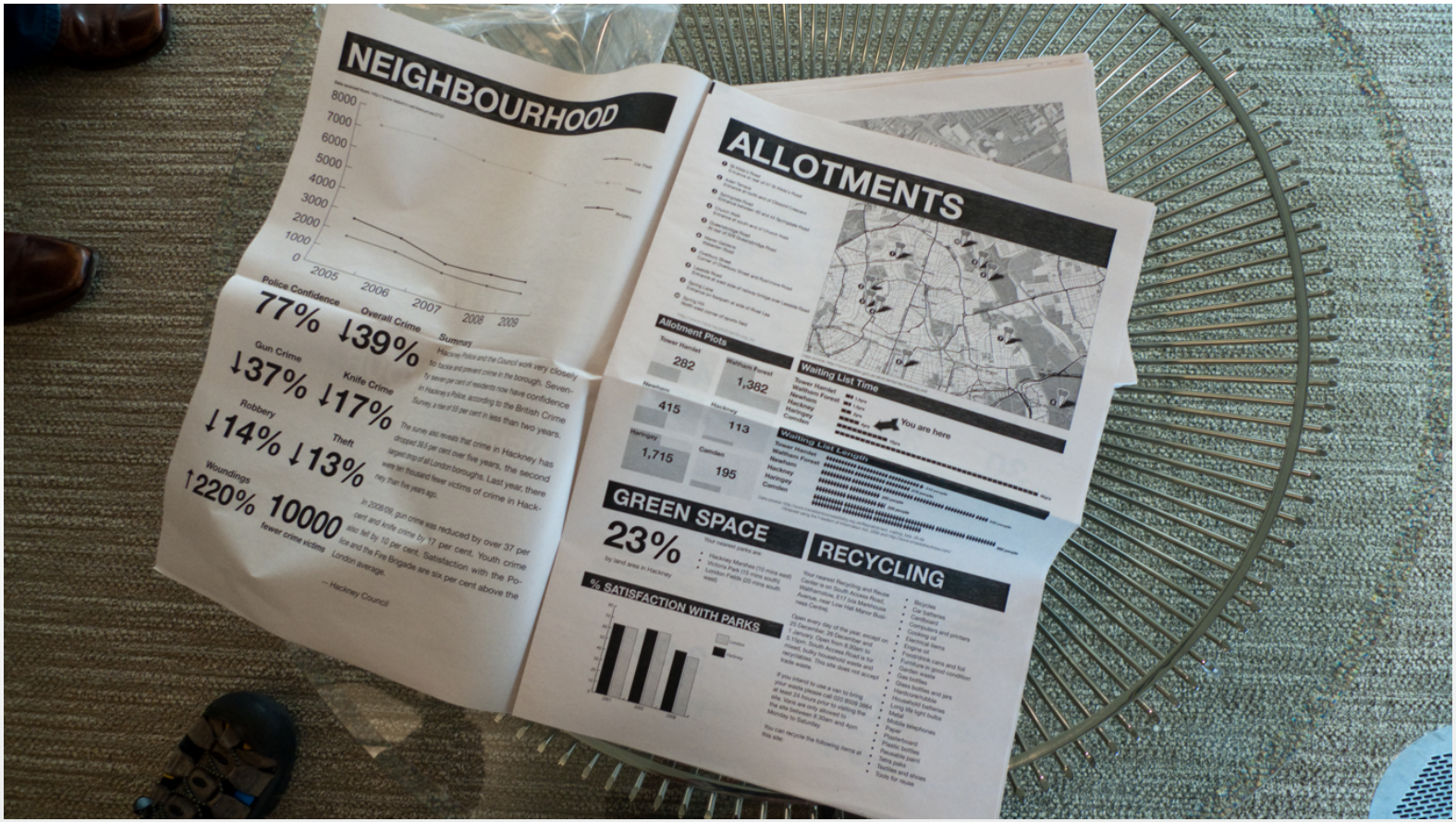
Click on any of the images to see a larger version



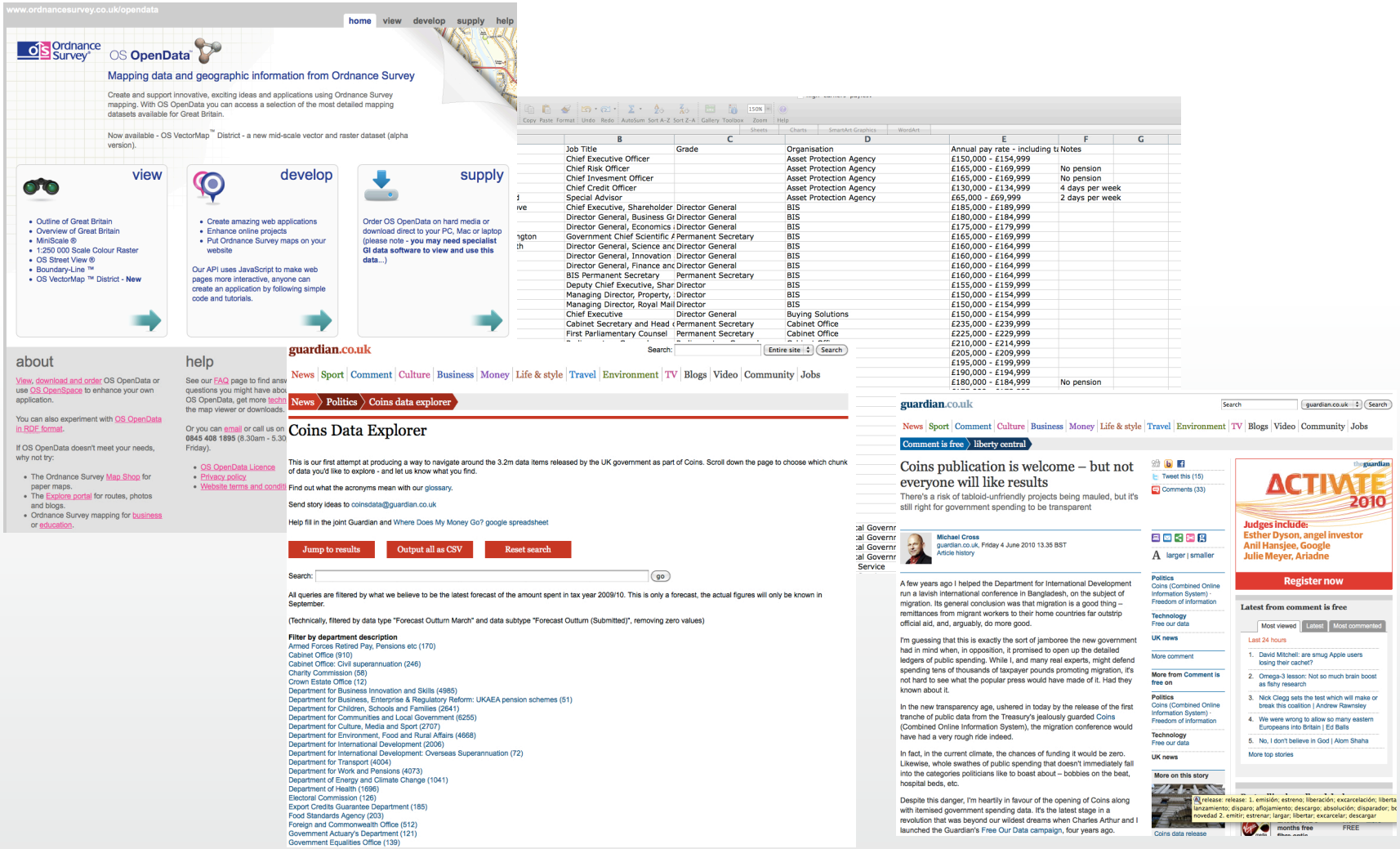


Buy on iTunes

The Power of OGD – Post Code Paper

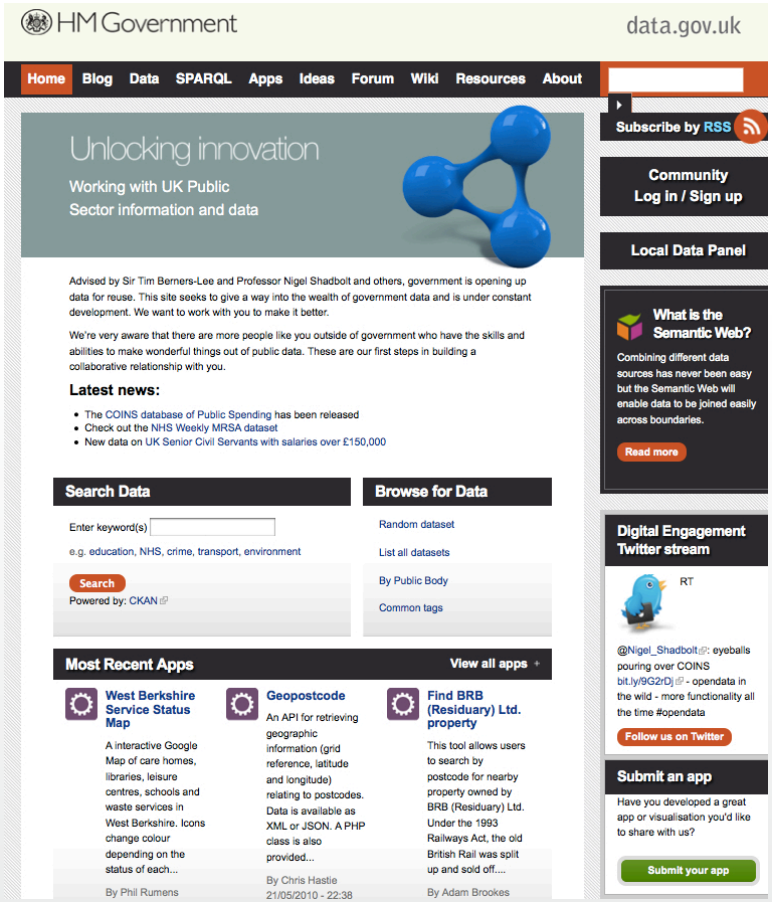


The Power of OGD – Maps, Spending…



How data.gov.uk

- data.gov.uk > 4000 datasets on it
  - open source and open standards
  - key data sets released inc OS OpenData
- a new open licence for Crown Copyright information
- a network of ‘guerrillas’ within Whitehall
- a community of external data users and developers
- a group of experts on creating, transparent open and standardised formats for data



How? Publish and Praise

http://www.w3.org/DesignIssues/LinkedData.html

★	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

How? Public Data Principles

1. 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;
2. Public data will be published in reusable, machine-readable form;
3. Public data will be released under the same open licence which enables free reuse, including commercial reuse;
4. Public data will be available and easy to find through a single easy to use online access point (http://www.data.gov.uk)
5. :

How? Public Data Principles

1. 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;
2. Public data will be published in reusable, machine-readable form;
3. Public data will be released under the same open licence which enables free reuse, including commercial reuse;
4. Public data will be available and easy to find through a single easy to use online access point (http://www.data.gov.uk)
5. Public data will be published using open standards and following the recommendations of the World Wide Web Consortium;
6. Public data underlying the Government's own websites will be published in reusable form for others to use;
7. Public data will be timely and fine grained;
8. Release data quickly, and then republish it in linked data form;
9. Public data will be freely available to use in any lawful way;
10. Public bodies should actively encourage the re-use of their public data; and
11. Public bodies should maintain and publish inventories of their data holdings.

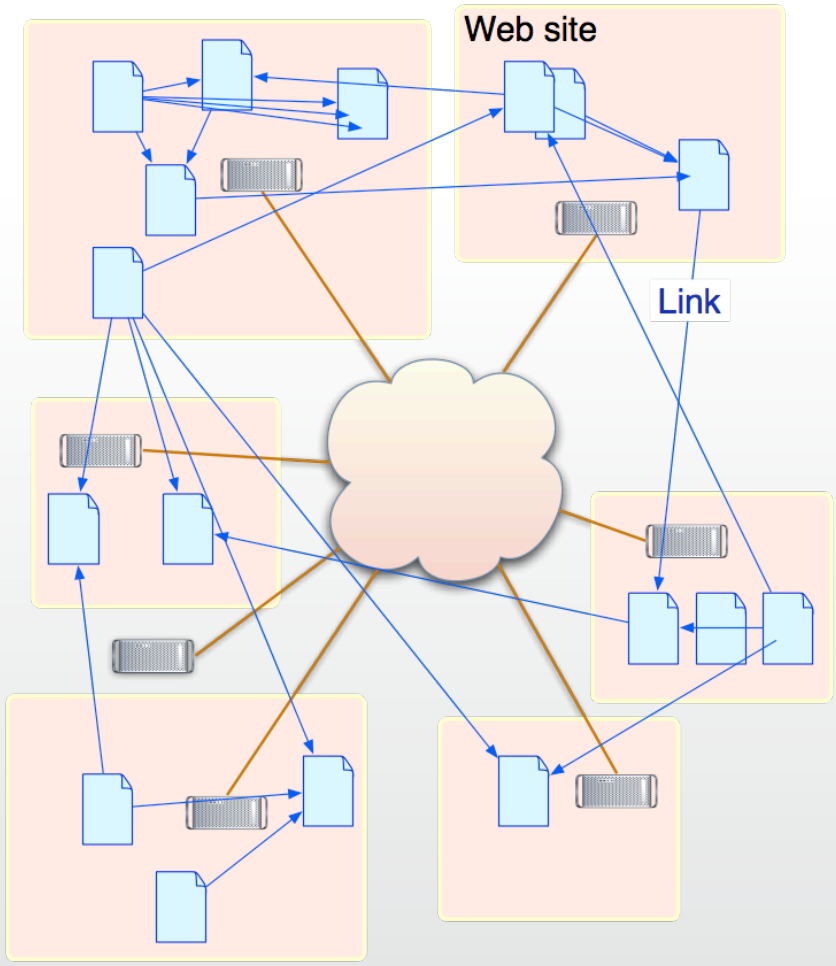
Why Linked Data

- Light weight ontologies
- Lots of them - mix vocabs
- Four Micro Principles
  - URIs
  - That are derefenceable
  - RDF/XML
  - Link

- Good for the hygiene of single datasets
- National digital infrastructure being built
- URIs for schools, roads, bus stops, post codes, admin bondaries...
- The link points will have a long tail
- Key data link points exist

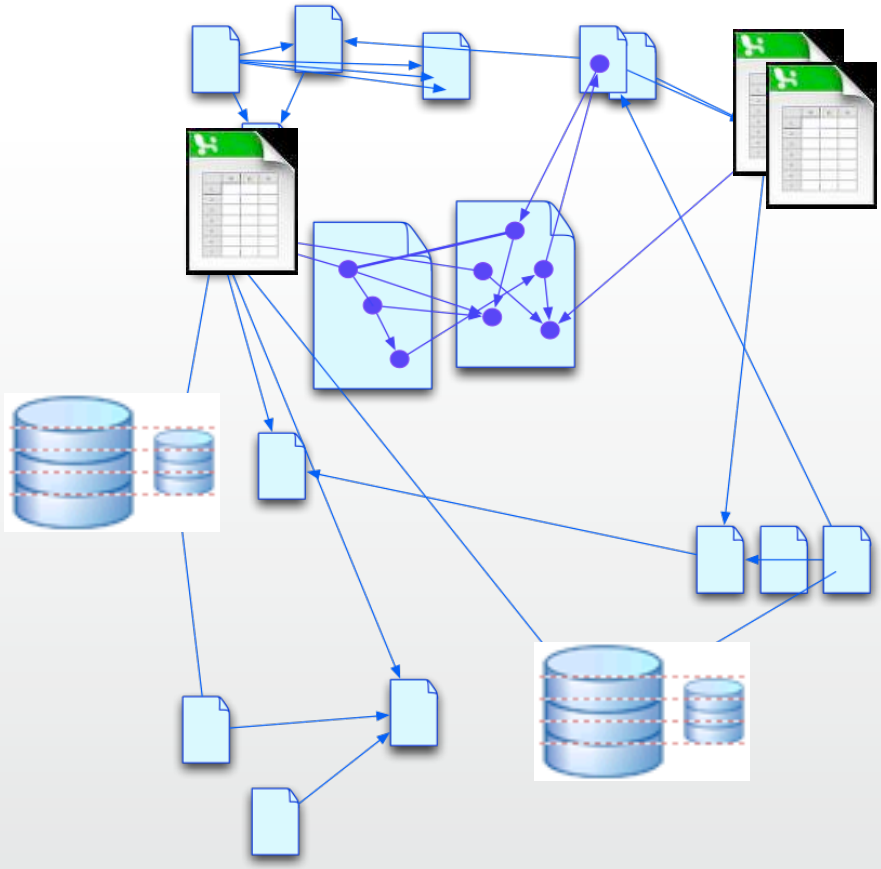
Web Science

anticipation – from a Web of Documents



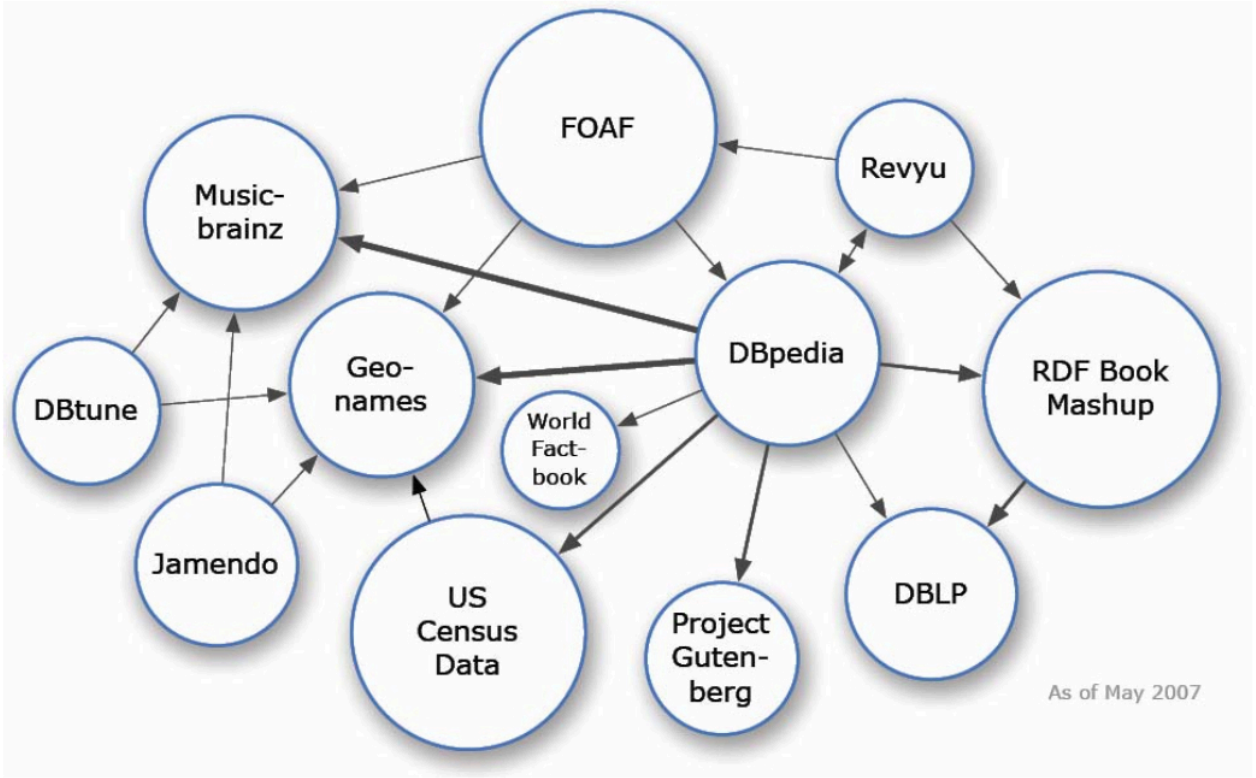
Web Science

anticipation – to a Web of Linked Data



Web Science

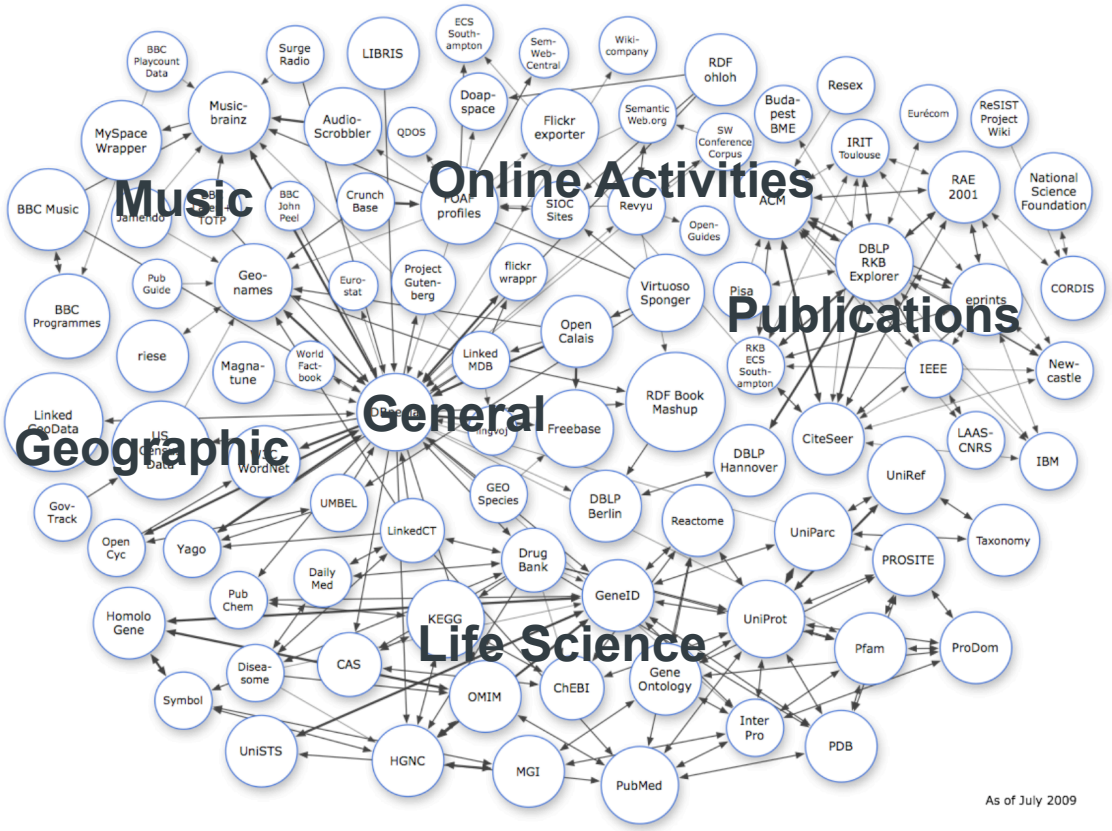
anticipation – linked data



Open Linked Data on the Web:
May 2007
500 million triples
120 million of links

Web Science

anticipation – linked data



Richard Cyganiak

Open Linked Data on the Web:
July 2009
Billions of triples and hundreds of millions of links

OGD - UK Linked Data

How good is my area ?

feedback ... please email us

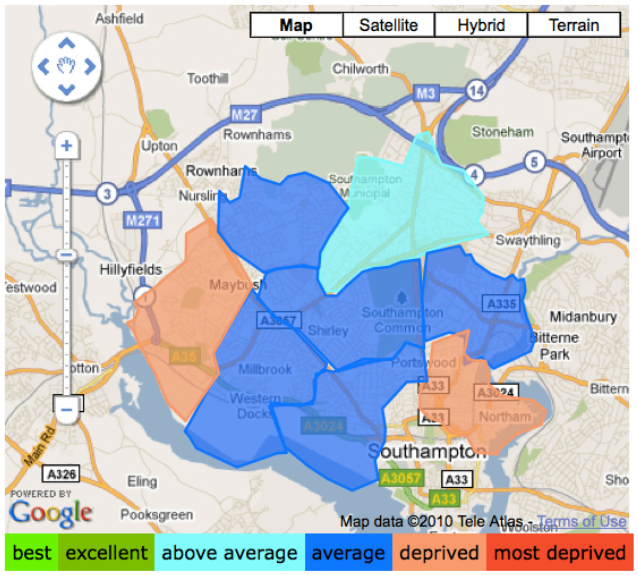
Sources:
statistics.data.gov.uk, data.ordnancesurvey.co.uk,
IMD Database, Boundary data from OS Open Data

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

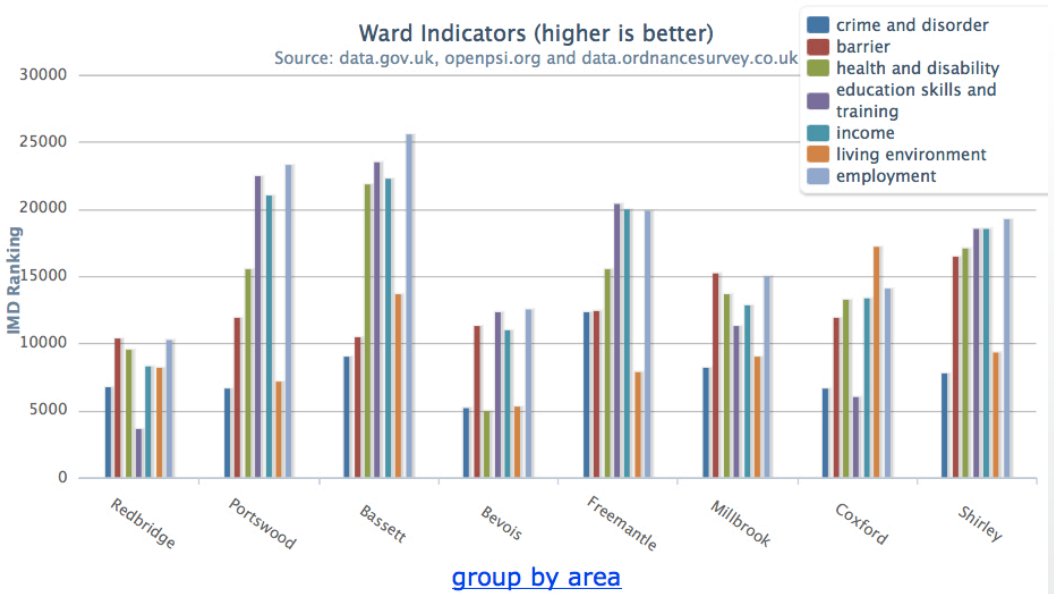
It displays a region and its surroundings showing how good are those areas based on the IMD database ranking.

shirley find

introduce an area name i.e: Portswood, Chilworth, Camden, ...



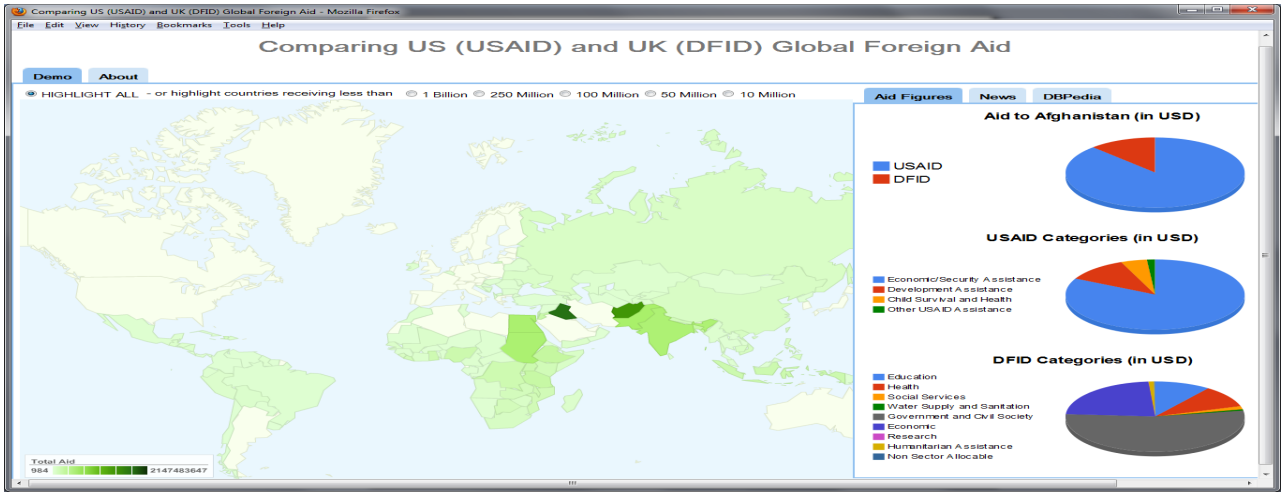
Area: Bevois



group by area

Linked OGD between Governments

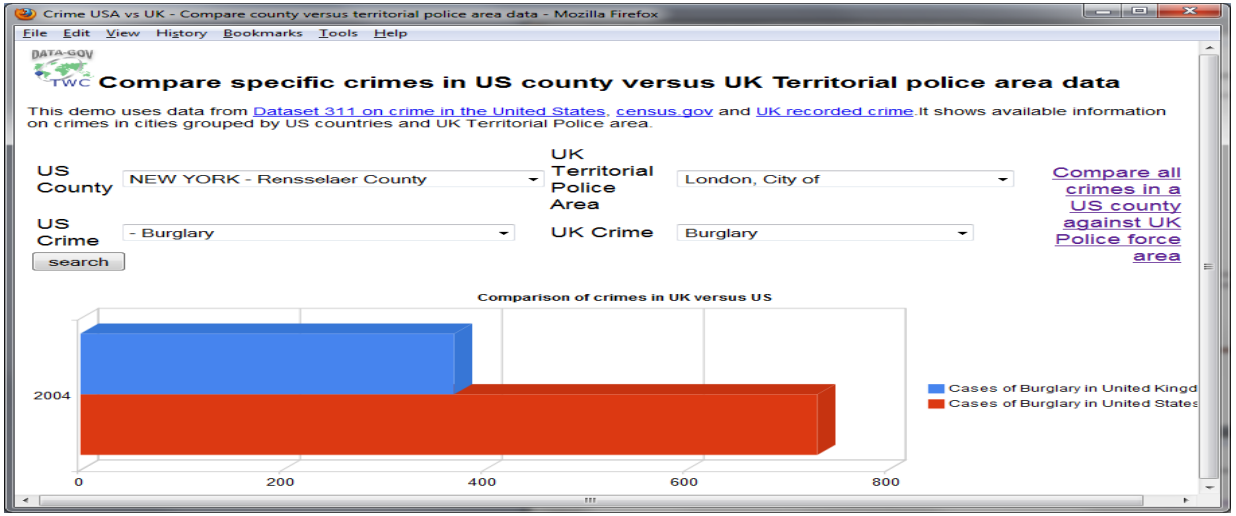
This application presents a mashup of foreign aid data (represented in US Dollars) from the United States Agency for International Development (USAID) and UK Department for International Development (DFID) for the 2007 US Fiscal Year.



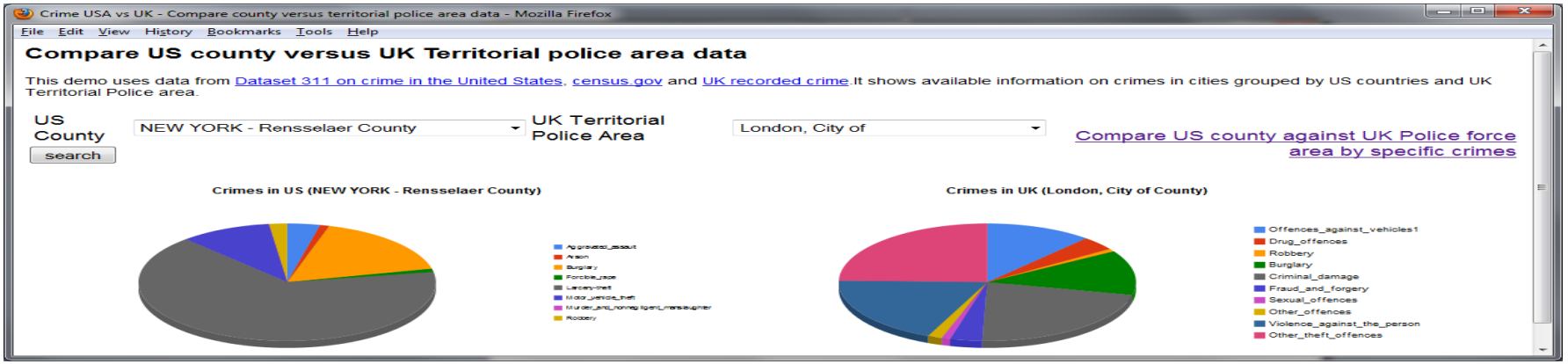
Users may retrieve foreign aid data for specific countries by clicking on a provided world map (shaded based on total combined contributions for USAID and DFID). Upon clicking on a desired country, three kinds of information are presented: Aid Figures, New York Times news, and wikipedia description.

http://data-gov.tw.rpi.edu/demo/linked/aidviz-1554-10030.html

Linked OGD between Governments



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



Why not

Loss of licensing revenue

Loss of control

Legal challenges

Unwelcomed exposure

Procedural changes

Privacy

National security

Complexity

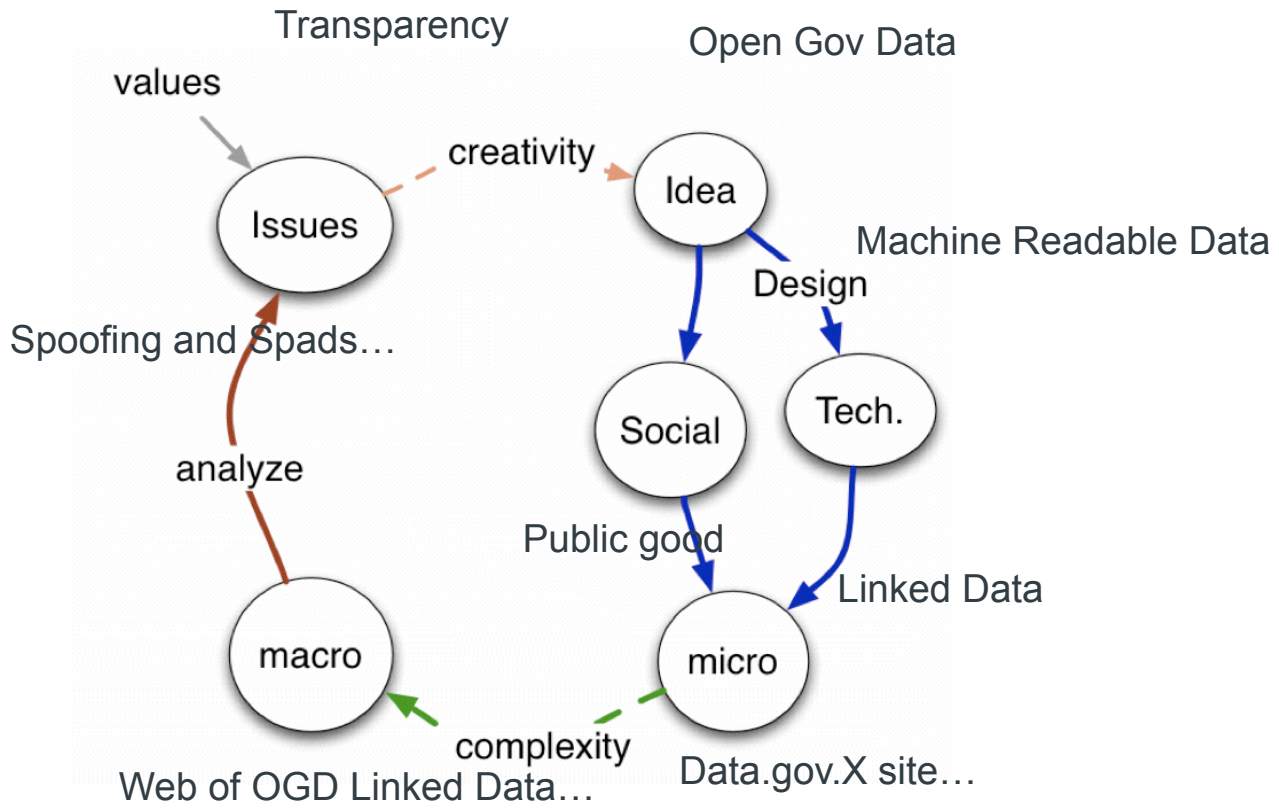
Investment

Capacity building required

Authenticity, quality

Customer service

Why

- More transparency – what is going on?
- More accountability – and is this acceptable?
- More localism – because it matters to me, my family, my locale
- More economic and social capital – which generates exploitation opportunities
- More engagement – and we can collect, contribute and improve Government's own data and information
- More argument –supporting evidenced based policy

OGD - Web Science Perspective