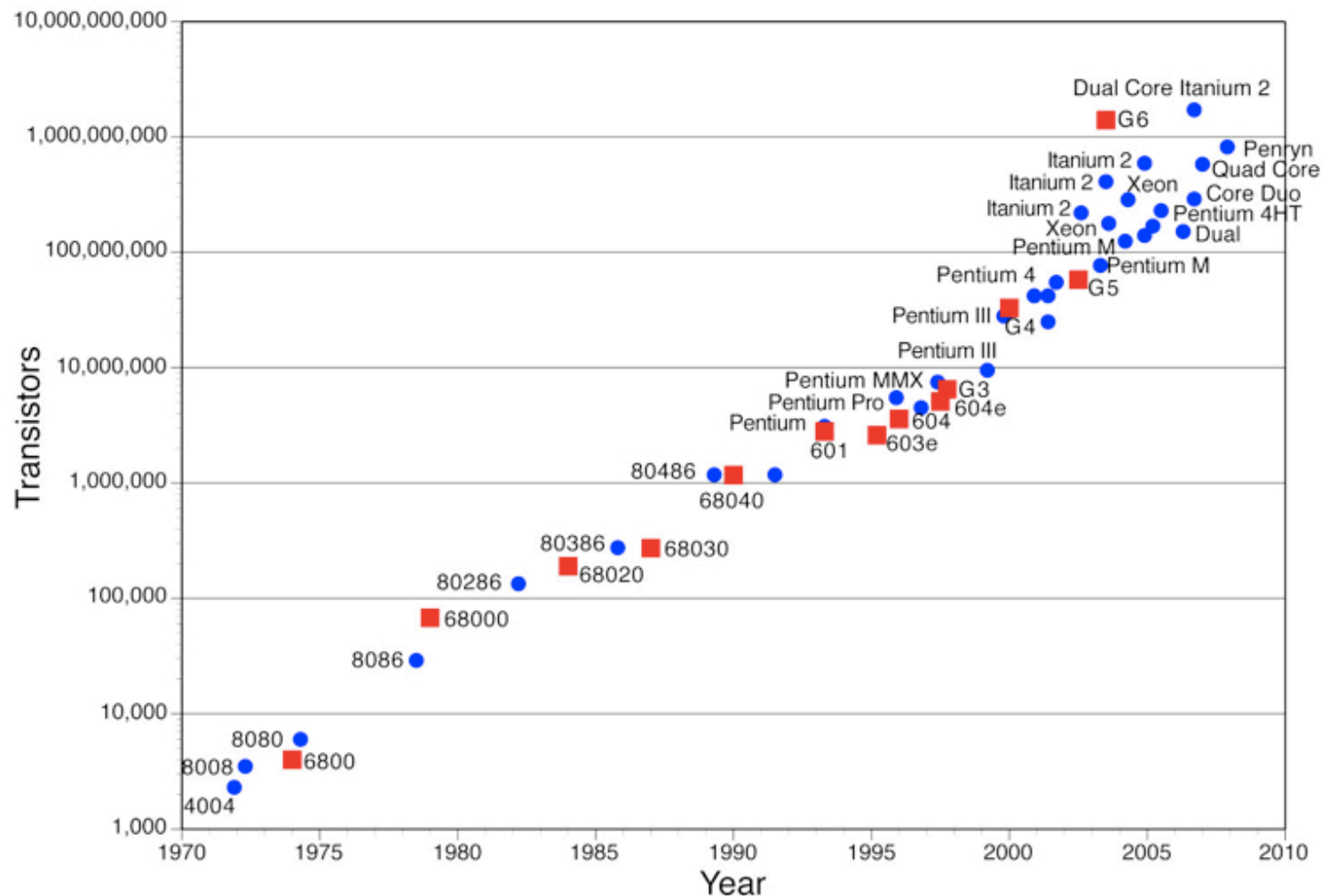


From Data to Decisions: The Power of Information in the Age of the Web of Linked Data

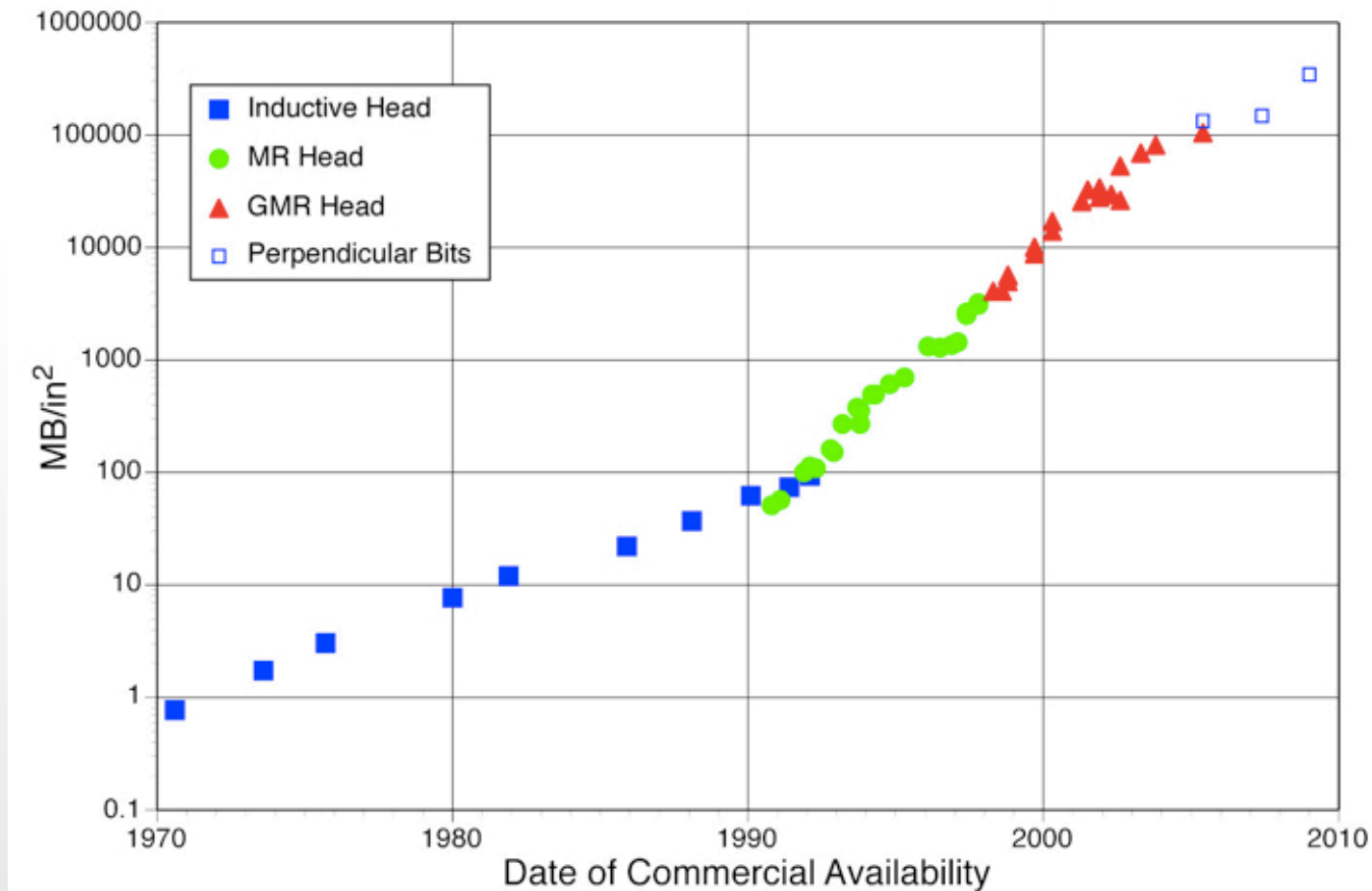
Professor Nigel Shadbolt FREng

HQS Wellington
19th October 2010

Technology: Powers of 10



Technology: Powers of 10



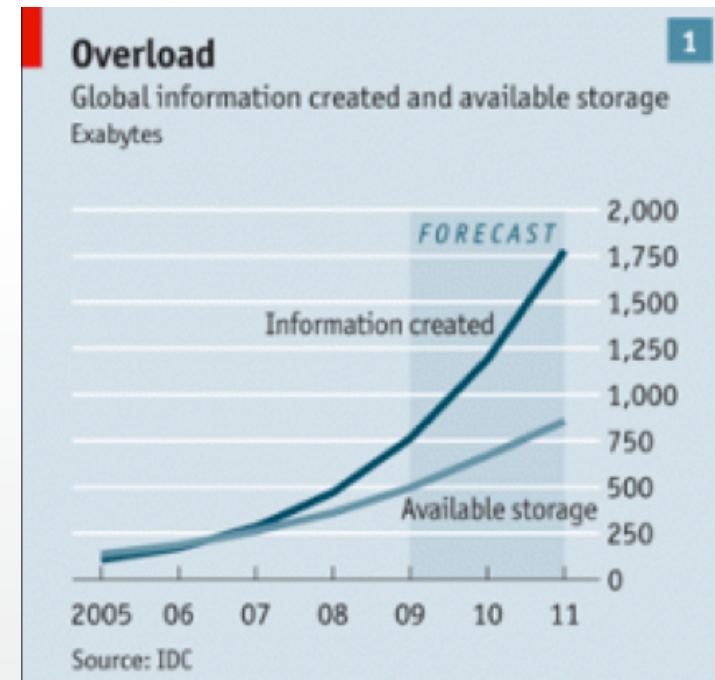
Data: Powers of 10



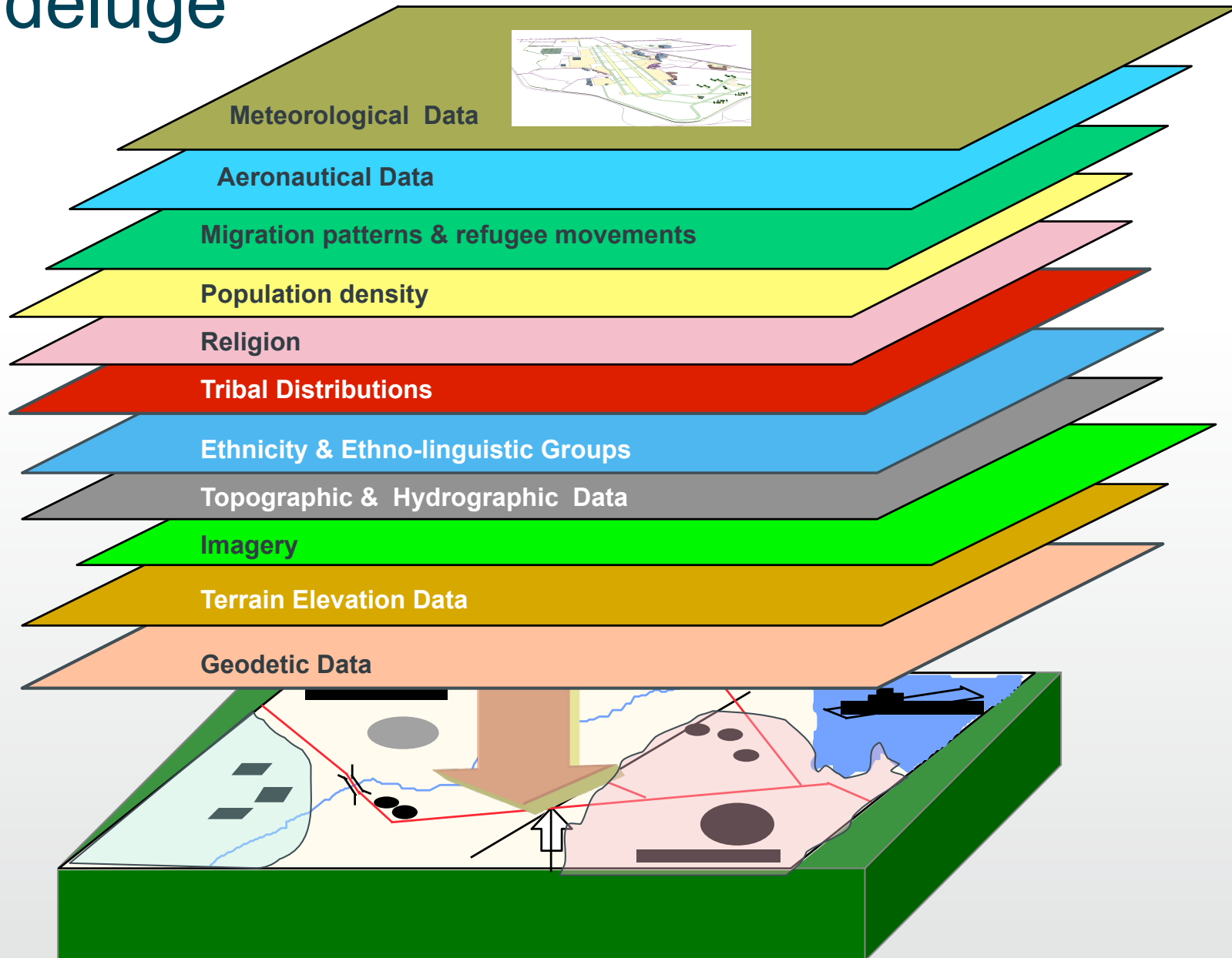
- 2007 – 281 EB (1 exabytes 1×10^{18} bytes))
- Or 281 trillion digitised novels
- 2010 – 1200 EB
- 2011 – 1.8 ZB

Data: Powers of 10

2011 amount of digital content created will be twice the amount of available storage

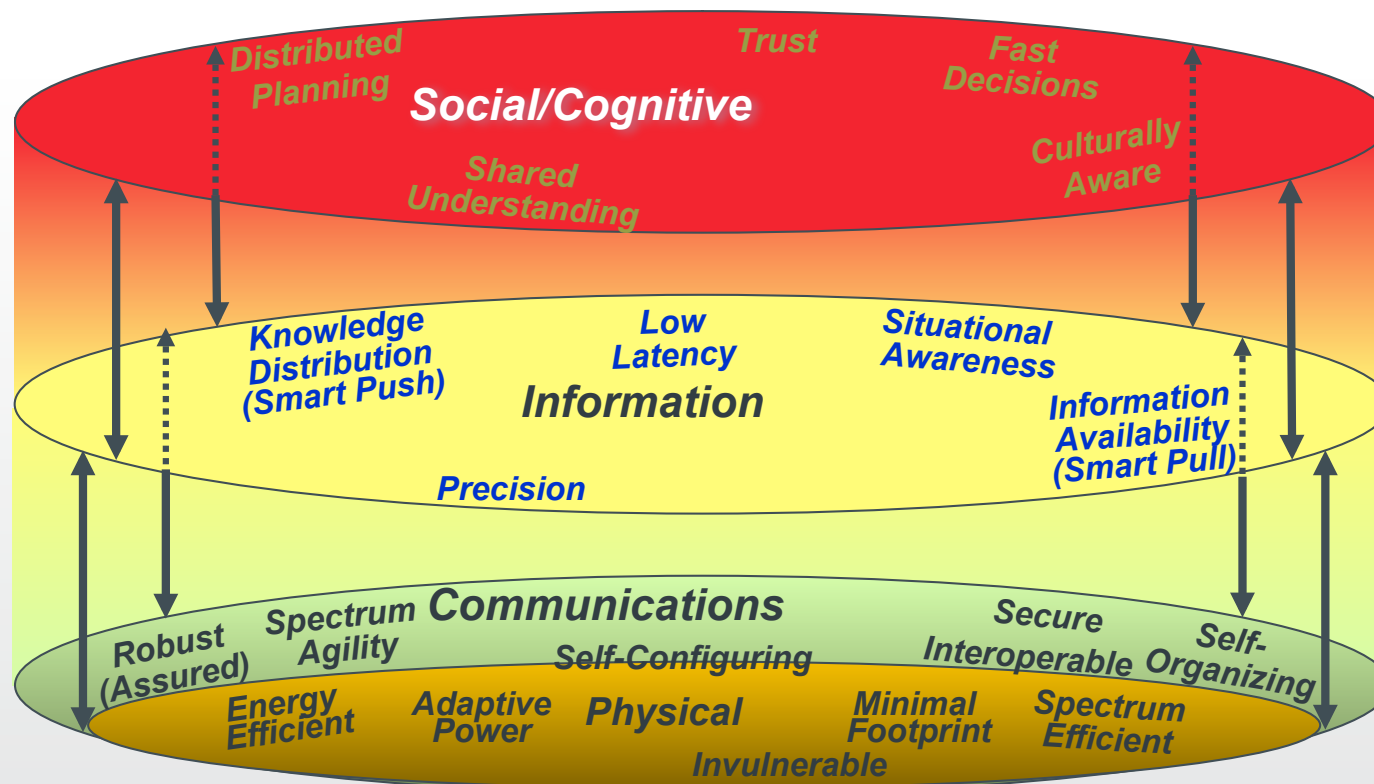


Data: A deluge



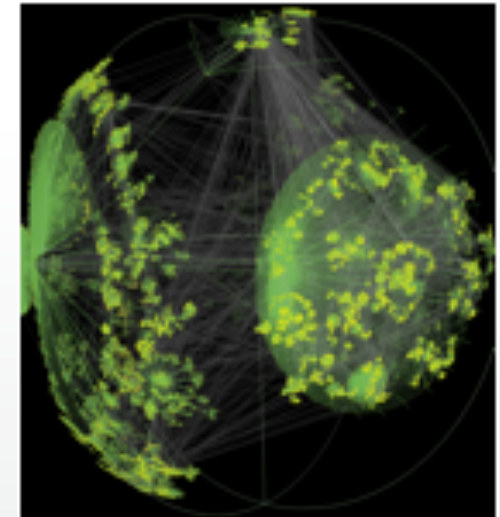
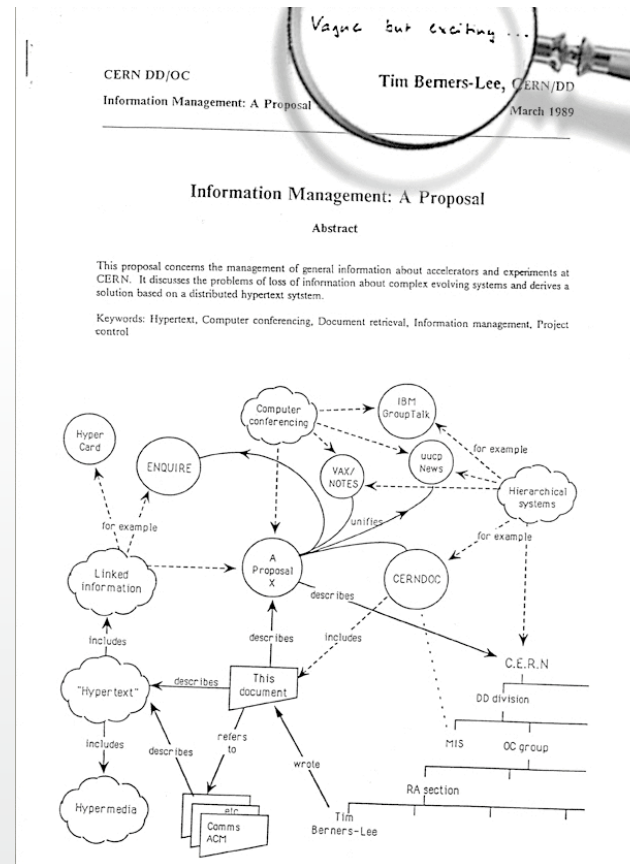
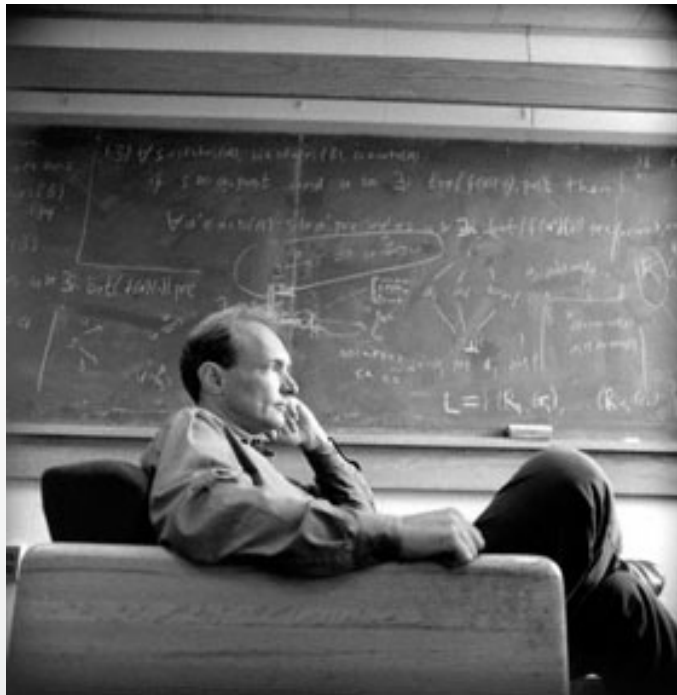
Networks: A plethora

Command and Control \rightleftharpoons Collaborate and Connect

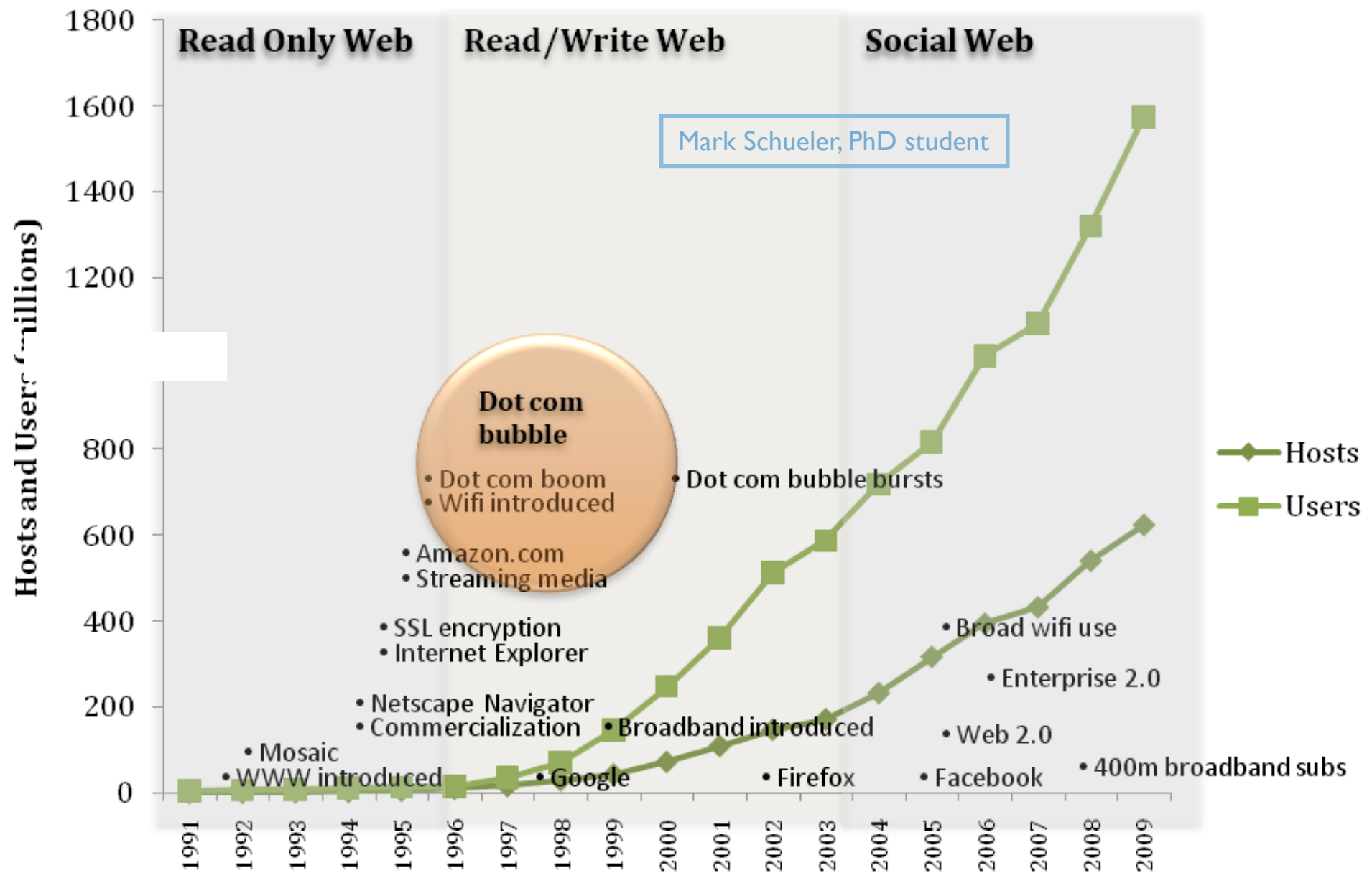


How to adapt?

The Web – most successful information architecture in history

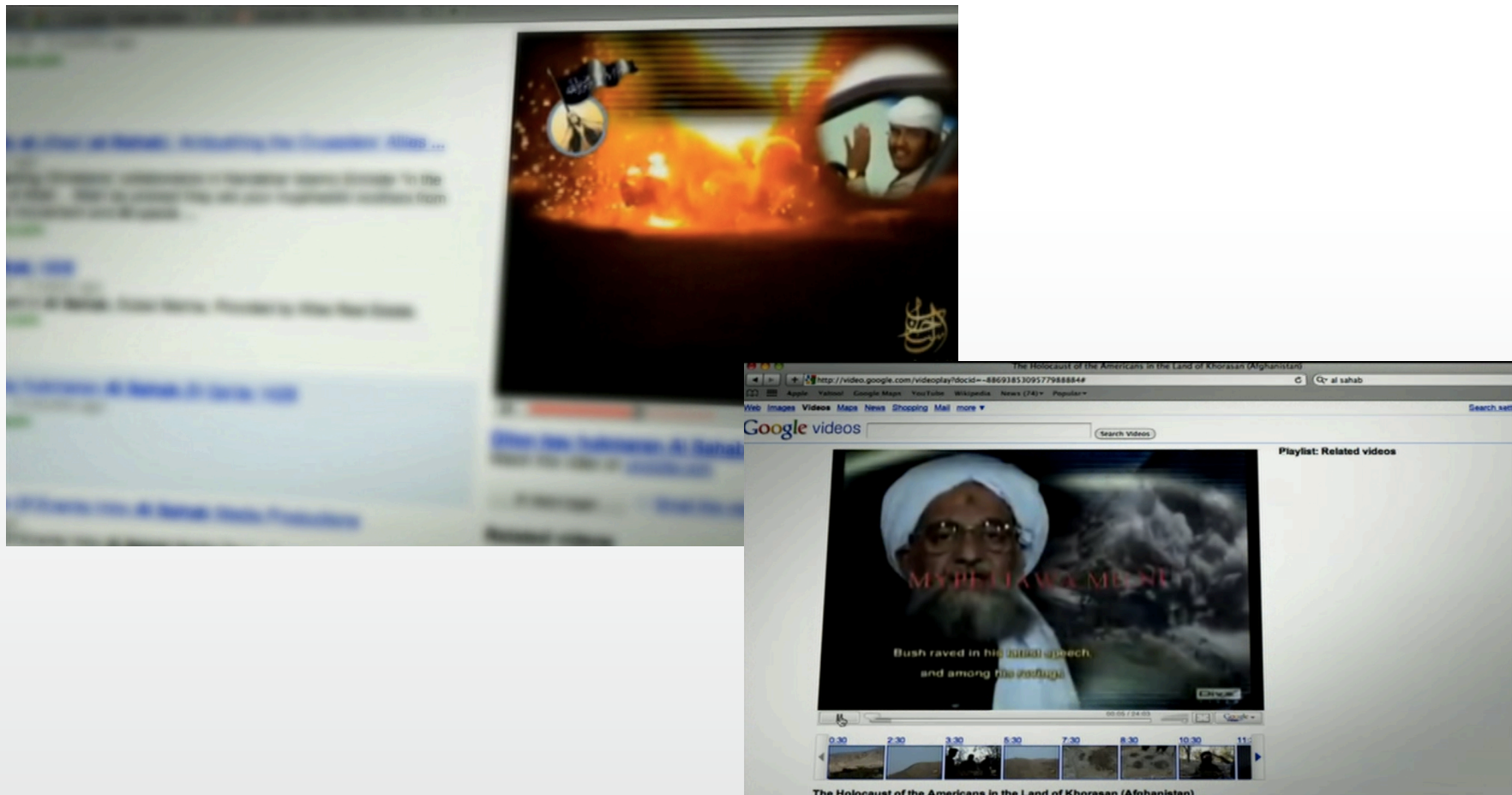


Internet Growth - Usage Phases - Tech Events



Note - events shown relate to the time axis only.

The Web in your World



The Web in your World



Collateral Murder

WikiLeaks has released a classified US military video depicting the indiscriminate slaying of over a dozen people in the Iraqi suburb of New Baghdad – including two Reuters news staff. Reuters has been trying to obtain the video through the Freedom of Information Act, without success since the time of the attack. The video, shot from an Apache helicopter gun-sight, clearly shows the unprovoked slaying of a wounded Reuters employee and his rescuers. Two young children involved in the rescue were also seriously wounded. For further information please visit the special project website www.collateralmurder.com.

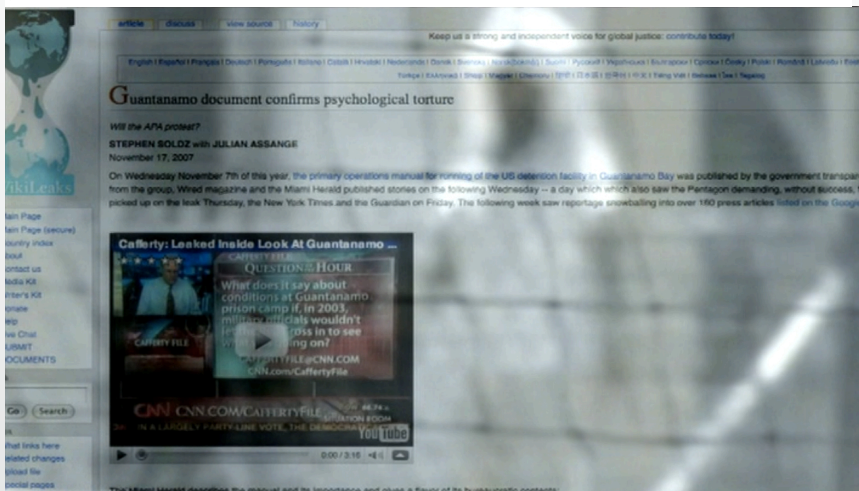
Fundraising drive

We have received hundreds of thousands of pages from corrupt banks, the US detainee system, the Iraq war, China, the UN and many others that we do not currently have the resources to release to a world audience. [You can change that](#) and by doing so, change the world. Even \$10 will pay to put one of these reports into another ten thousand hands and \$1000, a million.

We have raised just over \$370,000 for this year (our yearly budget is [around \\$800,000](#)).

The Sunshine Press (WikiLeaks) is a non-profit organization funded by human rights campaigners, investigative journalists, technologists and the general public. Through your support we have exposed significant injustice around the world – successfully fighting off over 100 legal attacks in the process. Although our work produces reforms daily and is the recipient of numerous prestigious awards, including the 2008 Index on Censorship-Economic Freedom of Expression Award as well as the 2009 Amnesty International New Media Award, these accolades do not pay the bills. Nor can we accept government or corporate funding and maintain our absolute integrity. It is [your strong support alone](#) that preserves our continued independence and strength.

We are releasing some time sensitive disclosures on this page until the moment of our re-launch.



SECRET//NOFORN



[ACIC Home](#)

(U) Wikileaks.org—An Online Reference to Foreign Intelligence Services, Insurgents, or Terrorist Groups?

NGIC-2381-0617-08

Information Cutoff Date: 28 February 2008

Publication Date: 18 March 2008

National Security Information

Unauthorized Disclosure Subject to Criminal Sanctions

Derived from: Multiple sources

Declassify on: Source documents marked 25X1

Date of source: 20060725

This Counterintelligence Analysis Report is published under the auspices of the Department of Defense Intelligence Analysis Program (DIAP).

Prepared by:

Michael D. Horvath
Cyber Counterintelligence Assessments Branch
Army Counterintelligence Center

External Coordination: National Ground Intelligence Center^[1]

This product responds to HQ, Department of Army, production requirement C764-97-0005.

ACIC Product Identification Number is RB08-0617.

[\[Back to Table of Contents\]](#)

(U) Purpose

(U) This special report assesses the counterintelligence threat posed to the US Army by the Wikileaks.org Web site.

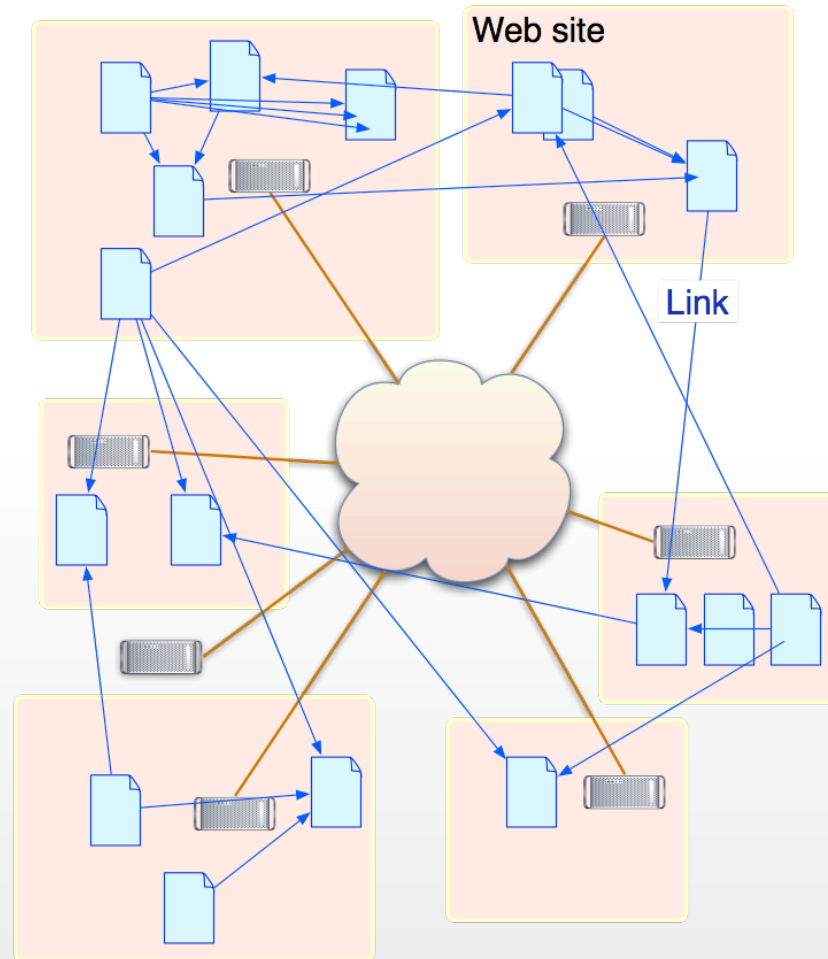
[\[Back to Table of Contents\]](#)

SECRET//NOFORN

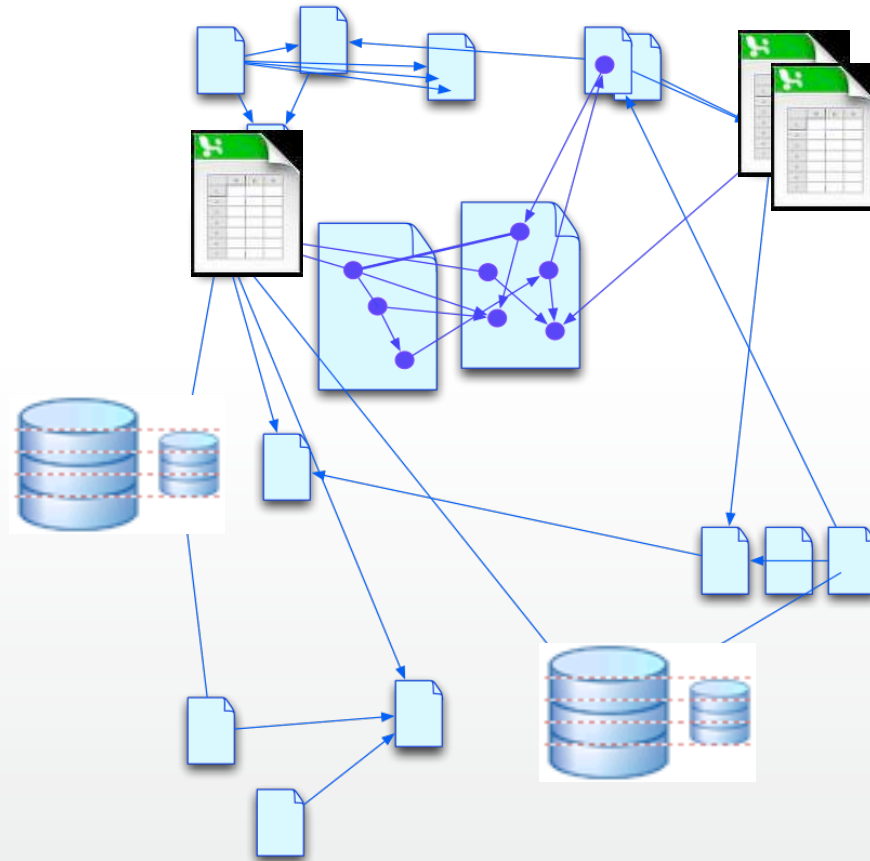
The Web evolves fast



A Web of documents...



Also a Web of data...

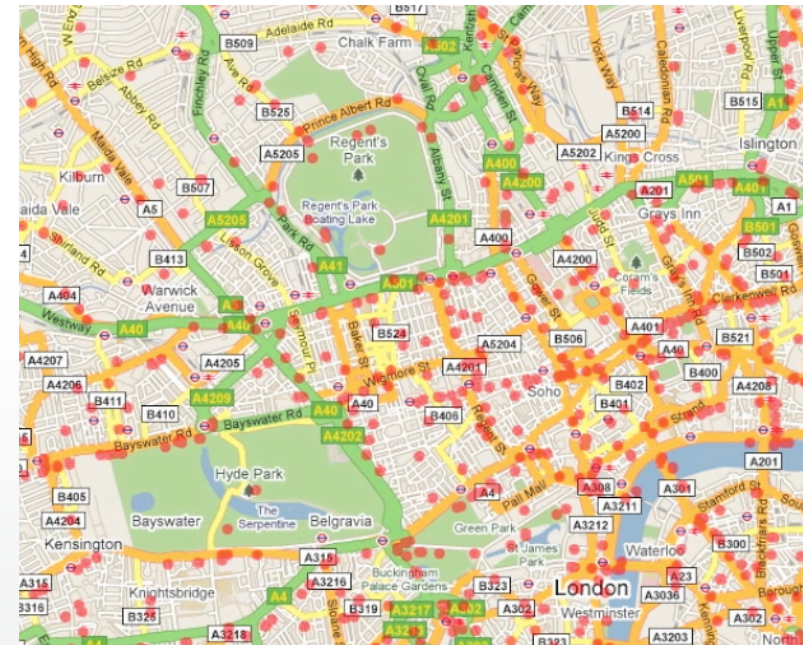


The 5 'O's

The Power of Open Data



**Cholera
cases**



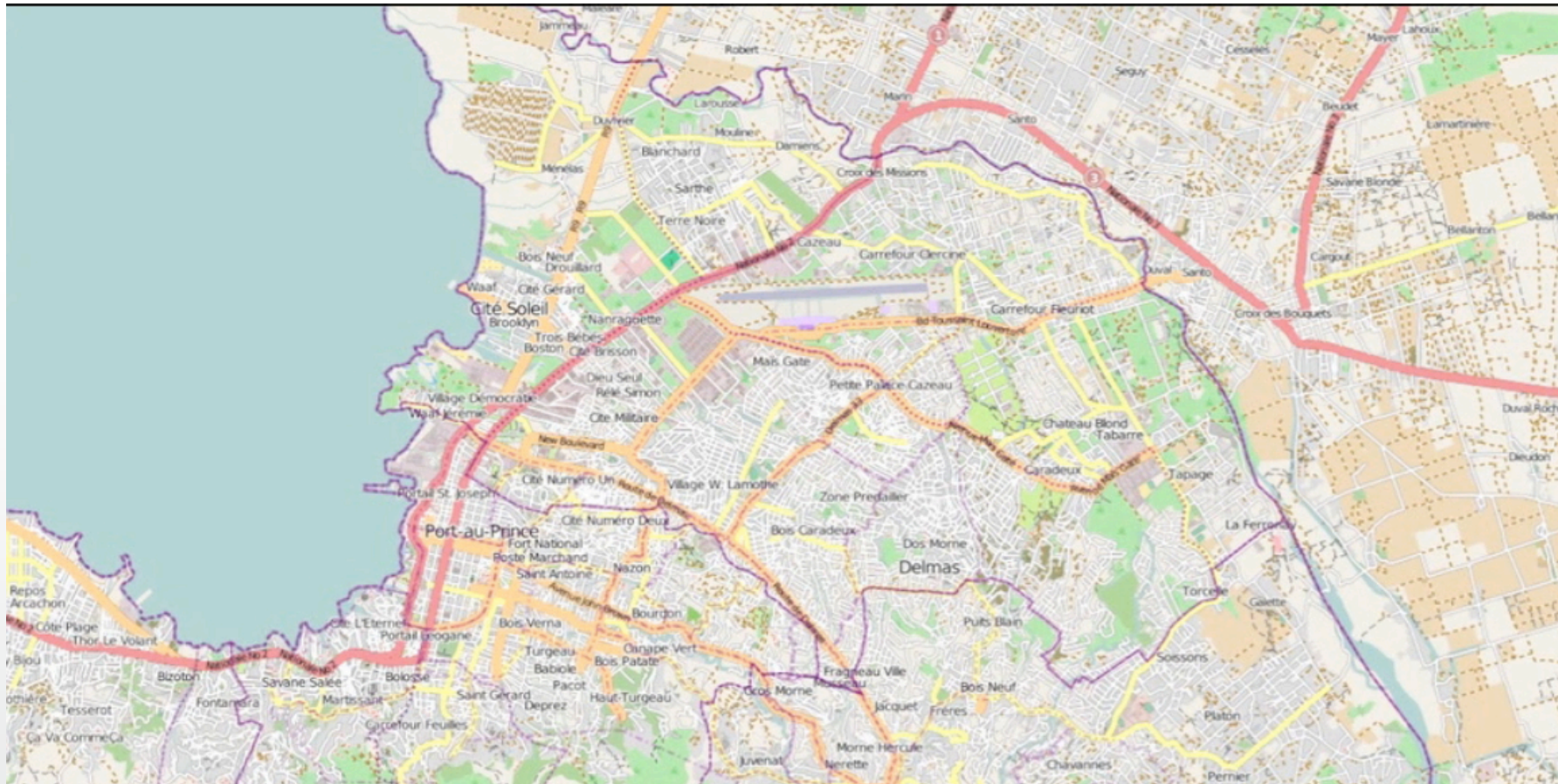
**Bicycling traffic
accidents**

The Map of Port-au-Prince 13th Jan 2010

The Power of Open Date - from this...



to this... in 13 days



How did this happen?

- Open Data
 - Open Standards
 - Open Source
 - Open Licences
 - Open Minds
-
- What lessons might you draw from this?
 - Web can be understood as extranet or intranet – http, html, etc....

Open Data

Open 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

OGD – poster child for open data

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



data.gov.uk

- 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 and the 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.

The screenshot shows the data.gov.uk website with the following content:

- Header:** HM Government logo, data.gov.uk URL, a search bar, and navigation links: Home, Data, Apps, Ideas, Forum, Wiki, Blog, Transparency, Linked Data, Resources. There are also links for Log In and Sign up.
- Main Article Title:** "New Public Sector Transparency Board and Public Data Transparency Principles". It was posted on 25/06/2010 and has 22 comments.
- Text:**
 - The Public Sector Transparency Board, which was established by the Prime Minister, met yesterday for the first time.
 - The Board will drive forward the Government's transparency agenda, making it a core part of all government business and ensuring that all Whitehall departments meet the new tight deadlines set for releasing key public datasets. In addition, it is responsible for setting open data standards across the whole public sector, listening to what the public wants and then driving through the opening up of the most needed data sets.
 - Chaired by Francis Maude, the Minister for the Cabinet Office, the other members of the Transparency Board are Sir Tim Berners-Lee, inventor of the World Wide Web, Professor Nigel Shadbolt from Southampton University, an expert on open data, Tom Steinberg, founder of mySociety, and Dr Rufus Pollock from Cambridge University, an economist who helped found the Open Knowledge Foundation.
 - In the words of Francis Maude:
"In just a few weeks this Government has published a whole range of data sets that have never been available to the public before. But we don't want this to be about a few releases, we want transparency to become an absolutely core part of every bit of government business. That is why we have asked some of the country's and the world's greatest experts in this field to help us take this work forward quickly here in central government and across the whole of the public sector."
 - At their first meeting yesterday they discussed some new Public Data Transparency Principles.
- Working definition of "Public 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.
- Draft 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.
- Featured blogs:**
 - Sir Tim Berners-Lee
 - Professor Nigel Shadbolt
 - Number 10
 - Digital Engagement
 - Digipov (COI)
 - Digital Britain
 - Jeni Tennison
- Have you used our data to create an application?**

Let us know...
Share your app
- Have you got a good idea for the data?**

We can help you find a developer...
Share your idea
- Do you want more data?**
Request new data

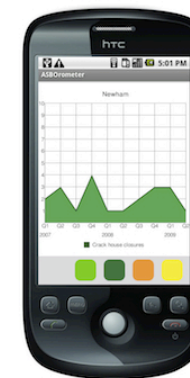
Early examples of OGD - ASBOrometer



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 local 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 Gilfelt](#) and made possible by the [data.gov.uk](#) initiative, which is opening up UK government data for public reuse.



OGD – location, location, location...

www.ordnancesurvey.co.uk/opendata

home view develop supply help

OS OpenData

Mapping data and geographic information from Ordnance Survey

Create and support innovative, exciting ideas and applications using Ordnance Survey mapping. With OS OpenData you can access a selection of the most detailed mapping datasets available for Great Britain.

Now available - OS VectorMap™ District - a new mid-scale vector and raster dataset (alpha version).

view

- Outline of Great Britain
- Overview of Great Britain
- MiniScale ©
- 1:250 000 Scale Colour Raster
- OS Street View ©
- Boundary-Line™
- OS VectorMap™ District - **New**

develop

- Create amazing web applications
- Enhance online projects
- Put Ordnance Survey maps on your website

Our API uses JavaScript to make web pages more interactive, anyone can create an application by following simple code and tutorials.

supply

Order OS OpenData on hard media or download direct to your PC, Mac or laptop (please note - **you may need specialist GI data software to view and use this data...**)

about

[View, download and order](#) OS OpenData or use [OS OpenSpace](#) to enhance your own application.

You can also experiment with [OS OpenData in RDF format](#).

If OS OpenData doesn't meet your needs, why not try:

- The Ordnance Survey [Map Shop](#) for paper maps.
- The [Explore portal](#) for routes, photos and blogs.
- Ordnance Survey mapping for [business](#) or [education](#).

help

See our [FAQ](#) page to find answers to any questions you might have about OS OpenData, get more [technical help](#) with the map viewer or downloads.

Or you can [email](#) or call us on **0845 408 1895** (8.30am - 5.30pm, Monday to Friday).

- [OS OpenData Licence](#)
- [Privacy policy](#)
- [Website terms and conditions](#)

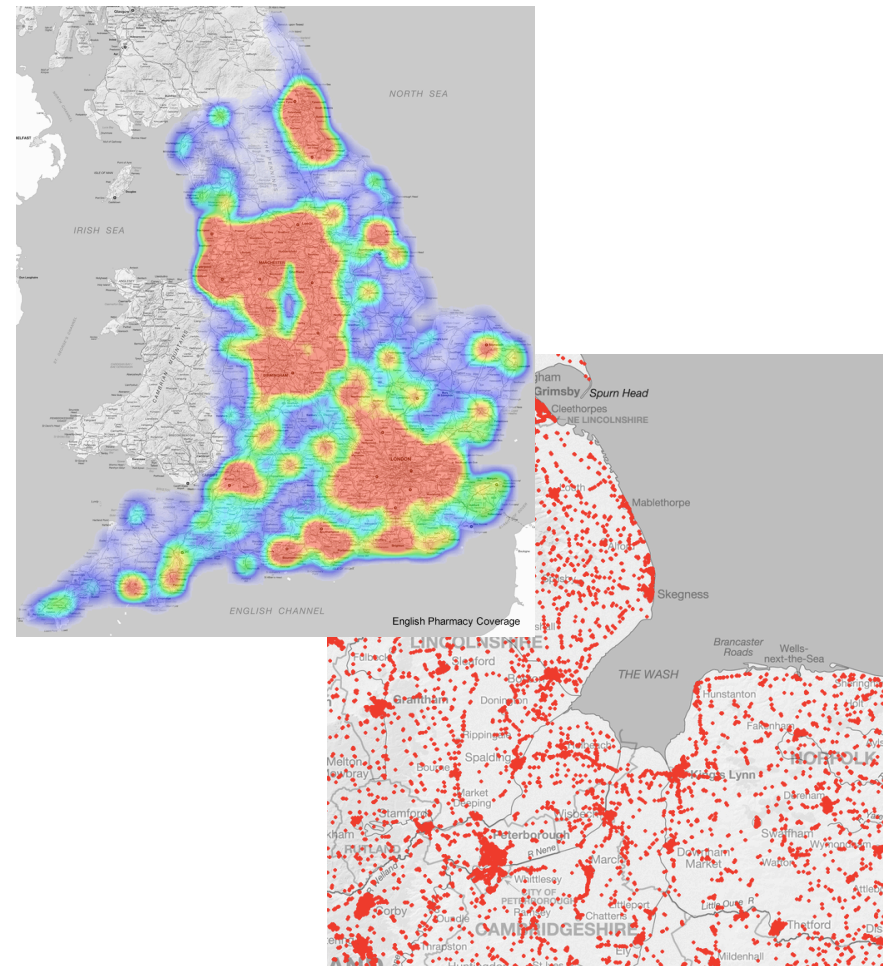
keep in touch

Follow us on Twitter, Facebook or YouTube.

[YouTube](#) [Twitter](#) [Facebook](#) [Blog](#)

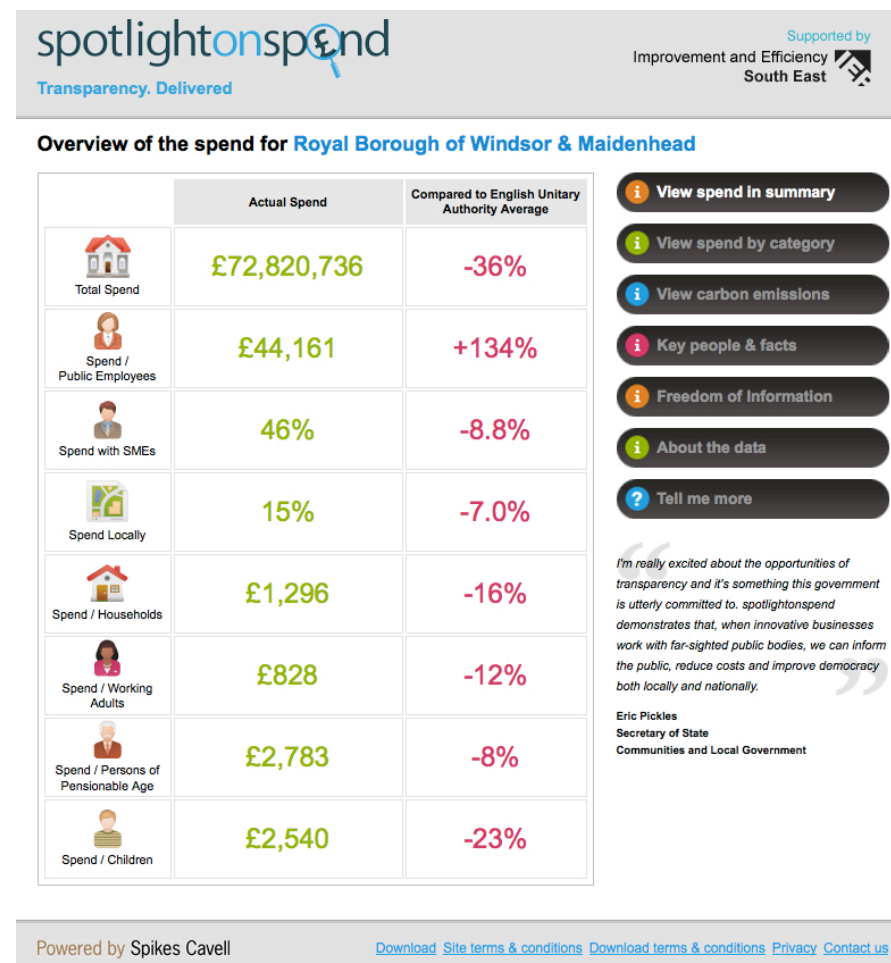
Featured blogs:

[Ordnance Survey Blog](#), [Sir Tim Berners-Lee](#), [Professor Nigel Shadbolt](#), [Digital Engagement](#)



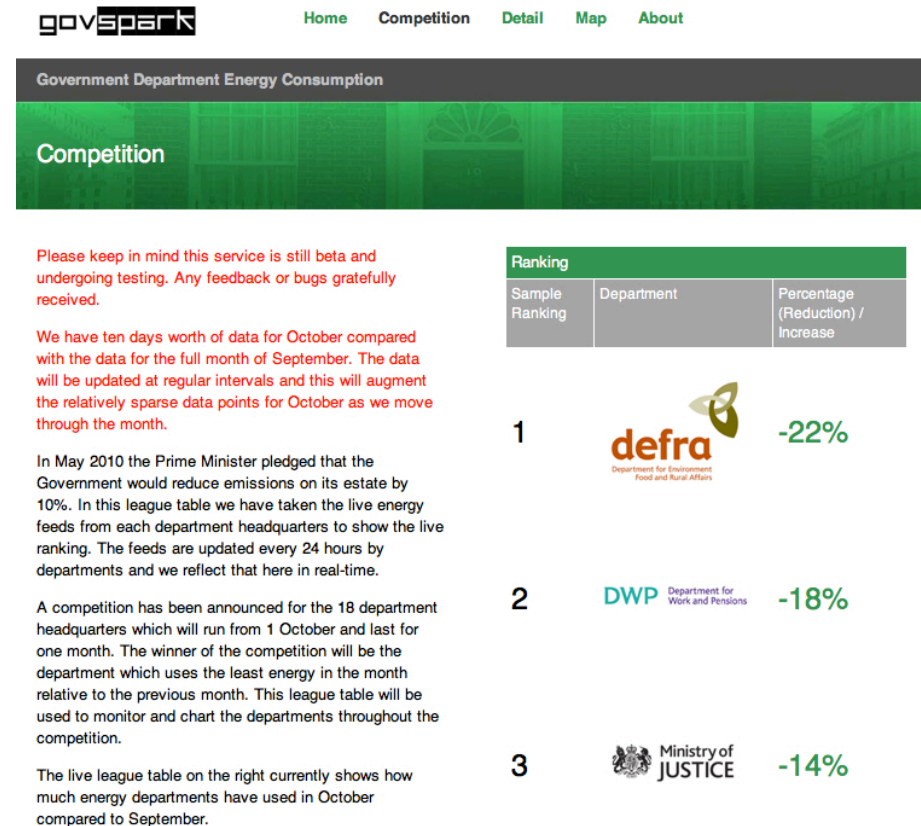
Data to decisions - spending...

New items of Local government spending over £500 – council by council from Jan 2011



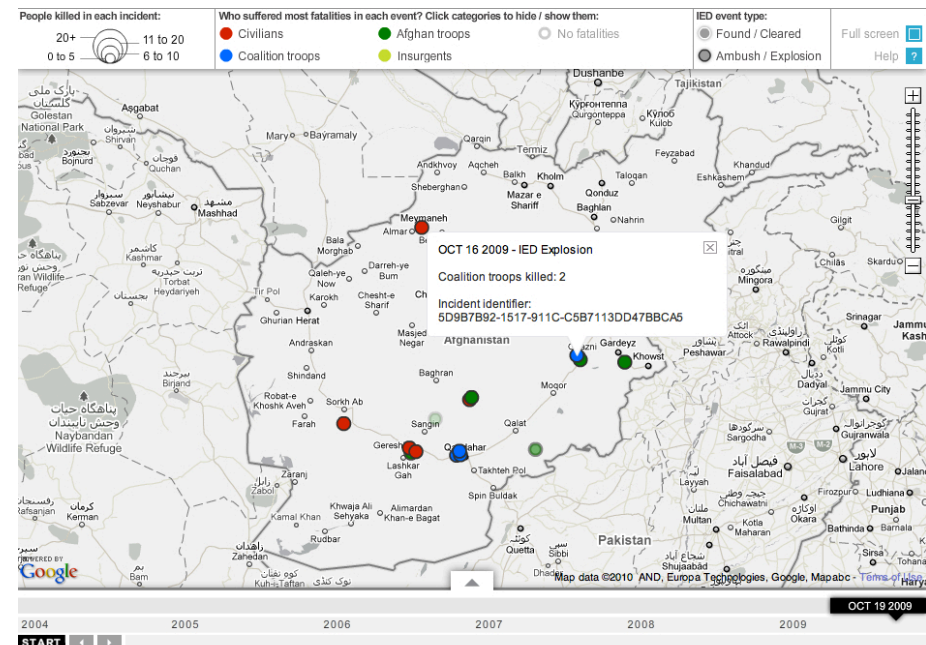
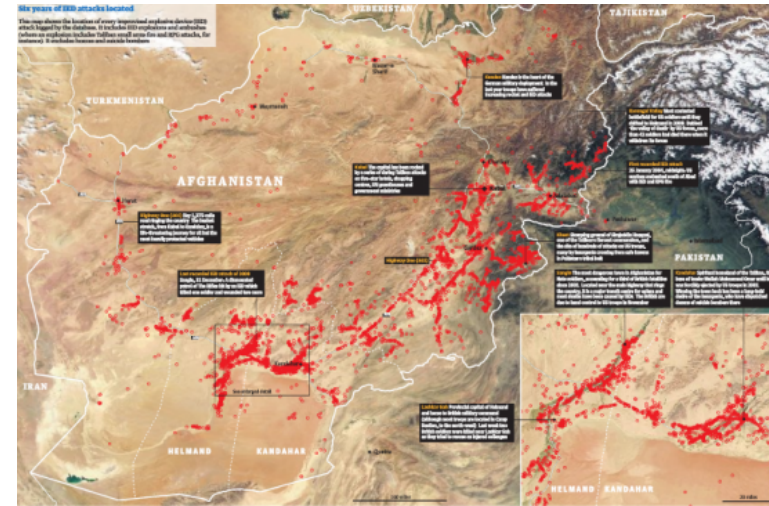
Data to decisions - behaviour...

Real time energy data is changing behaviours – the decisions people make



Data to decisions - behaviour...

IEDS from Afghanistan War
Logs – raw data available
from Guardian




Open Standards

Why 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

Ordinance Survey - Great Britain's national mapping agency See all our sites [Login](#)

 **Ordinance Survey**

[home](#) [business](#) [leisure](#) [map shop](#) [explore](#) [magazine](#) [education](#) [about us](#) [data](#)

Burghclere, Highclere and St. Mary Bourne

A description of the resource identified by <http://data.ordnancesurvey.co.uk/id/7000000000017789>

Label	Burghclere, Highclere and St. Mary Bourne
Name	Burghclere, Highclere and St. Mary Bourne
Type	District Ward
Area	11445.724
Area code	DIW
Census code	24UBKE
Unit i d	17789
Borders	Harewood Bourne Valley Kingsclere Overton, Laverstoke and Steventon Alamein Greenham St. Mary's Whitchurch Falkland East Woodhay Kintbury
Contains	Litchfield and Woodcott St. Mary Bourne Ecclesdown, Sydmonton and Bishops Green Burghclere Highclere Ashmansworth Newtown

The data for this description was obtained from the SPARQL service at <http://api.talis.com/stores/ordnance-survey/services/sparql>. This data is also available as: [RDF/XML](#), [Turtle](#) and [JSON](#). A free text search service is available at <http://api.talis.com/stores/ordnance-survey/items>

View this data using these other browsers:

- [Marbles](#)
- [Zitgist Data Viewer](#)
- [Disco](#)
- [Dipper](#)
- [OpenLink Data Explorer](#)

[Top of page](#)

Ordinance Survey is Great Britain's national mapping agency, providing the most accurate and up-to-date geographic data, relied on by government, business and individuals.

Ordinance Survey © Crown copyright 2009
[Jobs](#) | [Help](#) | [Accessibility](#) | [Contact Us](#) | [Sitemap](#) | [Legal & Privacy](#)

Why 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

Litchfield and Woodcott	label	Litchfield and Woodcott
Burghclere, Highclere and St. Mary Bourne	Area	11445.724
	Area Code	DIW
	Census Code	24UBKE
	Unit ID	17789
	borders	Greenham
		Falkland
		Kintbury
		Overton, Laverstoke and Steventon
		Harewood
		St. Mary's
		Whitchurch
		Kingsclere
		East Woodhay
		Bourne Valley
		Alamein
	contains	St. Mary Bourne
		Ecchinswell, Sydmonton and Bishops Green
		Newtown
		Burghclere
		Highclere
		Litchfield and Woodcott
		Ashmansworth
		District Ward
		Burghclere, Highclere and St. Mary Bourne
		Burghclere, Highclere and St. Mary Bourne
		Highclere
		Harewood
		Alamein
		Ecchinswell, Sydmonton and Bishops Green
		Burghclere
		Greenham
		St. Mary Bourne
		Newtown
		Bourne Valley
		District Ward
		East Woodhay
		Falkland
		Kingsclere
		Kintbury
		St. Mary's
		Whitchurch
		Ashmansworth
		Overton, Laverstoke and Steventon
http://data.ordnancesurvey.co.uk/doc/7000000000017789	Has Format	http://data.ordnancesurvey.co.uk/doc/7000000000017789.html
		http://data.ordnancesurvey.co.uk/doc/7000000000017789.json
		http://data.ordnancesurvey.co.uk/doc/7000000000017789.rdf
		http://data.ordnancesurvey.co.uk/doc/7000000000017789.ttl
		Linked Data for Burghclere, Highclere and St. Mary Bourne
		http://purl.org/dc/dcmitype/Text
		Document
		Burghclere, Highclere and St. Mary Bourne
http://data.ordnancesurvey.co.uk/doc/7000000000017789.rdf	primary topic	application/rdf+xml
	Format	http://data.ordnancesurvey.co.uk/doc/7000000000017789
	Is Format Of	Linked Data in RDF/XML format for Burghclere, Highclere and St. Mary Bourne
	Title	http://purl.org/dc/dcmitype/Text
	type	Document
http://data.ordnancesurvey.co.uk/doc/7000000000017789.html	primary topic	Burghclere, Highclere and St. Mary Bourne
	Format	text/html

RKBExplorer: The Power of Linked Data

RESIST RESILIENCE FOR SURVIVABILITY IN IST
KNOWLEDGE BASE EXPLORER
VERSION 1.0.2

People Research Areas Publications Projects Search Recently Viewed Reset Help

N. R. Shadbolt Restore

11 / 10

Detail
Name: N. R. Shadbolt
Other searches: [Google Scholar](#)

People
G. Rugg
A. M. Burton
T. Elliott
R. K. Stobart

Research Areas
I.2.6. Learning
I.2.4. Knowledge Representation Formalisms and Methods
J.7. COMPUTERS IN

Publications
Empowering Resource Providers to Build the Semantic Grid
Multiplicative synaptic normalization and a nonlinear Hebb rule underlie a neurotrophic

Projects
No projects found
[About](#) [Acknowledgements](#)

Find: Cult [Next](#) [Previous](#) [Highlight all](#) [Match case](#) [Reached end of page, continued from top](#)

Linked Data - Technical Principles

The four micro principles of the Semantic Web

1. All entities of interest, such as information resources, real-world objects, and vocabulary terms should be identified by URI references.
2. URI references should be dereferenceable, meaning that an application can look up a URI over the HTTP protocol and retrieve RDF data about the identified resource.
3. Data should be provided using the RDF/XML syntax
4. Data should be interlinked with other data.

The Semantic Web Revisited

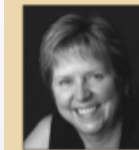
Nigel Shadbolt and Wendy Hall, *University of Southampton*
Tim Berners-Lee, *Massachusetts Institute of Technology*



Nigel Shadbolt is a professor of artificial intelligence in the School of Electronics and Computer Science at Southampton University. Contact him at nrs@ecs.soton.ac.uk.



Tim Berners-Lee is the director of the World Wide Web Consortium, a senior researcher at the Massachusetts Institute of Technology's Computer Science and Artificial Intelligence Laboratory, and a professor of computer science in the Department of Electronics and Computer Science at Southampton University. Contact him at timbl@w3.org.



Wendy Hall is a professor of computer science in the School of Electronics and Computer Science at Southampton University. Contact her at wh@ecs.soton.ac.uk.

Principle 1: URIs

- Uniform Resource identifier
- A Web based identifier
- You can dereference them – get something back
- Information about objects, relations, properties and values

<http://rdf.ecs.soton.ac.uk/person/2686>

<http://rdf.ecs.soton.ac.uk/project/464>

<http://rdf.ecs.soton.ac.uk/publication/11065>

<http://education.data.gov.uk/doc/school/120805>

[http://southampton.rkbexplorer.com/id/
person-02686](http://southampton.rkbexplorer.com/id/person-02686)

http://dbpedia.org/resource/Nigel_Shadbolt

Principle 2: Dereferencable

- Uniform Resource identifier
- A Web based identifier
- You can dereference them – get something back
- Information about objects, relations, properties and values

▼ Southampton ECS People: Professor Nigel R Shadbolt		
Southampton ECS People: Professor Nigel R Shadbolt	Description	This rdf document contains information about a person in the Department of Southampton.
Nigel Shadbolt	Source	http://www.ecs.soton.ac.uk/people/nrs
	Title	Southampton ECS People: Professor Nigel R Shadbolt
	Date Created	2009-11-26T10:53:20Z
	type	Ontology
	label	Southampton ECS People: Professor Nigel R Shadbolt
	Appellation	Professor
	Family Name	Shadbolt
	Full Name	Professor Nigel R Shadbolt
	Given Name	Nigel R
	Role	http://id.ecs.soton.ac.uk/role/2686
	member Of	http://id.ecs.soton.ac.uk/project/221
		http://id.ecs.soton.ac.uk/project/293
		http://id.ecs.soton.ac.uk/project/312
		http://id.ecs.soton.ac.uk/project/353
		http://id.ecs.soton.ac.uk/project/361
		http://id.ecs.soton.ac.uk/project/381
		http://id.ecs.soton.ac.uk/project/395
		http://id.ecs.soton.ac.uk/project/443
		http://id.ecs.soton.ac.uk/project/44
		http://id.ecs.soton.ac.uk/project/45
		http://id.ecs.soton.ac.uk/project/463
		http://id.ecs.soton.ac.uk/project/464
		http://id.ecs.soton.ac.uk/project/465
		http://id.ecs.soton.ac.uk/project/466
		http://id.ecs.soton.ac.uk/project/467
		http://id.ecs.soton.ac.uk/project/508
		http://id.ecs.soton.ac.uk/project/623
		http://id.ecs.soton.ac.uk/project/629
		http://id.ecs.soton.ac.uk/project/630
	type	Person
	family name	Shadbolt
	givenname	Nigel R
	homepage	http://users.ecs.soton.ac.uk/nrs/
	img	http://www.ecs.soton.ac.uk/image.php?id=person_2686&checksum=6c2158b33a0c10e328ec62f9ed1cc732&maxw=250
	mbox	mailto:nrs@ecs.soton.ac.uk
	name	Professor Nigel R Shadbolt

Principle 3: RDF back

- Uniform Resource identifier
- A Web based identifier
- You can dereference them – get something back
- Web Knowledge Representation called RDF
- Information about objects, relations, properties and values

▼ Southampton ECS Projects: MIMEX: Multivariant Information Management and Exploitation		
Southampton ECS Projects: MIMEX: Multivariant Information Management and Exploitation	Description	This rdf document contains information about a project in the Department of Electronics and Computer Science, University of Southampton.
	Source	http://www.ecs.soton.ac.uk/research/projects/443
	Title	Southampton ECS Projects: MIMEX: Multivariant Information Management and Exploitation
	Date Created	2009-11-26T10:58:31Z
	type	Ontology
	label	Southampton ECS Projects: MIMEX: Multivariant Information Management and Exploitation
http://id.ecs.soton.ac.uk/project/443	associated With	http://id.ecs.soton.ac.uk/UoS
	Name	General Dynamics UK Limited
	type	Organization
	Name	University of Cardiff
	type	Organization
	Name	Ministry of Defence
	type	Organization
	associated With Unit	http://id.ecs.soton.ac.uk/UoS/ECS
	Name	MIMEX: Multivariant Information Management and Exploitation
	Theme	http://id.ecs.soton.ac.uk/theme/106
		http://id.ecs.soton.ac.uk/theme/127
		http://id.ecs.soton.ac.uk/theme/130
		http://id.ecs.soton.ac.uk/theme/38
		http://id.ecs.soton.ac.uk/theme/66
	Web Page	https://www.edefence.org/mimex
	run By Group	http://id.ecs.soton.ac.uk/group/iam
	begins	in Calendar Clock
		day
		hour
		minute
		month
		year
		type
		in Calendar Clock Data Type
		2006-07-03T00:00:00Z
		Instant
	ends	in Calendar Clock
		day
		hour
		minute
		month
		year
		type
		in Calendar Clock Data Type
		2009-03-31T00:00:00Z
		Instant

Principle 4: Link data

<http://rdf.ecs.soton.ac.uk/person/2686> SameAs

<http://southampton.rkbexplorer.com/id/person-02686> SameAs

http://dbpedia.org/resource/Nigel_Shadbolt

The screenshot shows the SIG.MA Semantic Informa Mashup interface. The main content area displays a profile for Nigel R Shadbolt, including his given name, family name, and a list of works he is the creator of. The works list includes titles like "AKT Selected Papers 2004", "INSTRUCTA-Phase II", "Techniques of Temporal Representation", and "Expert Systems - A Natural History". A sidebar on the right lists 15 sources, each with a date and a link to the source. The bottom of the page has a search bar and navigation links.

The screenshot shows the SIG.MA Semantic Informa Mashup interface. The main content area displays a profile for Nigel Shadbolt, including his identifier, known for, label, and is means of. The is means of section lists various categories like "British computer scientists", "Living people", "People from London", "Fellows of the Royal Academy of Engineering", and "Fellows of the British Computer Society". A sidebar on the right lists 15 sources, each with a date and a link to the source. The bottom of the page has a search bar and navigation links.

Wikipedia as Data: DBpedia

- 2.9 million things
- > 282,000 persons
- 339,000 places
- 88,000 music albums
- 44,000 films
- 15,000 video games
- 119,000 organizations
- 130,000 species
- 4400 diseases.



About / News
Applications
Use Cases
Datasets
Online Access
Downloads
ChangeLog
Interlinking
Framework
Publications
Community
Blog
Credits
Contact / Imprint

DBpedia is a community effort to extract structured information from Wikipedia and to make this information available on the Web. DBpedia allows you to ask sophisticated queries against Wikipedia, and to link other data sets on the Web to Wikipedia data. We hope this will make it easier for the amazing amount of information in Wikipedia to be used in new and interesting ways, and that it might inspire new mechanisms for navigating, linking and improving the encyclopaedia itself.

News

[Open Knowledge Conference 2010](#)

OKCon, now in its fifth year, is the interdisciplinary conference that brings together individuals from across the open knowledge spectrum (such as also DBpedia in particular and Linked Open Data in general) for a day of presentations and workshops. Open knowledge promises significant social and economic benefits in a wide range of areas from governance [...]

[DBpedia in ReadWriteWeb's Top 10 Semantic Web Products of 2009](#)

The new year is slowly approaching and people start compiling their top x lists of 2009, with x usually ranging between 10 and 365. The popular Web technology blog ReadWriteWeb has chosen x with value 10 and picked DBpedia as one of their top Semantic Web products of 2009. Its actually the only non-commercial [...]

[German government proclaims Faceted Wikipedia/DBpedia Search one of the 365 most innovative ideas in Germany](#)

The German federal government has proclaimed Faceted Wikipedia Search as one of the 365 most innovative ideas in Germany in the context of the Deutschland 7 Land der Ideen competition. The competition showcases innovative ideas in areas such as science and technology, business, education, art and ecology. The patron of the competition is the German President Horst Köhler. Faceted [...]

The DBpedia Knowledge Base

Knowledge bases are playing an increasingly important role in enhancing the intelligence of Web and enterprise search and in supporting information integration. Today, most knowledge bases cover only specific domains, are created by relatively small groups of knowledge engineers, and are very cost intensive to keep up-to-date as domains change. At the same time, Wikipedia has grown into one of the central knowledge sources of mankind, maintained by thousands of contributors. The DBpedia project leverages this gigantic source of knowledge by extracting structured information from Wikipedia and by making this information accessible on the Web under the terms of the [Creative Commons Attribution-ShareAlike 3.0 License](#) and the [GNU Free Documentation License](#).

SPARQL

- A data access language for RDF on the Web
- Queries across diverse data sources
- SPARQL contains capabilities for querying required and optional patterns

The screenshot shows the SPARQL Explorer interface for dbpedia.org. The query is a SELECT statement filtering for people born before 1800-01-01. The results table lists 12 individuals with their names, birth dates, death dates, and dbpedia URIs.

SPARQL Explorer for <http://dbpedia.org/sparql>

SPARQL:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfa: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX : <http://dbpedia.org/resource/>
PREFIX dbpedia2: <http://dbpedia.org/property/>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dbo: <http://dbpedia.org/ontology/>

SELECT ?name ?birth ?death ?person WHERE {
  ?person dbpedia2:birthPlace <http://dbpedia.org/resource/London> .
  ?person dbo:birthDate ?birth .
  ?person foaf:name ?name .
  ?person dbo:deathDate ?death
  FILTER (?birth < "1800-01-01"^^xsd:date) .
}

ORDER BY ?name
```

Results:

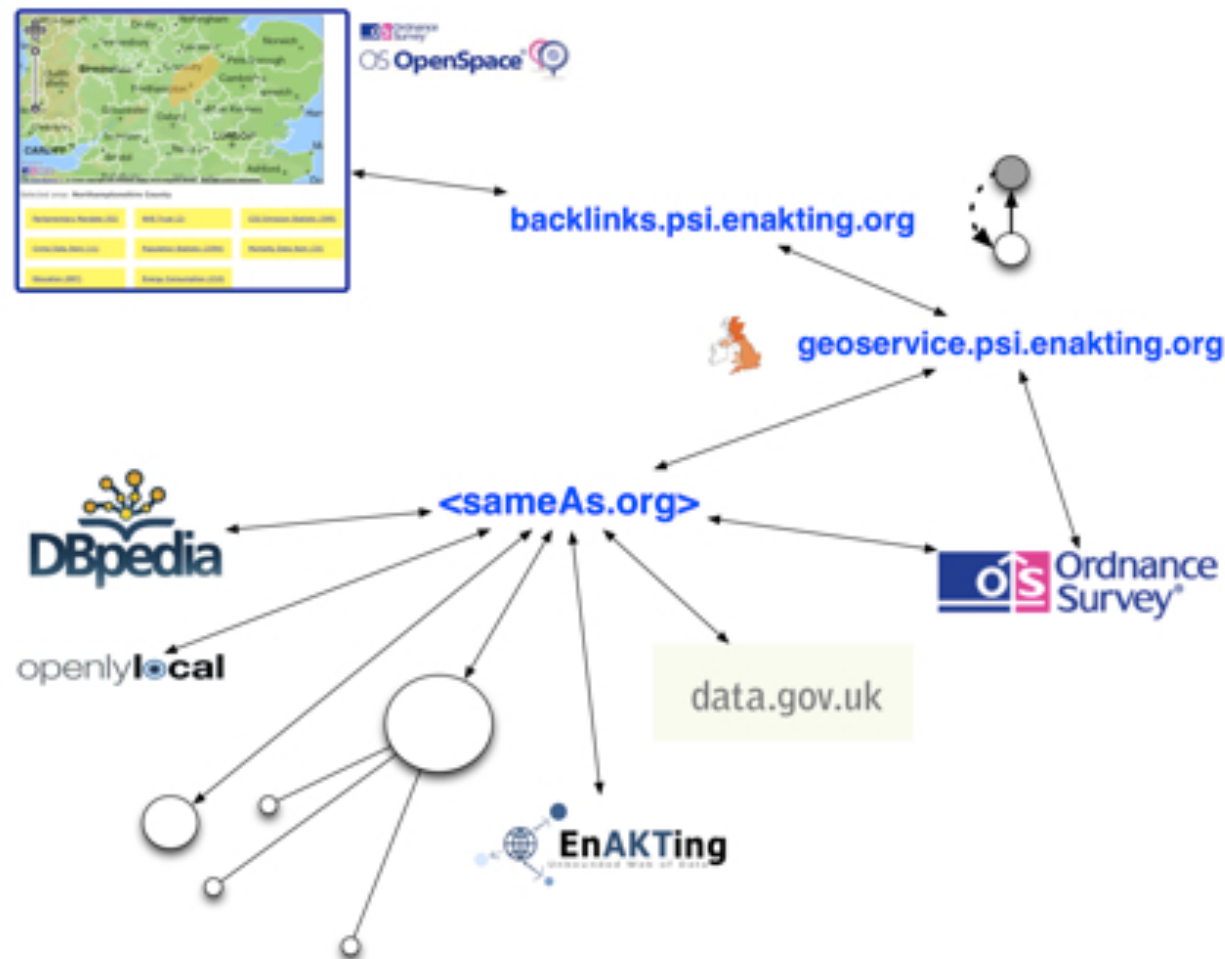
SPARQL results:

name	birth	death	person
"<small>Countess of Chatham"	"1720-11-08"^^xsd:date	"1803-04-09"^^xsd:date	:Hester_Pitt%2C_Countess_of_Chatham ↗
"Aaron Hart"	"1724-08-16"^^xsd:date	"1800-12-28"^^xsd:date	:Aaron_Hart_%28businessman%29 ↗
"Alexander Pope"	"1688-05-21"^^xsd:date	"1744-05-30"^^xsd:date	:Alexander_Pope ↗
"Alexander Schomberg"	"1720-01-01T00:00:00-05:00"^^xsd:gYear	"1804-01-01T00:00:00-05:00"^^xsd:gYear	:Alexander_Schomberg ↗
"Anne Killigrew"	"1660-01-01T00:00:00-05:00"^^xsd:gYear	"1685-06-16"^^xsd:date	:Anne_Killigrew ↗
"Aubrey George Spencer"	"1795-02-12"^^xsd:date	"1872-02-24"^^xsd:date	:Aubrey_George_Spencer ↗
"Ben Jonson"	"1572-06-11"^^xsd:date	"1637-08-06"^^xsd:date	:Ben_Jonson ↗
"Benjamin Aislable"	"1774-01-01T00:00:00-05:00"^^xsd:gYear	"2"	:Benjamin_Aislable ↗
"Benjamin Aislable"	"1774-01-01T00:00:00-05:00"^^xsd:gYear	"6"	:Benjamin_Aislable ↗
"Benjamin Aislable"	"1774-01-01T00:00:00-05:00"^^xsd:gYear	"1842-01-01T00:00:00-05:00"^^xsd:gYear	:Benjamin_Aislable ↗
"Benjamin"	"1784-03-18"^^xsd:date	"1809-01-01T00:00:00-05:00"^^xsd:gYear	:Benjamin_Bathurst_%28diplomat%29 ↗

Find: ☒ Match case ☐ Phrase not found

Joining up the dots...

<http://map.psi.enakting.org/how>



Joining up the dots – UK Linked Data

<http://myarea.psi.enakting.org/>

How good is my area ?

feedback ... please [email us](#)

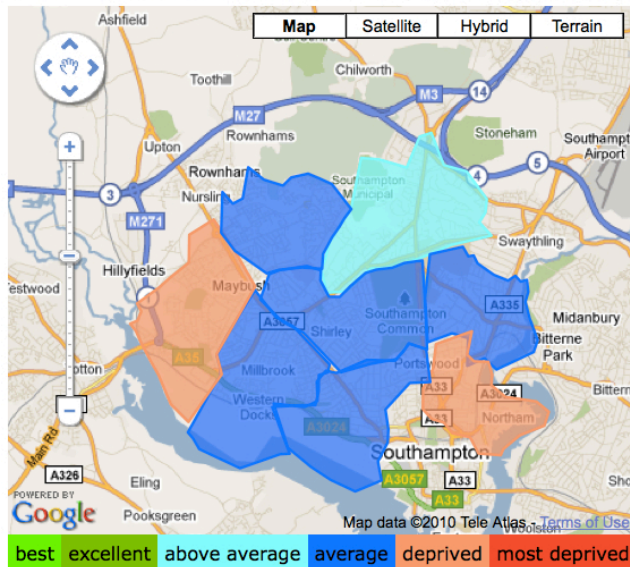
Sources:

[statistics.data.gov.uk](#), [data.ordnancesurvey.co.uk](#),
[IMD Database](#), [Boundary data from OS Open Data](#)

This is a **BETA** app developed in 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](#).

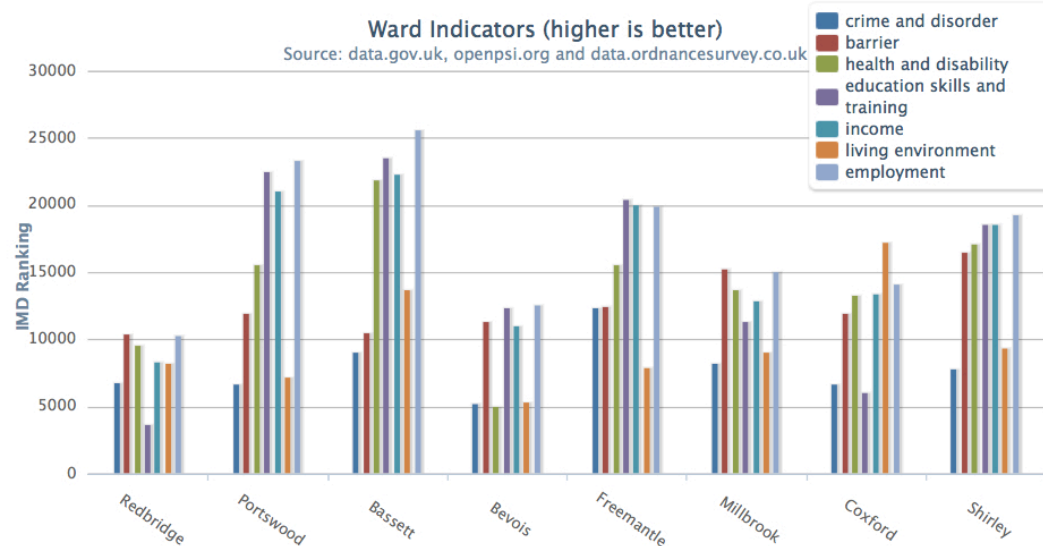
introduce an area name i.e: Portwood, Chilworth, Camden, ...



Area: Bevois

Ward Indicators (higher is better)

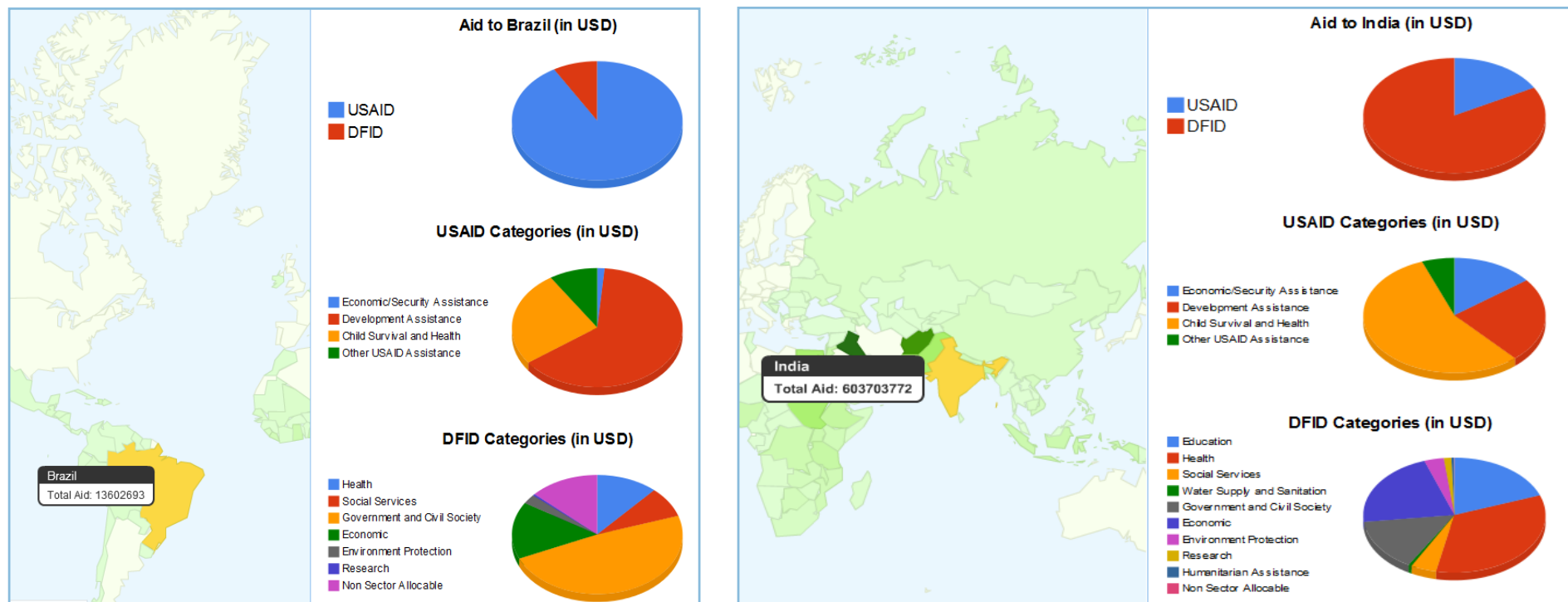
Source: [data.gov.uk](#), [openpsi.org](#) and [data.ordnancesurvey.co.uk](#)



[group by area](#)

US and UK Foreign AID FY2007:

Integrating data from two countries



Data Sources:



	AID	Major aids from US	Major aids from UK
Brazil	US > UK	Development Assistance	Gov & civil society, Economic
India	UK > US	Child Survival and Health	Health, Economic

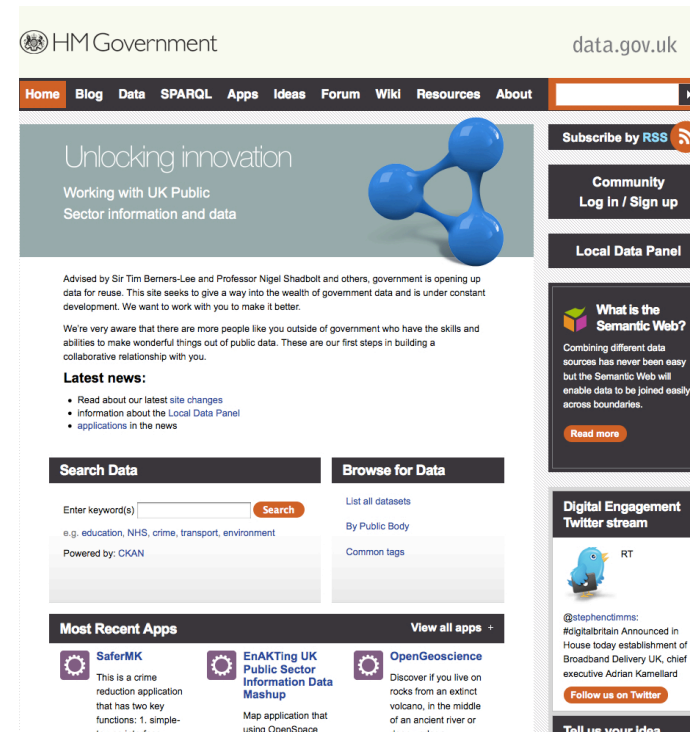
[Spatial Mashup] Data.gov (USAID) + Data.gov.uk (DFID)

Being extended to allow user annotations

Open Source

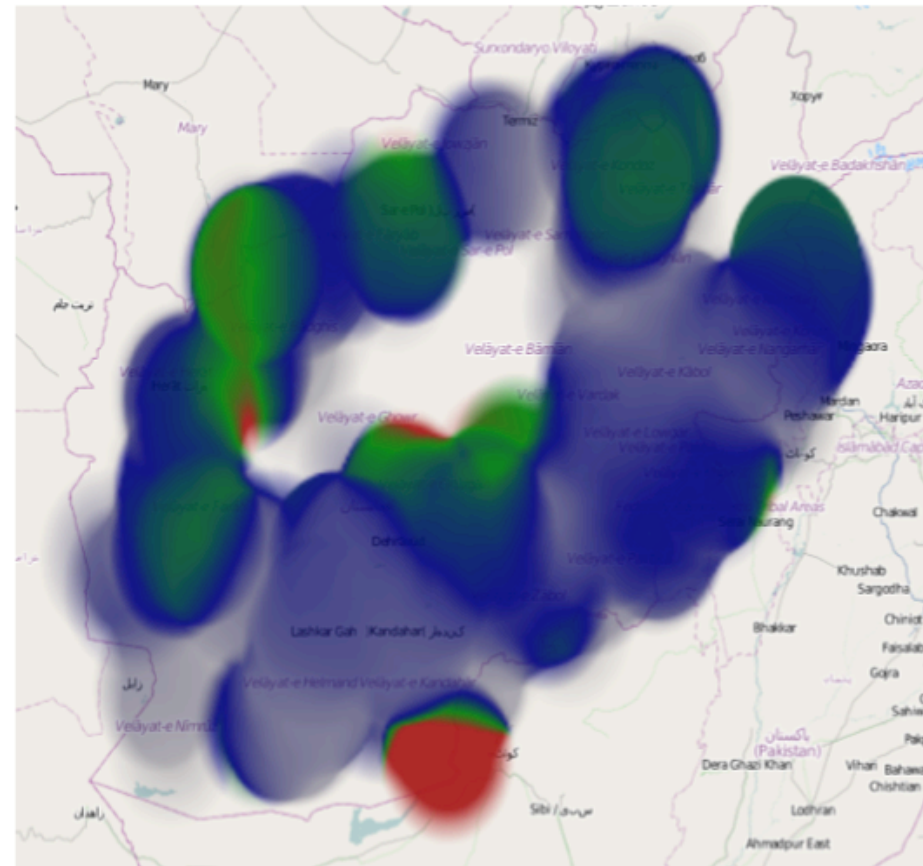
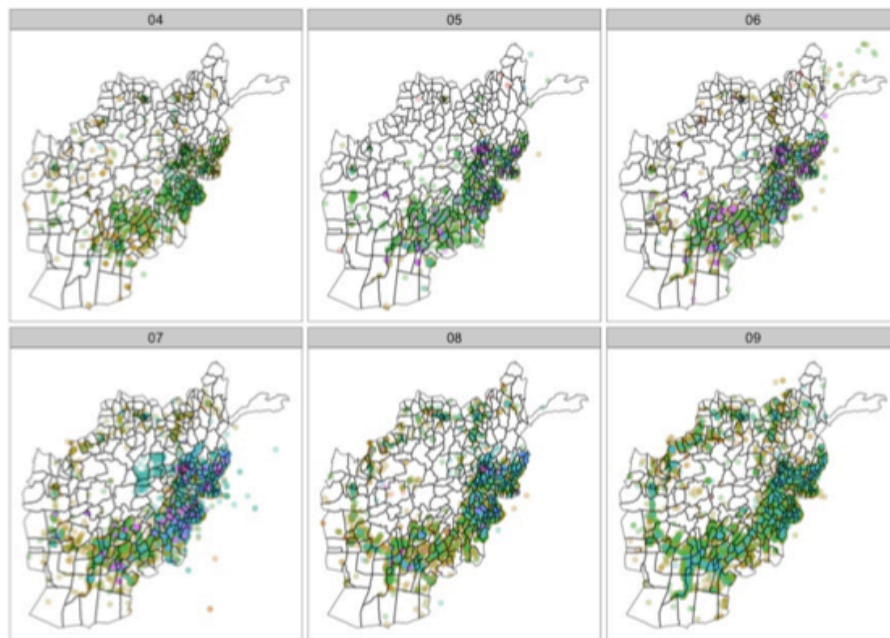
data.gov.uk

- Drupal and Media Wiki
- Cheap
- Robust
- Resilient
- Maintained



IED Data

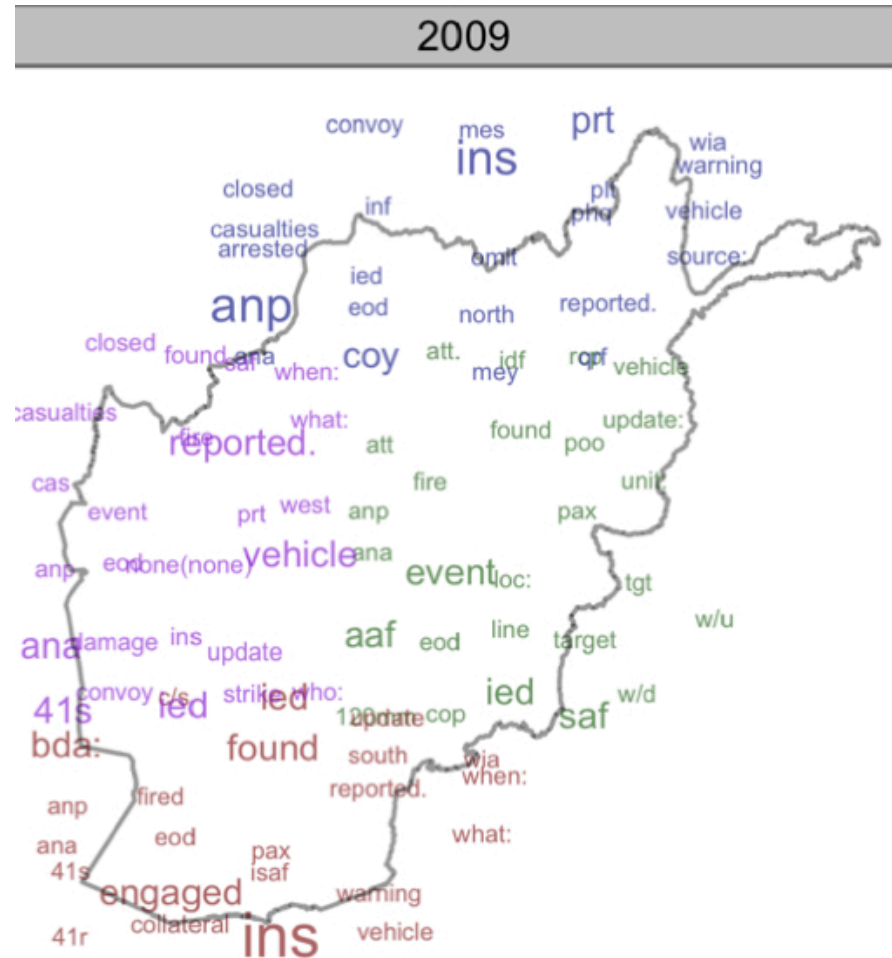
<http://www.wired.com/dangerroom/2010/08/open-source-wikileaks-docs-illustrated-afghan-meltdown/>



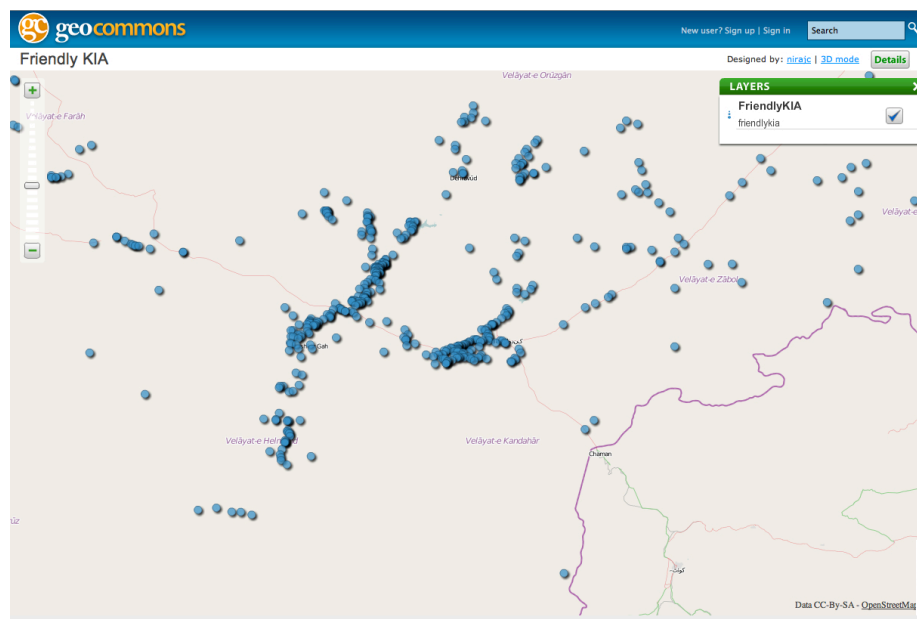
IED Data

```
100644 | 227 lines (200 sloc) | 9.271 kb raw | blame | histor

1 # File-Name:      wikileaks_tm.R
2 # Date:           2010-08-11
3 # Author:         Drew Conway
4 # Email:          drew.conway@nyu.edu
5 # Purpose:        Perform basic text mining operations on WL data
6 # Data Used:      see load_data.R
7 # Packages Used:
8 # Output File:
9 # Data Output:
10 # Machine:        Drew Conway's MacBook Pro
11
12 # Copyright (c) 2010, under the Simplified BSD License.
13 # For more information on FreeBSD see: http://www.opensource.org/licenses/bsd-license.php
14 # All rights reserved.
15
16 source("load_data.R")
17 source("utils.R")
18
19 ## LDA topic models
20 library(topicmodels)
21
22 # Create corpus from Summary column for all regions
23 create_lda<-function(afg.df, num.topics=3, num.words=10, stemming=FALSE, stopwords = TRUE, minWordLength = 3) {
24   # Create nested list of Summary topic models for all Year/Region pairs
25   ## WARNING: with above dtm.control settings this takes a VERY LONG TIME
26   ## To speed the process set stemming=FALSE and stopwords=FALSE, however,
27   ## results will be different/worse.
28   afg.terms<-list()
29   regions<-levels(as.factor(afg.df$Region))
30   years<-levels(as.factor(afg.df$Year))
31   dtm.control<-list(stemming = stemming, stopwords = stopwords, minWordLength = minWordLength)
32   for(y in years){
33     y.sub<-subset(afg.df,afg.df$Year==y)
34     for(r in regions){
35       r.sub<-subset(y.sub,y.sub$Region==r)
36       yr.corp<-Corpus(VectorSource(r.sub$Summary))
37       dtm<-DocumentTermMatrix(yr.corp,control=dtm.control)
38       dtm<-removeSparseTerms(dtm,0.95)
39       afg.terms[[y]][[r]]<-get_terms(LDA(dtm,control = list(alpha = 0.05), k = num.words),num.topics)
40     }
41     names(afg.terms[[y]])<-regions
42   }
43   names(afg.terms)<-years
44   return(afg.terms)
45 }
46
47 ## Only run this portion if you have not yet generated LDA
48 afg.terms<-create_lda(afg,stopwords=c(stopwords("english"),"report","reported","reports","report"),"delayed","s//rel","ro
49
```



Geo Commons



The screenshot shows the 'Niraj's Library' page on Geo Commons. The page has a blue header with the 'geo commons' logo, a search bar, and links for 'New user? Sign up | Sign in'. Below the header, there's a navigation bar with 'home', 'Niraj's Library', 'Layers', and 'Maps'. The main content area is titled 'Niraj's Library' and contains a list of overlays. Each overlay entry includes the name, the uploader, the upload date, and buttons for 'Download', 'Make a Map', and 'Details'. The overlays listed are 'FriendlyKIA', 'HostNationKIA', 'EnemyKIA', and 'CivilianKIA'. To the right of the list, there's a 'Sort By' section with options: 'Newest First', 'Oldest First', 'Name - A to Z', and 'Name - Z to A'. Below this is a 'Subscribe' section with a 'Subscribe to the Atom Feed' button and a 'View the GoogleEarth KML' button. At the bottom, there's an 'About Tags' section with a 'Tip: Use Tags to Collect and Share' section. The footer contains links for 'Need Help?', 'Follow Us', 'About Us', 'We Love Feedback', and 'Enterprise Solutions'.

Emerging DoD Interest

- <http://www.forge.mil/>



The screenshot shows the Forge.mil website homepage. At the top is the Forge.mil logo and a banner with the text "Transforming the Way DoD Innovates IT" and "Forge.mil is a collaborative environment to accelerate the development and deployment of dependable software and services within the Department of Defense." Below the banner is a navigation bar with links: HOME, COMMUNITY, FAQs, NEWS, RESOURCES, and SUPPORT. A button indicates "Browse over 300 projects & collaborate with over 6000 DoD members".

Top 10 FAQs

- ▶ What is the Forge.mil Program?
- ▶ What is SoftwareForge?
- ▶ What is ProjectForge?
- ▶ What is the difference between SoftwareForge and ProjectForge?
- ▶ What capabilities are available in the system?
- ▶ What are the guidelines for participating?
- ▶ How can I get access to Forge.mil (CAC users)?
- ▶ How can I get access to Forge.mil (ECA Certificate users)?
- ▶ How do I order a RACE Node?
- ▶ Where do I go to download DoD/ECA Root and Intermediate Certificates?

About This Program

Forge.mil is a family of services provided to support the DoD's technology development community. The system enables the collaborative development and use of open source and DoD community source software. For programs and projects that require greater access control, the system supports private collaborative development with an on demand, fee for service offering. These initial software development capabilities are growing to support the full system life-cycle and enable continuous collaboration among all stakeholders including Project Managers, developers, testers, certifiers, operators, and users.

Tools & Resources

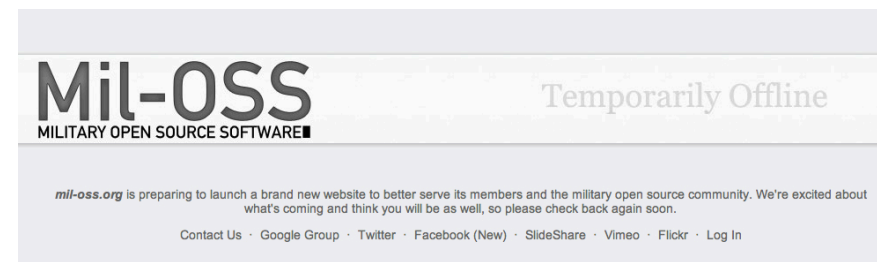
- Forge.mil Brochure & ProjectForge Info Sheet
- ECA Registration
- PKI Information
- Subversion Client Instructions
- Download DoD Root & Intermediate Certificates
- Download DoD Mozilla Firefox & Thunderbird Add-ons
- Accessibility & Section 508 Information

Join Community

Latest News

- ▶ Forge.mil Project Director, Rob Vietmeyer, and Community Manager, Guy Martin, named in 2010 Federal 100 Winners
- ▶ Forge.mil selected as 2010 Excellence.Gov Finalist

More FAQs



The banner for Mil-OSS (Military Open Source Software) features the text "Mil-OSS" in large, bold letters, with "MILITARY OPEN SOURCE SOFTWARE" underneath. To the right, it says "Temporarily Offline". Below this, a message states: "mil-oss.org is preparing to launch a brand new website to better serve its members and the military open source community. We're excited about what's coming and think you will be as well, so please check back again soon." At the bottom, there are links for "Contact Us", "Google Group", "Twitter", "Facebook (New)", "SlideShare", "Vimeo", "Flickr", and "Log In".

Open Licence



Open Government Licence
for public sector information

delivered by
The National Archives
[Back to The National Archives](#)

You are encouraged to use and re-use the Information that is available under this licence, the Open Government Licence, freely and flexibly, with only a few conditions.

Using information under this licence

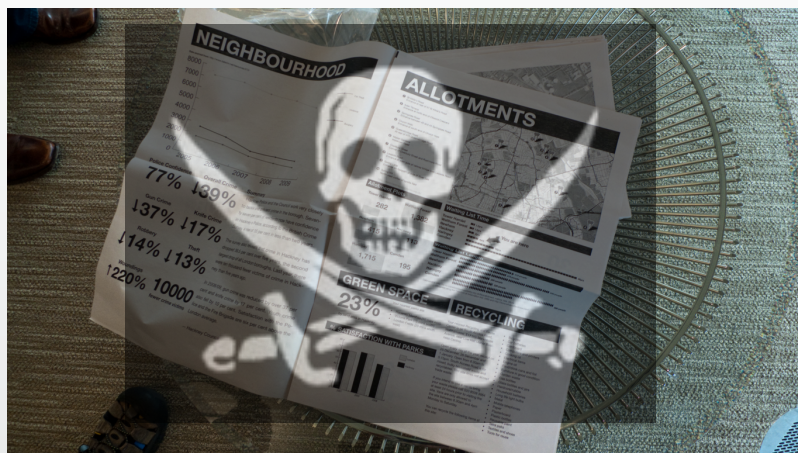
Use of copyright and database right material expressly made available under this licence (the 'Information') indicates your acceptance of the terms and conditions below.

The Licensor grants you a worldwide, royalty-free, perpetual, non-exclusive licence to use the Information subject to the conditions below.

You are free to:

- ✓ copy, publish, distribute and transmit the Information;
- ✓ adapt the Information;
- ✓ exploit the Information commercially for example, by combining it with other Information, or by including it in your own product or application.

You must, where you do any of the above:



Open Licencing

- OPEN SOURCE AND THESE UNITED STATES
- Major Justin Seiferth

[http://www.dtic.mil/cgi-bin/GetTRDoc?
Location=U2&doc=Get
TRDoc.pdf&AD=ADA3
98898](http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA398898)



Open Minds

Challenges

- Cultural & Organisational
 - Data hugging
 - We have released it - sort of...
 - Licence impediments
- Worries about
 - Confidentiality
 - How it will be Interpreted
 - QoI
 - New IT procurement
- Capability
 - What assets have we got
 - Who can do this work
 - How to share best practice
- Policy
 - Adopting and extending the principles
 - Consistency in what is published and its format
 - Range of exceptions do need to be appreciated
 - Confidentiality, accuracy, liability and reputation

Web of Linked Data

- The Web of Linked Data is next
- Can change
 - Government
 - Business
 - Cities
 - Universities
 - :
- Exploitation of
 - open data
 - open standards
- Can apply to data intranets as well as extranets

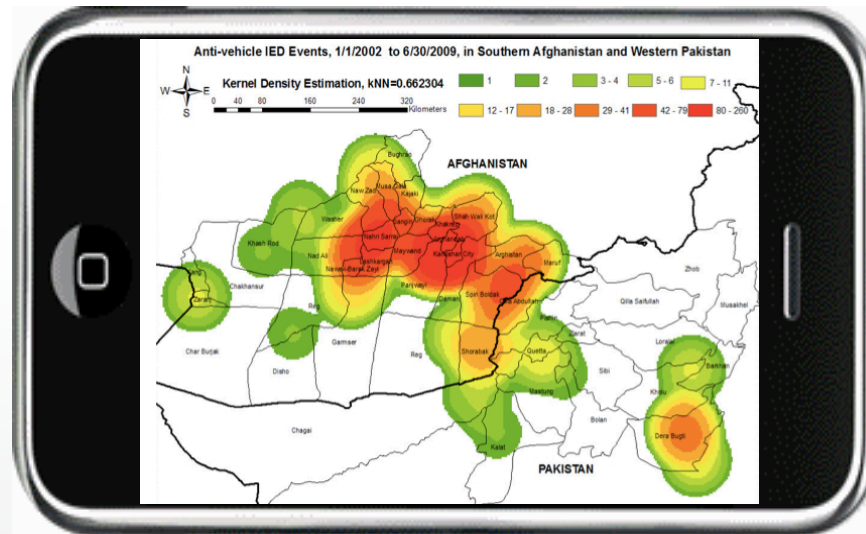


Why does this matter to you

- Supports information interoperability
- Sum is greater than the parts
- Exploitation of
 - open data
 - open standards
 - open source
 - open licence
- Can apply to data intranets as well as extranets
- Military and intelligence contexts are classic examples where there is heterogeneous information that could be shared, linked, mapped and enriched using this approach



Challenge – this in 2 days?



Challenge – this in 2 days?



Thank you

- Colleagues with me to also answer questions!
- Any further information or follow ups regarding research please email nrs@ecs.soton.ac.uk
- Any further information or follow ups regarding application development or consultancy please email tim.organ@seme4.com

www.seme4.com

