Data, Government and the Open Web

Dr Leslie Carr,
Web Science Doctoral Training Centre & EPrints Open Access Software Centre
Web and Internet Science Research Group,
University of Southampton, UK
Introduction

• Computer Science lecturer
• Research interests: hypertext, knowledge management, open access, open data, open education, extended cognition, distributed information systems
• The Web!
• Open Source software supporting Open Access to Scientific and Scholarly Knowledge
  – EPrints
Web & Internet Science Research Group

• 20 faculty, 70 PhDs, 60 masters
• Masters courses in
  – Web Science
  – Web Technology
• Doctoral Training Centre
Why Web Science Matters

The Web has become an essential part of humanity's current and future economy, science and technology.

- Understanding the Web is a challenge as big as any other global cause.
Quiz: Where’s This?
Quiz: What’s This?
Remember These 2006 Headline?

Good news! Black hole won't destroy Earth
Fears raised collider would create black holes that could swallow planet

Worst Case: Collider Spawns Planet-Devouring Black Hole

World's Largest Supercollider Could Destroy the Universe
Birth of the Web

- From an underground nuclear bunker on the Swiss border, something did escape
- In 1989 the Web took over the academia, industry and the rest of the world
## Less successful Webs from history

<table>
<thead>
<tr>
<th>Sponsor</th>
<th>System</th>
<th>Scope</th>
<th>Real</th>
<th>Date</th>
<th>Important Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finance / Press</td>
<td>Reuters</td>
<td>Professional, centralised</td>
<td>✔</td>
<td>1850</td>
<td>News &amp; stock information (originally carrier pigeon and subsequently telegraph)</td>
</tr>
<tr>
<td>Private Institution</td>
<td>Mundaneum</td>
<td>Public, centralised</td>
<td>✔</td>
<td>1920</td>
<td>Based on indexing technology (the library card)</td>
</tr>
<tr>
<td>Military</td>
<td>Memex</td>
<td>Scholarly, individual, centralised</td>
<td>☒</td>
<td>1945</td>
<td>Aimed at Scientists and Technologists in WWII</td>
</tr>
<tr>
<td>Media</td>
<td>Xanadu</td>
<td>Public, decentralised</td>
<td>☒</td>
<td>1960</td>
<td>Focused on DRM, reuse and writing for “creatives”</td>
</tr>
<tr>
<td>Media</td>
<td>CEEFAX</td>
<td>Public, national, centralised</td>
<td>✔</td>
<td>1970</td>
<td>Broadcast, linked, not participatory</td>
</tr>
<tr>
<td>Government</td>
<td>Minitel</td>
<td>Public, national, centralised</td>
<td>✔</td>
<td>1980</td>
<td>Commercial services and information</td>
</tr>
<tr>
<td>Academy (CS &amp; HEP)</td>
<td>FTP / Archie / Anarchie</td>
<td>Public, decentralised</td>
<td>✔</td>
<td>1985</td>
<td>Downloaded resources (papers, reports) to hard drives and printed on LaserWriters.</td>
</tr>
<tr>
<td>Commerce</td>
<td>Hypercard, HyperTIES</td>
<td>Private, centralised</td>
<td>✔</td>
<td>1988</td>
<td>Personal applications, sometimes tied to multimedia resources on CDROMs / video disks</td>
</tr>
<tr>
<td>Academy (HEP)</td>
<td>WWW</td>
<td>Public, global, decentralised</td>
<td>✔</td>
<td>1990</td>
<td>Universal naming, linking, interoperability, participative. However no writing, no indexing.</td>
</tr>
<tr>
<td>Academy (CS)</td>
<td>Microcosm</td>
<td>Private, centralised</td>
<td>✔</td>
<td>1990</td>
<td>Sophisticated linking and openness for personal information stores</td>
</tr>
<tr>
<td>Academy (CS)</td>
<td>HyperG</td>
<td>Public, centralised</td>
<td>✔</td>
<td>1990</td>
<td>Extension of Web for with support for writing, indexing and consistency management.</td>
</tr>
<tr>
<td>Commerce</td>
<td>AOL, CompuServ</td>
<td>Public, centralised</td>
<td>✔</td>
<td>1990</td>
<td>Dialup access to email, forums, chat rooms and information resources</td>
</tr>
</tbody>
</table>
Society is Diverse. One Web Fits All?

The development of society as a whole (nuanced and structured and refined) is inextricably related to the technology of information provision, consumption and dissemination (e.g. writing, reading, printing, education). Different parts of society have different objectives and hence incompatible Web requirements, e.g. openness, security, transparency, privacy.
Expansion of the Web

• The Web spread the conditions of its initial creation throughout the whole of society as it underwent an initial inflationary phase.

• The academy
  – government patronage
  – large-scale co-operation
  – sharing of intellectual property
Openness Agendas

• Open Access to Research Outputs
  – Supported by all UK Research Councils
  – Heavily promoted by JISC

• Open Research Data
  – Demanded by government, private funders & UK Research Councils
  – Move from closed VRE / EScience environment

• Open Educational Resources
  – Move from closed VLEs to sharing resources
Open Research Publications

Institutional repository for the London School of Economics
http://eprints.lse.ac.uk/

One of 145 UK Institutional Research repositories
- Norway 47
- Sweden 58
- Denmark 9
- Finland 9
Open Scientific Data

A repository of chemical data JISC EBank project (2003)

Scientific, not bibliographic, metadata.

Run by the UK Crystallographic Service and latterly an international consortium.
Open Educational Resources

- JISC EdSpace Project
- IEEE LOM metadata
- Web 2, informally tagged
- Sufficient for discoverability and usefulness

JISC Jorum (full LOM) & JorumOpen (lightweight OA approach, DSpace)
Directions for Researchers

- Open Access, Open Data, Open Educational Resources
  - Web Technology joins the High Moral Ground
- vs the established economic model for ensuring continuity of information production
  - trading of privately held information through subscription products such as journals or magazines
Climate science must be more open, say MPs

By Roger Harrabin
Environment analyst, BBC News

MPs investigating the climate change row at the UK's University of East Anglia (UEA) have demanded greater transparency from climate scientists.

The Commons Science and Technology Committee criticised UEA authorities for failing to respond to requests for data from climate change sceptics.

"It is not standard practice in climate science to publish the raw data and the computer code in academic papers," its report said.

"We consider that climate scientists should take steps to make available all the data that support their work, including raw data... and full methodological workings, including the computer codes."

"When the prices to pay are so large, the knowledge on which these kinds of decisions are taken had better be right," the report stated.
Open Government Data

- Same principles work on the government
  - Transparency
  - Openness
  - Innovation
  - Enabling new value & new businesses
data.gov.uk (8,300 data sets)

Bespoke apps produced by independent developers for government data

Shared ideas: Michael built a 'planning proposals' web app. Some turned it into an RSS feed. Someone else turned that into a Twitter feed. Someone else used scraperwiki...
Over 8,300 datasets now
Frequent deposit activity of medium sized dataset deposits
Signs of a healthy data catalogue
- 406 departments supplying data
- These include national Government departments, local government, and local council

**Departments Depositing >1% of Total Datasets in data.gov.uk**

- Department of Health [11410]
- Department for Communities and Local Government [11401]
- Office for National Statistics [11606]
- Department for Environment Food and Rural Affairs [11403]
- Department for Children Schools and Families [11398]
- Home Office [11402]
- Ministry of Defence [11404]
- Department for Business Innovation and Skills [11399]
- Department for Work and Pensions [11409]
- Ministry of Justice [11400]
- NHS Information Centre for Health and Social Care [16121]
- Department for Transport [11406]
- Department for Culture Media and Sport [11419]
- Department for Education [11424]
- Warwickshire County Council [11428]
Data.gov.uk - Who’s Using It

- 197 Applications available on Data.gov.uk
- Ranging over multiple topics of interest
  - Transparency Apps
  - Transport Apps
  - Public Health Apps

![Number of Open Data Applications Published on data.gov.uk](chart)

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Apps</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>6</td>
</tr>
<tr>
<td>2010</td>
<td>102</td>
</tr>
<tr>
<td>2011</td>
<td>75</td>
</tr>
<tr>
<td>2012</td>
<td>9</td>
</tr>
</tbody>
</table>
Institutional Open Data

Open Data Service

Open Data Homepage

5 Star Data
Frequently Asked Questions
Apps
Data Catalogue
Places
Phonebook
Academic Programmes
Organisation
Jargon
Points of Service
Products & Services
Bus Routes
SPARQL Endpoint
Feedback
Suggestion
Report a Problem
Register an App

Datasets

Here is a list of datasets we currently have available. At this time, the list is subject to alteration because we are learning as we go. If you are going to rely on the data, please let us know.

- Apps using our data
- Buildings and Places
- Catering
- Common Learning Spaces Extra Information
- EPrints Repository
- EdShare
- EdShare Video
- Electronics and Computer Science EPrints Repository
- Events Diary
- International Links
- International Links DBpedia Data
- JACS Codes
- Links to DisabledGo Access Information
- Local Amenities
- Open Data Catalog
- Open Days July 2011
- Organisation
- Payments 2010-11 to 2011-01
- Photographer of University of Southampton Things
Institutional Apps

Application created by 4 students using the following datasets:

- Buildings and Places
- Catering
- Local Amenities
- Southampton bus-routes
- Southampton Bus-stops
- Teaching Room Features
- Ordnance Survey Data
How Does Open Data Work?

• Make catalogue of all the things you know
• Give each thing a permanent web name (URL)
  Publish information about that thing
  ★ on the Web with an open license
  ★★★ Use a machine readable format
  ★★★★ Use a standard data format
  ★★★★★ Use an open standard data format
  ★★★★★★ Mention other things using their web URLs
“Open” Platforms

Internet
“Open” Platforms

Internet

Web
“Open” Platforms

Internet

Web

Wiki  Facebook  Twitter  YouTube  Google
“Open” Platforms

Internet

Web

Wiki
Facebook
Twitter
YouTube
Google

Wikipedia
Social Marketing
Media Channel
Why Closed Knowledge Systems are Bad

Evaluation

Did Germany make the right choice in relocating their post-unification capital to Bonn?

Synthesis

Create a set of guidelines for choosing alternative national capitals in a post-climate change Europe.

Analysis

What were the overriding factors in the choice of sites for the European parliament and commission?

Application

Choose the most appropriate capital for the Isle of Wight

Comprehension

Why is Paris the capital of France?

Knowledge

What is the capital of France?

Bloom’s Educational Taxonomy demonstrates why closed knowledge systems are bad
Open Source, Systems, Data + Us = Open Society

• The Web isn't a thing but an activity
  – the creation of a network of information by a network of individuals.

• The Web wasn't invented by Tim Berners-Lee,
  – it is still being invented by all of us as we gradually adapt our tools and change our practices.