

The Foundation for Science and Technology

Foundation Lecture

8th December, 2010
Held at The Royal Society

Professor Nigel Shadbolt FEng

Professor of Artificial Intelligence
School of Electronics and Computer Science
Southampton University

The Wonder of the Web

The Foundation is very grateful to Sir William Francis CBE FEng for support for this lecture.

The slides from the presentation are below.

10 Carlton House Terrace
London
SW1Y 5AH

Company Number: 1327814
Charity Number: 274727

The Wonder of the Web

Professor Nigel Shadbolt

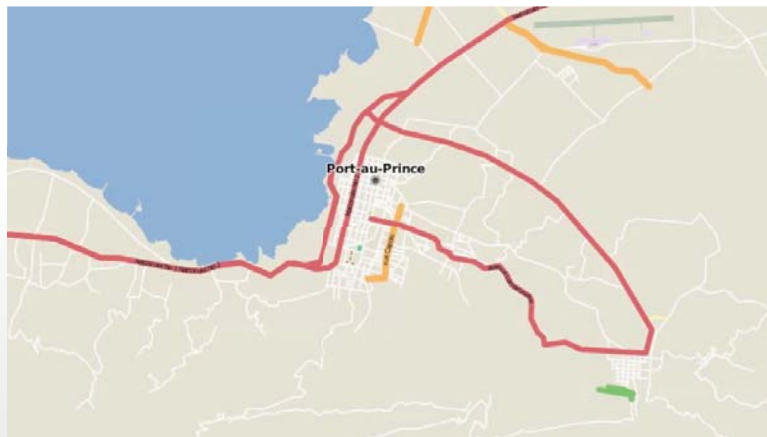
8th December 2010
FST, Royal Society, London



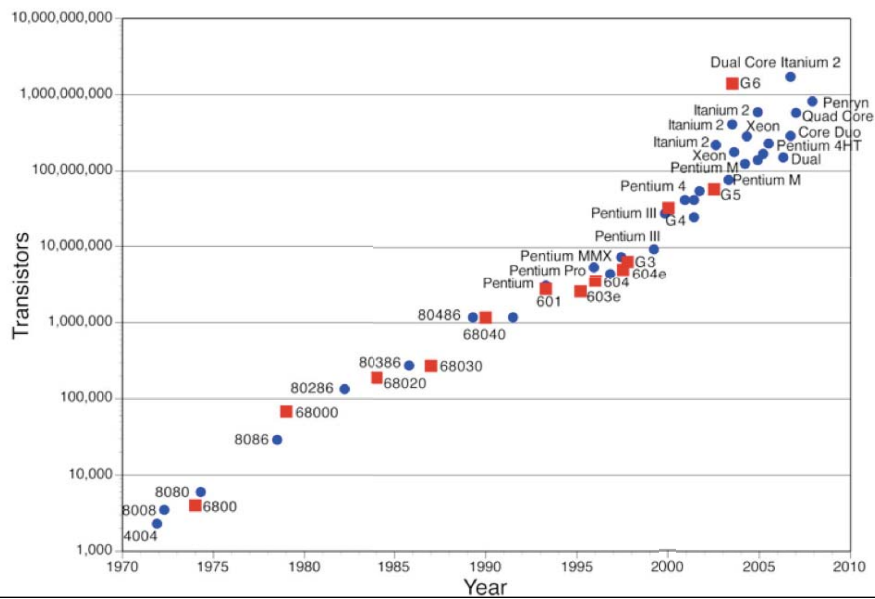
An appropriate sense of awe



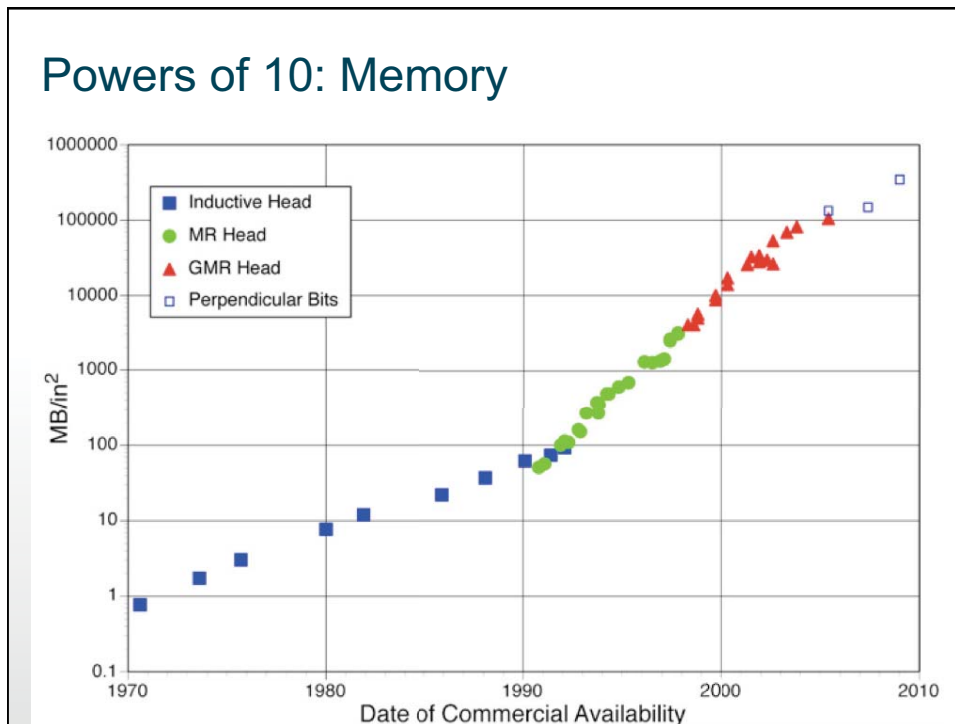
An appropriate sense of awe



Powers of 10: Processors



Powers of 10: Memory

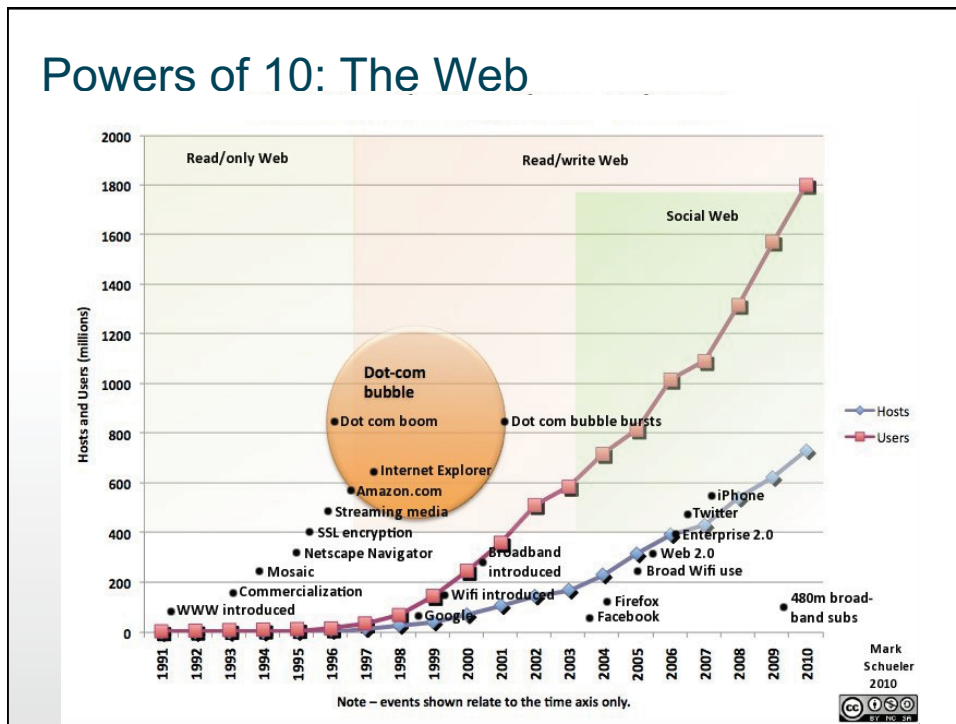


Powers of 10: Data

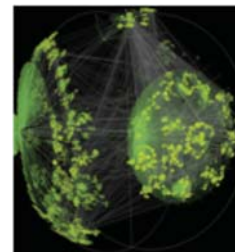
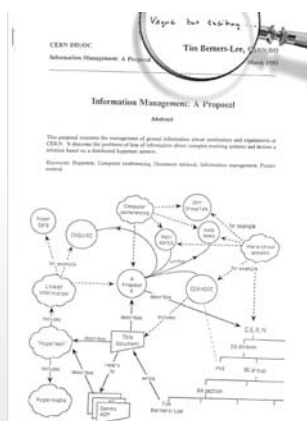
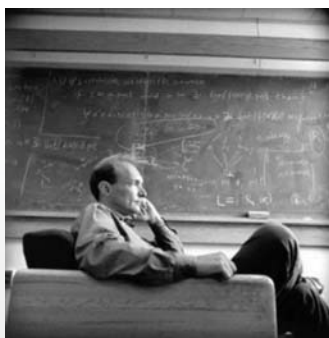


- 2007 – 281 EB (1 exabytes 1×10^{18} bytes)
- Or 281 trillion digitised novels
- 2010 – 1200 EB
- 2011 – 1.8 ZB

Powers of 10: The Web

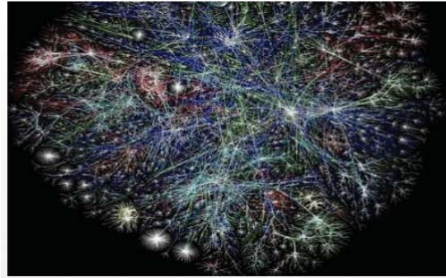


The Web – most successful information architecture in history

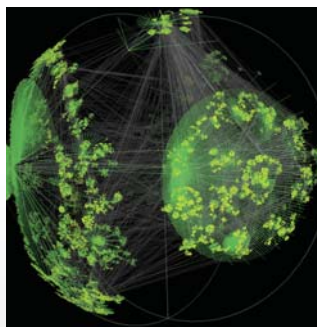


The Web – we need to understand it

- Web architecture is simple set of protocols
- These give rise to complex macro phenomena
- Need systems oriented view of the Web and its ecosystem – Web Science
- One that acknowledges social and technical components



The Web – emergent shape and structure



Scale-free

- Some nodes are of high degree most are low degree

Power laws

- The degree distribution follows a power law, with an exponent $\beta > 2$.

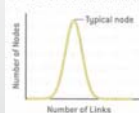
Small worlds

- The average distance (or diameter) is much smaller than the order of the graph.

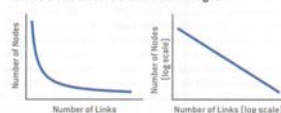
Hubs and authorities

- The number of distinct bipartite cliques or cores is large when compared to a random graph with the same number of nodes and edges.

Bell Curve Distribution of Node Linkages



Power Law Distribution of Node Linkages



The Web – exploiting shape and structure



$$I^0 = c_1 v_1 + c_2 v_2 + \dots + c_n v_n$$

$$I^1 = S I^0 = c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \dots + c_n \lambda_n v_n$$

$$I^2 = S I^1 = c_1 \lambda_1^2 v_1 + c_2 \lambda_2^2 v_2 + \dots + c_n \lambda_n^2 v_n$$

$$\vdots$$

$$I^k = S I^{k-1} = c_1 \lambda_1^k v_1 + c_2 \lambda_2^k v_2 + \dots + c_n \lambda_n^k v_n$$

$$G I^k = \alpha H I^k + \alpha A I^k + \frac{1-\alpha}{n} 1 I^k$$

The Web – exploiting shape and structure

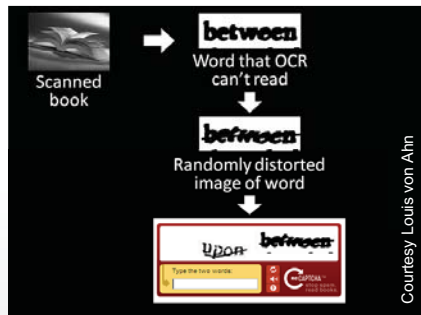
80% of the world's content is now user generated

huge amounts of it are unstructured

meaning based computing



The Web – emergent problem solving



The Web – exploiting emergent problem solving



amazonmechanical turk

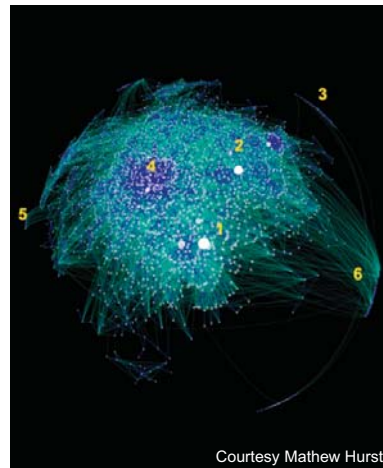
Mechanical Turk is a marketplace for work.

Make Money by working on HITs

Get Results from Mechanical Turk Workers

76,093 HITs available. View them now.

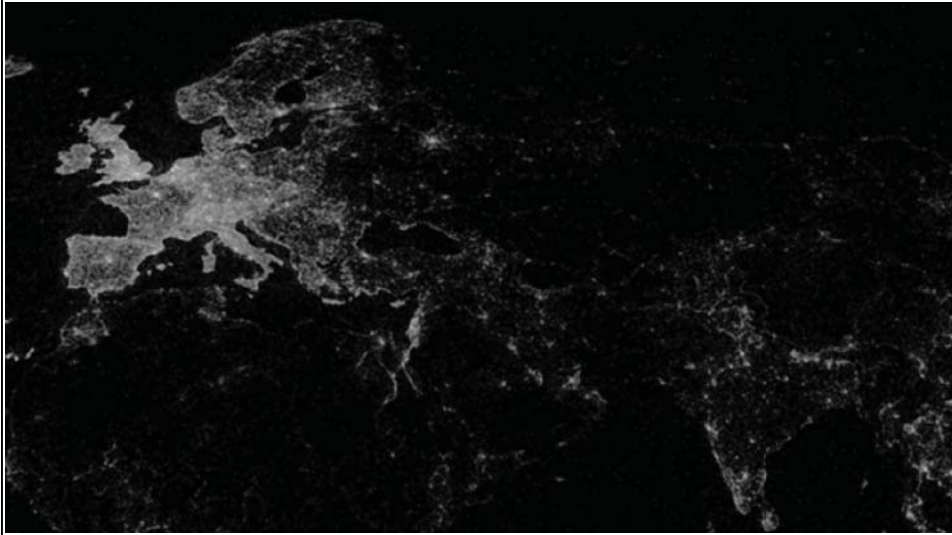
The Web – emergent networks



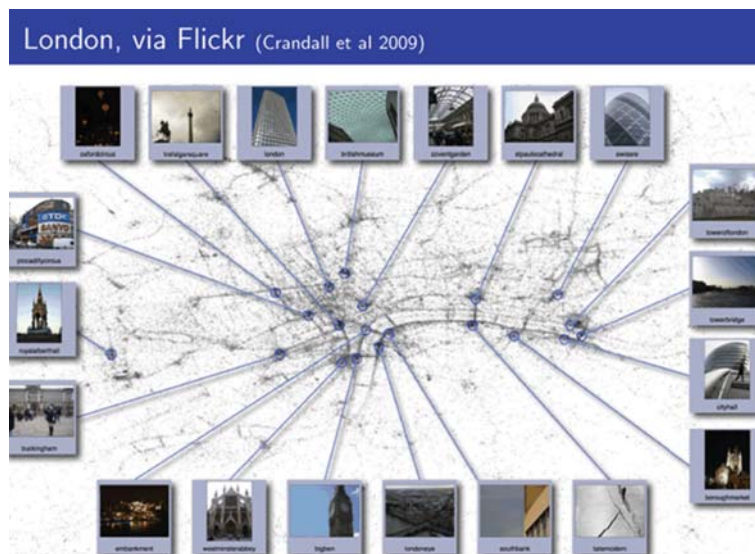
The Web – exploiting emergent networks



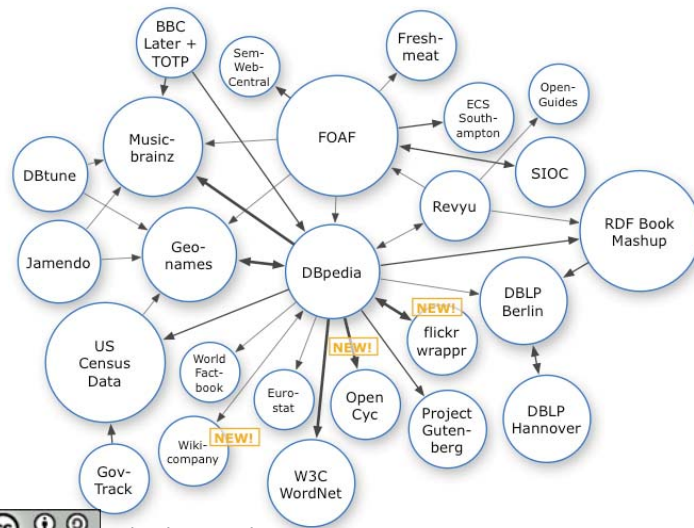
The Web – emergent data



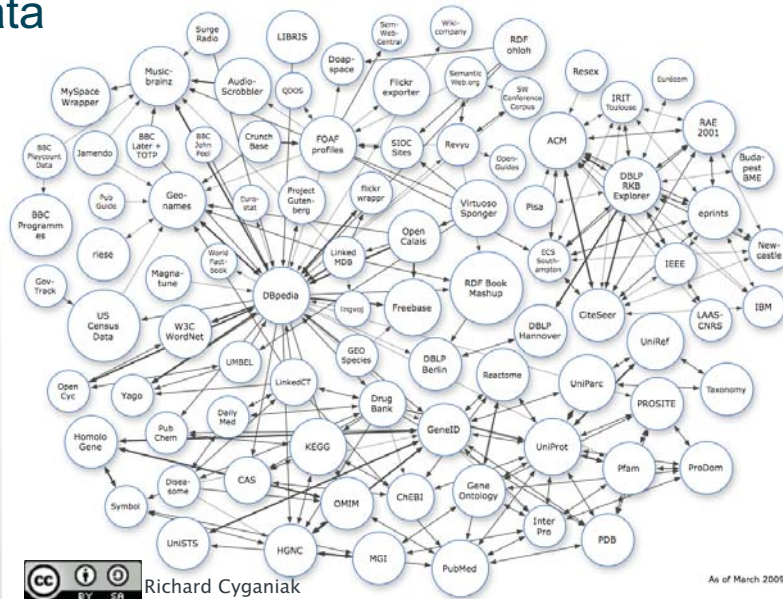
The Web – exploiting emergent data



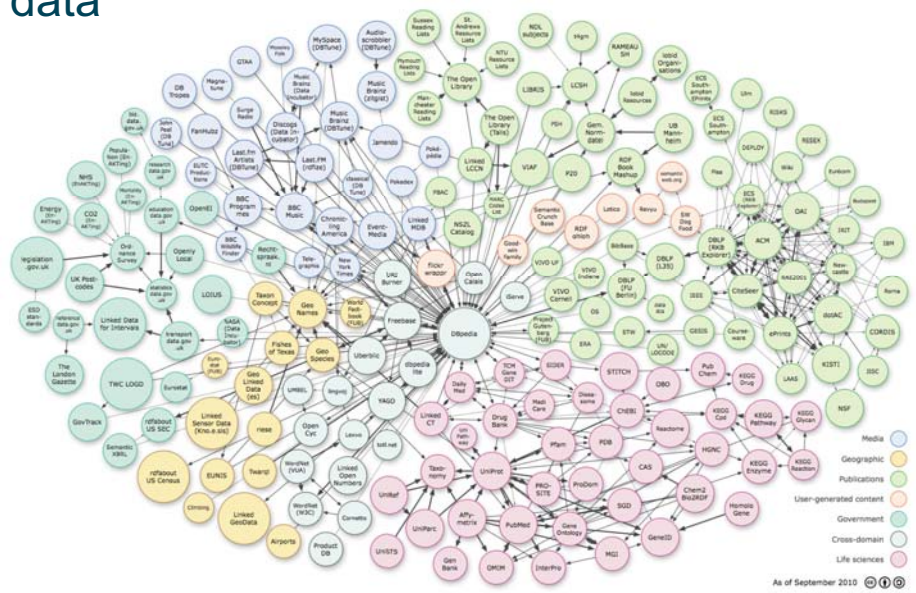
The Web – emergent networks of linked data



The Web – emergent networks of linked data



The Web – emergent networks of linked data



The Web - Linked Data Principles

The four micro principles of the Semantic Web

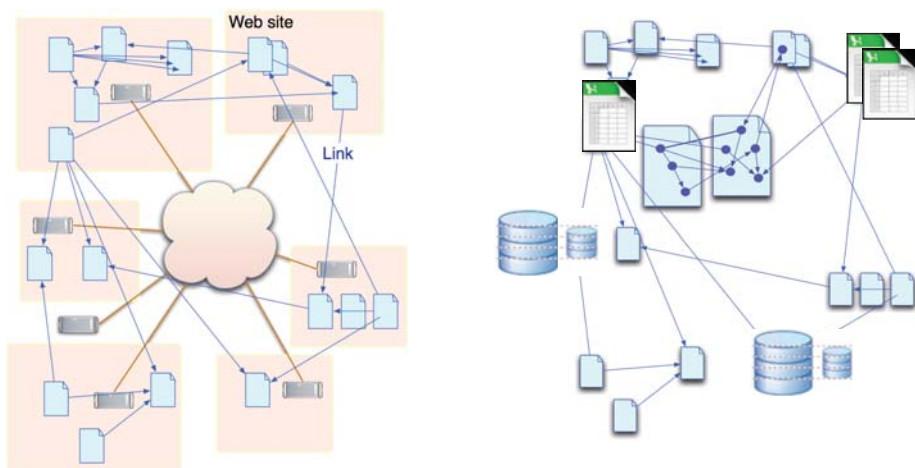
1. All entities of interest, such as information resources, real-world objects, and vocabulary terms should be identified by URI references.
2. URI references should be dereferenceable, meaning that an application can look up a URI over the HTTP protocol and retrieve RDF data about the identified resource.
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



The Web - of documents and data



Principle 1 URIs for Everything

- Uniform Resource Identifiers for all items of interest
- You can dereference them – use http protocol to get data back
- Using a simple Web Knowledge Representation Language – RDF
- Link URIs together

<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://dbpedia.org/resource/Nigel_Shadbolt

Principle 2 URIs will dereference

- Uniform Resource Identifiers for all items of interest
- You can dereference them – use http protocol to get data back
- Using a simple Web Knowledge Representation Language – RDF
- Link URIs together

*Southampton ECS People: Professor Nigel R Shadbolt

Property	Value
Description	This RDF document contains information about a person
Source	Southampton
Title	http://www.ecs.soton.ac.uk/people/nr/
Date Created	Southampton ECS People: Professor Nigel R Shadbolt
type	2009-11-26T10:53:20Z
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://rd.eecs.soton.ac.uk/role/2986
member of	http://rd.eecs.soton.ac.uk/project/2921 http://rd.eecs.soton.ac.uk/project/2923 http://rd.eecs.soton.ac.uk/project/2912 http://rd.eecs.soton.ac.uk/project/2923 http://rd.eecs.soton.ac.uk/project/2981 http://rd.eecs.soton.ac.uk/project/2985 http://rd.eecs.soton.ac.uk/project/2943 http://rd.eecs.soton.ac.uk/project/2944 http://rd.eecs.soton.ac.uk/project/2945 http://rd.eecs.soton.ac.uk/project/2923 http://rd.eecs.soton.ac.uk/project/2944 http://rd.eecs.soton.ac.uk/project/2945 http://rd.eecs.soton.ac.uk/project/2946 http://rd.eecs.soton.ac.uk/project/2947 http://rd.eecs.soton.ac.uk/project/2948 http://rd.eecs.soton.ac.uk/project/2949 http://rd.eecs.soton.ac.uk/project/2950 http://rd.eecs.soton.ac.uk/project/2951
type	Person
family name	Shadbolt
givenname	Nigel R
homepage	http://users.ecs.soton.ac.uk/nr/
img	http://www.ecs.soton.ac.uk/image.php?id=person_2981 http://www.ecs.soton.ac.uk/image.php?id=person_2981 http://www.ecs.soton.ac.uk/image.php?id=person_2981
inbox	mailto:nr@ecs.soton.ac.uk
name	Professor Nigel R Shadbolt

Principle 3 Get RDF Back

- Uniform Resource Identifiers for all items of interest
- You can dereference them – use http protocol to get data back
- Using a simple Web Knowledge Representation Language – RDF
- Link URIs together

*Southampton ECS Projects: MIMEX: Multivariant Information Management and Exploitation

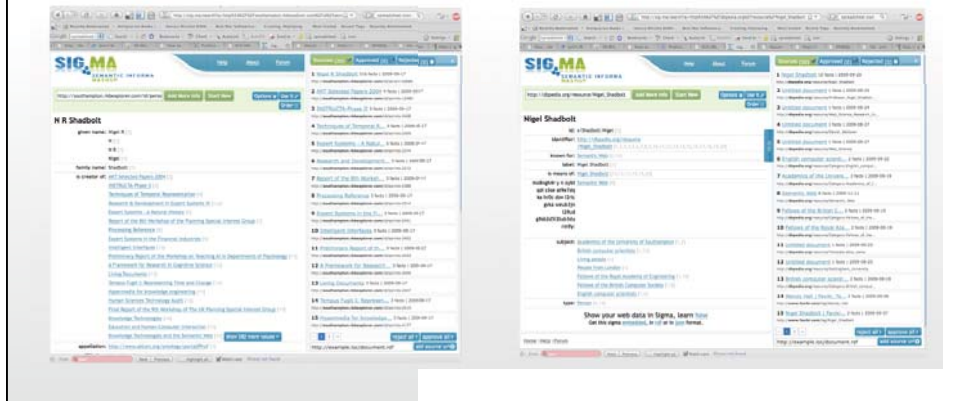
Property	Value
Description	This RDF document contains information about
Source	University of Southampton
Title	http://www.ecs.soton.ac.uk/research/projects/
Date Created	Southampton ECS Projects: MIMEX: Multivariant Information Management and Exploitation
type	2009-11-26T10:58:31Z
label	Southampton ECS Projects: MIMEX: Multivariant Information Management and Exploitation
associated With	http://rd.eecs.soton.ac.uk/2943
type	General Dynamic Organization
Name	University of Southampton
type	Organization
Name	Ministry of Defence
type	Organization
associated With URI	http://rd.eecs.soton.ac.uk/2943
Name	MIMEX: Multivariant Information Management and Exploitation
Theme	http://rd.eecs.soton.ac.uk/Theme/109 http://rd.eecs.soton.ac.uk/Theme/127 http://rd.eecs.soton.ac.uk/Theme/130 http://rd.eecs.soton.ac.uk/Theme/136 http://rd.eecs.soton.ac.uk/Theme/166 http://www.evidence.org/theses
Web Page run By Group	http://rd.eecs.soton.ac.uk/group/iam
begins	in Calendar Clock day hour minute month year type
ends	in Calendar Clock Date Type 2009-07-03T00:00:00Z type instant in Calendar Clock day hour minute month year type in Calendar Clock Date Type 2009-05-31T00:00:00Z type instant

Principle 4 Link Data URIs

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

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

http://dbpedia.org/resource/Nigel_Shadbolt



The Web – exploiting emergent content



The Web - Wikipedia as Linked Data

- 2,900,000 things
- >282,000 persons
- 339,000 places
- 130,000 species
- 4,400 diseases
- 88,000 music albums
- :



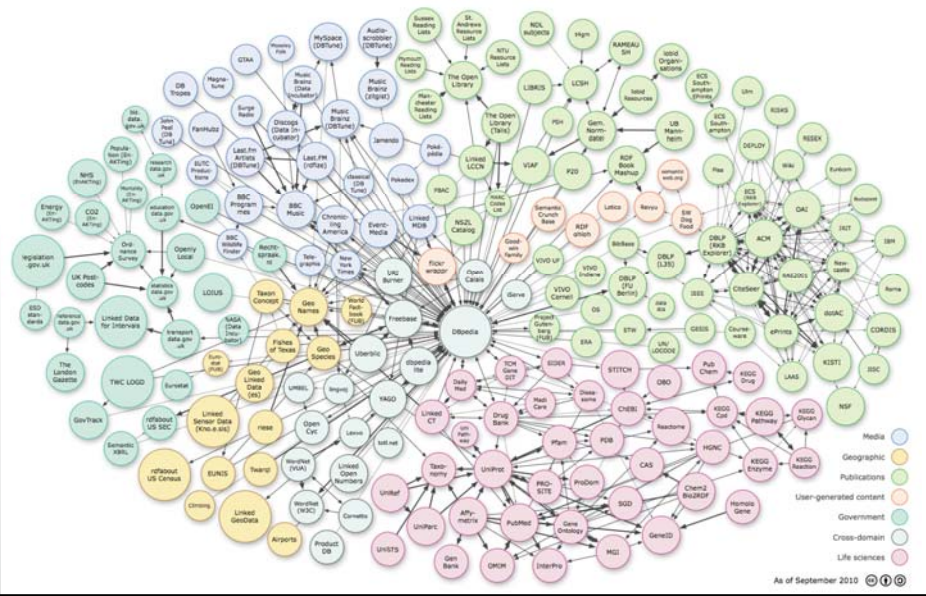
The Web - adding a little SPARQL

- A data access language for the Web of Linked Data
- Can query across diverse data sources
- SPARQL can query required and optional patterns

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX type: <http://dbpedia.org/class/yago/>
PREFIX prop: <http://dbpedia.org/property/>
SELECT ?country_name ?population
WHERE {
  ?country a type:LandlockedCountries ;
    rdfs:label ?country_name ;
    prop:populationEstimate ?population .
  FILTER (?population > 15000000) .
}
```

country_name	population
Ethiopia	82825000
Uganda	32710000
Nepal	29331000
Afghanistan	28150000
Uzbekistan	27606007
Burkina Faso	15757000
Niger	15290000
Malawi	15263000

The Web – Open Government Data



Open Government Data

Open City Data

The screenshot shows the London.gov.uk website's 'Open City Data' section. At the top, there's a search bar and navigation links. The main heading is 'A first step towards freeing London's data'. Below this, a 'Welcome to the London Datasource' section provides an overview. To the right, a 'Data Catalog' section lists various datasets, and a 'Browse Catalog' section shows a table of datasets with columns for Name, Release Date, and Format.

Open Local Authority Data

The screenshot shows the Redbridge i website. The main heading is 'Redbridge Conversation 2010'. Below this, there are sections for 'My Area - Stone', 'Your Neighbourhood', and 'Transparency'. The 'Transparency' section contains a detailed text document about transparency in the council's operations, including sections on 'Annual Information and Data Transparency (AIDT)', 'Freedom of Information Act (FOIA)', and 'Freedom of Information Act (FOIA) - 2005 Publications Scheme'.

Started in the UK



University of Southampton project to show how open data could deliver services across Government (2005-06)



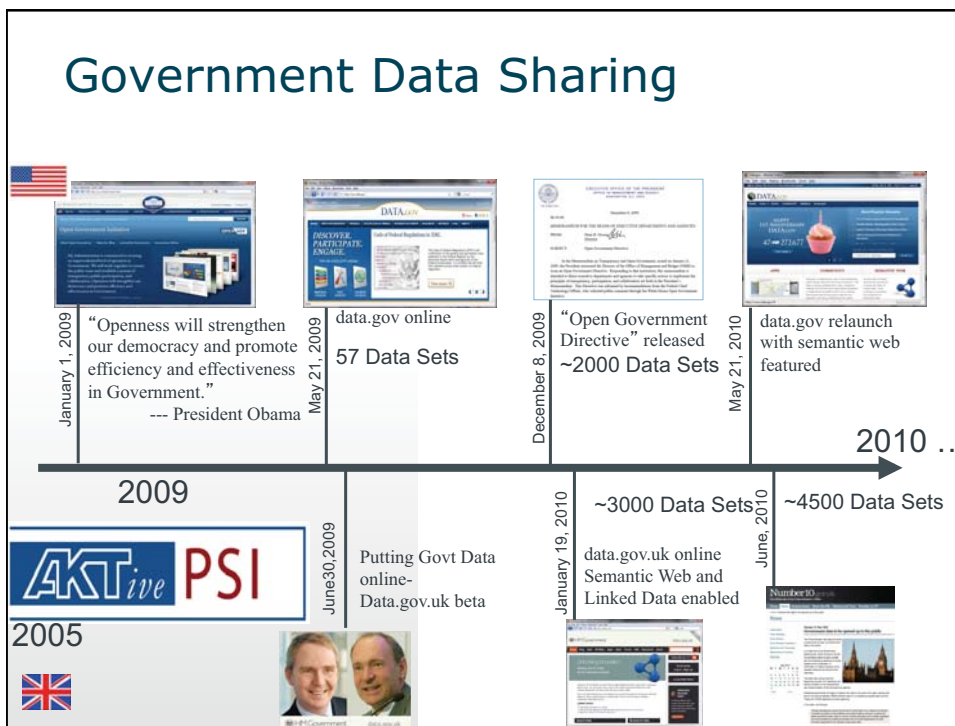
Reported to UK Parliament 2007



A Report by the Office of Public Sector Information
July 2007

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),²⁶ 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. It 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 technology²⁷. OPSI is using this technology in the following ways:



Why?

- Transparency, accountability and engagement
- Create economic and social value
- Improve Public Services
- e-Government – better for less

Early UK OGD Apps – ASBOs to Dentists



UK Dentists

Find your nearest NHS dentist quickly and easily!

The National Health Service has over 7,000 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 86.4% 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 click, name or postcode search. Dentists can be displayed in maps or thematic grids (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 License.

Application Screenshots

Click on any of the images to see a larger version

Buy on iTunes

UK OGD – Location...

OS OpenData

Mapping data and geographic information from Ordnance Survey

Creates 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.

New available: OS VectorMap™ District - a new mid-scale vector and raster dataset (please see news).

view **develop** **supply**

- Outline of Great Britain
- Overview of Great Britain
- MapGrids ES
- 1:250,000 Scale Colour Raster
- OS Street View ES
- Boundary-Line™
- OS VectorMap™ District - New

about **help** **keep in touch**

Use [download and create](#) OS OpenData or use [OS OpenData](#) to enhance your own application.

You can also experiment with [OS OpenData in ES2 format](#).

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

- The Ordnance Survey [Map Shop](#) for paper maps
- The [Solutions Centre](#) for mobile, phone and kiosk
- Ordnance Survey mapping for [business](#) or [education](#)

help

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

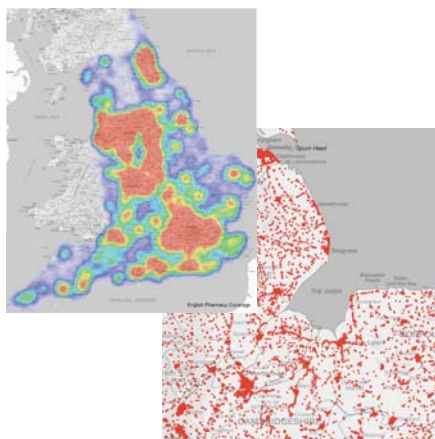
Do you can't find or still in on 0844 848 8888 (9.00am - 5.00pm, Monday to Friday).

keep in touch

Follow us on Twitter, Facebook or YouTube.

Feedback stage

Ordnance Survey Shop, Sir Tom Brimacombe, Professor Steve Bradburn, OGDge (equipment)



UK OGD – Spending...

The image shows two web dashboards. The left one is the Guardian.co.uk 'Coins Data Explorer', which allows users to search for specific OGD data points. It features a search bar, a list of departments to filter by, and a table of results with columns for department, amount, and percentage change. The right dashboard is 'spotlightsonspend', which provides an overview of spending for Royal Boroughs of Windsor & Maidenhead. It includes a table of various services with their respective costs and percentage changes, such as 'Council Services' at £72,820,736 (-36%) and 'Waste' at £44,161 (+134%).

UK OGD – Behaviour...

Real time data changes behaviour – the decisions and actions people take

The image shows a 'gov.uk' dashboard for the 'Competition' section. It features a green header and a main content area with text explaining the data. A 'Ranking' section lists three items: 'defra' with a 22% increase, 'DWP' with a -18% decrease, and 'HM JUSTICE' with a -14% decrease. Below the ranking is a map showing geographical data points. The text on the page discusses how real-time data is used to inform decisions and actions, specifically mentioning the impact of the 2010 Spending Review and the need for departments to be more efficient.

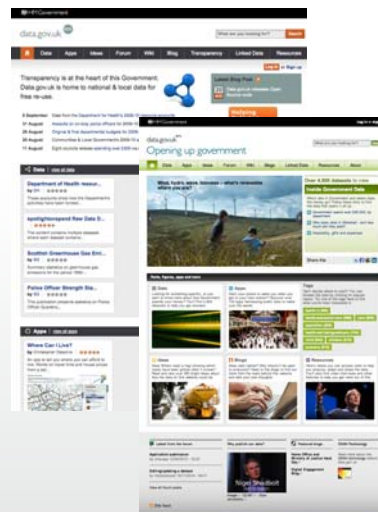
Public Data Principles

- Release driven by public and business
- Publish the data unless good reason not to
- Public data will be timely and fine grained
- Public data will be published in reusable, machine-readable form
- Release data quickly, and then re-publish it in linked data form
- Public data will be released under Open Government Licence



Public Data Principles

- Public data will be available through data.gov.uk
- Public bodies should maintain and publish inventories of their data holdings
- Public data underlying websites will be published in reusable form for others to use
- Public data will be freely available to use in any lawful way
- Public bodies should actively encourage the re-use of their public data



The road to stardom

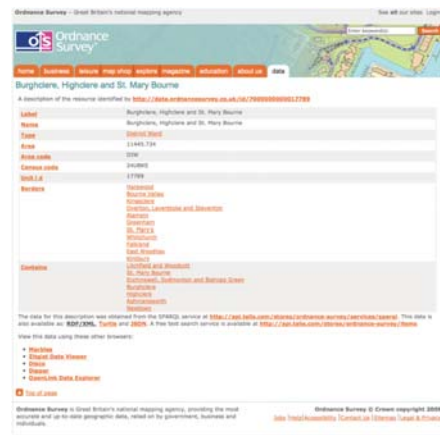
- ★ Put your data on the Web (any format)
- ★★ Make it available as structured data
- ★★★ Use open, standard formats
- ★★★★ Use URLs to identify things
- ★★★★★ Link your data to other people's data

From bus stops to ministers – a linked Web of data

The screenshot displays a web application interface for a linked data system. On the left, there are three panels for bus stops: 'Entry Hill (Dep)', 'St Marks School', and 'Windsor Bridge Road'. Each panel shows fields for 'name', 'lat', 'long', 'wgs84:lat', 'wgs84:long', 'wgs84:lat', and 'wgs84:long'. In the center, a map shows the location of 'St Marks School' with a list of nearby bus stops and a 'Sort by' dropdown menu. On the right, there are three panels for political figures: 'N Secretary in Northern Ireland Office', 'Dr Liam Fox', and 'Shailesh Vara'. Each panel shows fields for 'name', 'description', 'office', 'party', and 'role'. A 'View' dropdown menu is at the bottom right.

data.gov.uk 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



Challenges – Computation

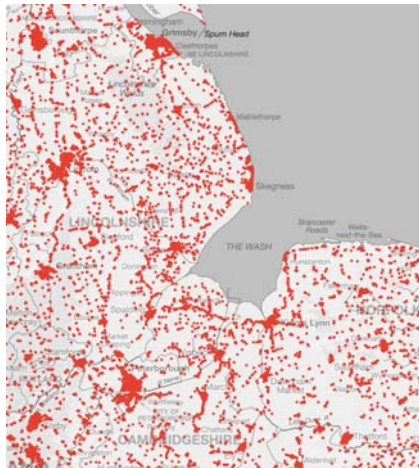
- There will be a very great deal of data....
- Who bears the cost of supporting it?
- Can we really treat the Web as a large decentralised database?
- Dotsam and netsam?



Shutterstock

Challenges – Quality

- NaPTAN (public access transport points)
- Includes 360,00 bus stops
- Around 18,000 errors



Challenges – Quality

- NaPTAN (public access transport points)
- Includes 360,00 bus stops
- Around 18,000 errors
- Which can be crowd source improved



Challenges – Interpretation

- Can't trust people with the data
- They might interpret it incorrectly
- Do they have the skills
- New levels of data literacy



Challenges – Security & Privacy

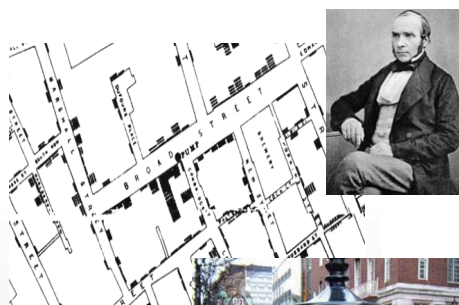


Challenges – Security & Privacy

- Practical obscurity no longer works – information triangulation
- Information triangulation
- Need social conventions – and legal agility



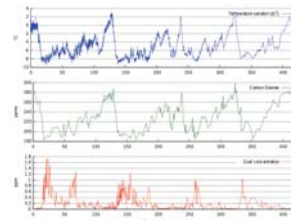
Open data - basis of our modern science




PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD

Vol. I.
For Anno 1665, and 1666.

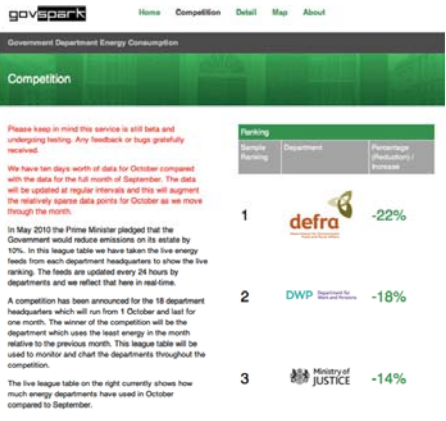
In the SPOT,
Printed by T. N. for John Murray at the Bell, a little without Temple-Bar, and Spence's Office in Duke-Lane,
Printers to the Royal Society.



Open Data - basis of modern government



Area Name	Current Spend	Percentage Change
Windsor & Maidenhead Council	£72,820,736	-36%
Windsor & Maidenhead Council	£44,161	+134%
Windsor & Maidenhead Council	46%	-8.8%
Windsor & Maidenhead Council	15%	-7.0%
Windsor & Maidenhead Council	£1,296	-16%
Windsor & Maidenhead Council	£828	-12%
Windsor & Maidenhead Council	£2,783	-6%
Windsor & Maidenhead Council	£2,540	-23%



govspark Home Competition Detail Map About

Government Department Energy Consumption

Competition

Please keep in mind this service is still beta and undergoing testing. Any feedback or bugs gratefully received.

We have ten days worth of data for October compared with the data for the full month of September. The data will be updated at regular intervals and this will augment the relatively sparse data points for October as we move through the month.

In May 2010 the Prime Minister pledged that the Government would reduce emissions on its estate by 10%. In this league table we have taken the live energy feeds from each department headquarters to show the live rating. The feeds are updated every 24 hours by departments and we reflect that here in real-time.

A competition has been announced for the 18 department headquarters which will run from 1 October and last for one month. The winner of the competition will be the department which uses the least energy in the month relative to the previous month. This league table will be used to monitor and chart the departments throughout the competition.

The live league table on the right currently shows how much energy departments have used in October compared to September.

Ranking	Energy Profile Rating	Department	Percentage (Percentage) Increase
1		defra	-22%
2		DWP	-18%
3		Ministry of JUSTICE	-14%

[Open] Linked Data – a basis for modern business?



Area Name	Current Spend	Percentage Change
Windsor & Maidenhead Council	£72,820,736	-36%
Windsor & Maidenhead Council	£44,161	+134%
Windsor & Maidenhead Council	46%	-8.8%
Windsor & Maidenhead Council	15%	-7.0%
Windsor & Maidenhead Council	£1,296	-16%
Windsor & Maidenhead Council	£828	-12%
Windsor & Maidenhead Council	£2,783	-6%
Windsor & Maidenhead Council	£2,540	-23%



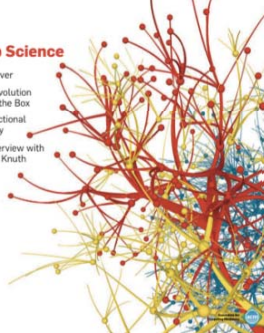



The Web – we need to understand it

COMMUNICATIONS OF THE ACM

Web Science

- XML Fever
- The Revolution Inside the Box
- Transactional Memory
- An Interview with Donald Knuth



INFORMATION TECHNOLOGY

Web Science EMERGES

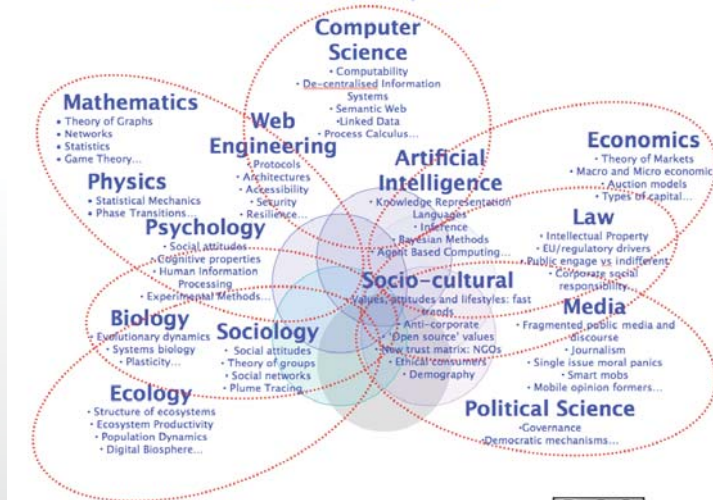
Studying the Web will reveal better ways to exploit information, prevent identity theft, restructure industry and manage our ever growing online lives

By Nigel Shadbolt and Tim Berners-Lee

Since the World Wide Web blossomed in the mid 1990s, it has revolutionised the way we live and work. The number of devices that are online has exploded and the amount of data being generated is growing exponentially. This data is being used to create new products and services, to improve existing ones and to create new ones. The Web is becoming a central part of our lives and it is important that we understand it better. This is the goal of Web Science, a new discipline that is emerging from the intersection of computer science, social science, psychology, economics, politics, law, medicine, biology, and other disciplines. Web Science is a multidisciplinary field that seeks to understand the Web as a complex system and to use this understanding to create new products and services. The Web is a complex system that is constantly changing and it is important that we understand it better. This is the goal of Web Science, a new discipline that is emerging from the intersection of computer science, social science, psychology, economics, politics, law, medicine, biology, and other disciplines. Web Science is a multidisciplinary field that seeks to understand the Web as a complex system and to use this understanding to create new products and services.

The components of Web Science

Web Science: Components



Web Science – why this matters

- an essential part of humanity
- an essential part of the current and future economy, science and technology
- understanding the Web is a major challenge as big as any other global cause



“Pivotal events of lasting significance, marking major developments of the greatest importance... an area of great importance for the next 20–30 years ”

The screenshot shows the Royal Society website page for an event titled "Web science: A new frontier". The page includes the Royal Society logo, event details (8:00am on 27 September 2010, 6:00pm on 28 September 2010), location (The Royal Society, London), and organizers (Professor Nigel Shadbolt, University of Southampton; Professor Carme Wernis, University of Southampton; Professor James Hendler, Memorial University; and Professor Bill Dunton, University of Oxford). It also features a "Watch a podcast of the event" button, a "Registration for this event is closed" notice, and a "NEW Live Webcast" section. The page layout includes a search bar, navigation links, and a sidebar with "Downloads" and "Add this event to your calendar" options.

<http://royalsociety.org/events-Web-Science-Presentations.aspx>