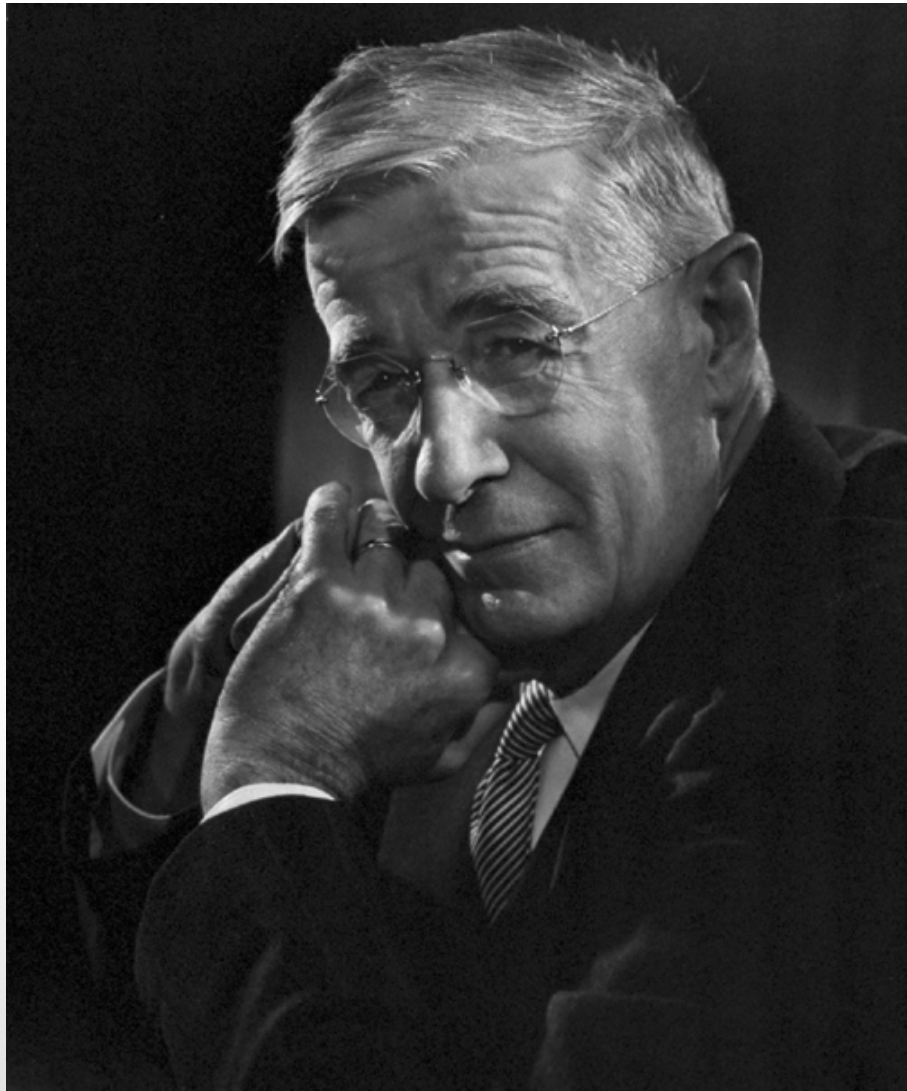


# The Emerging Science of the Web

And why it is important

Professor Dame Wendy Hall  
22 Jan 2010

# Inspiration



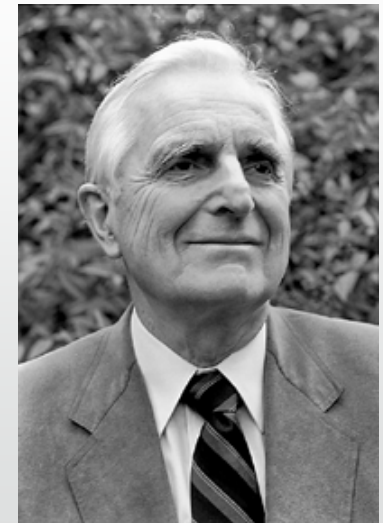
“As we may think”  
Vannevar Bush  
Atlantic Monthly  
July 1945



Everything is deeply intertwined

## Ted Nelson and Doug Engelbart

Augmenting human intellect





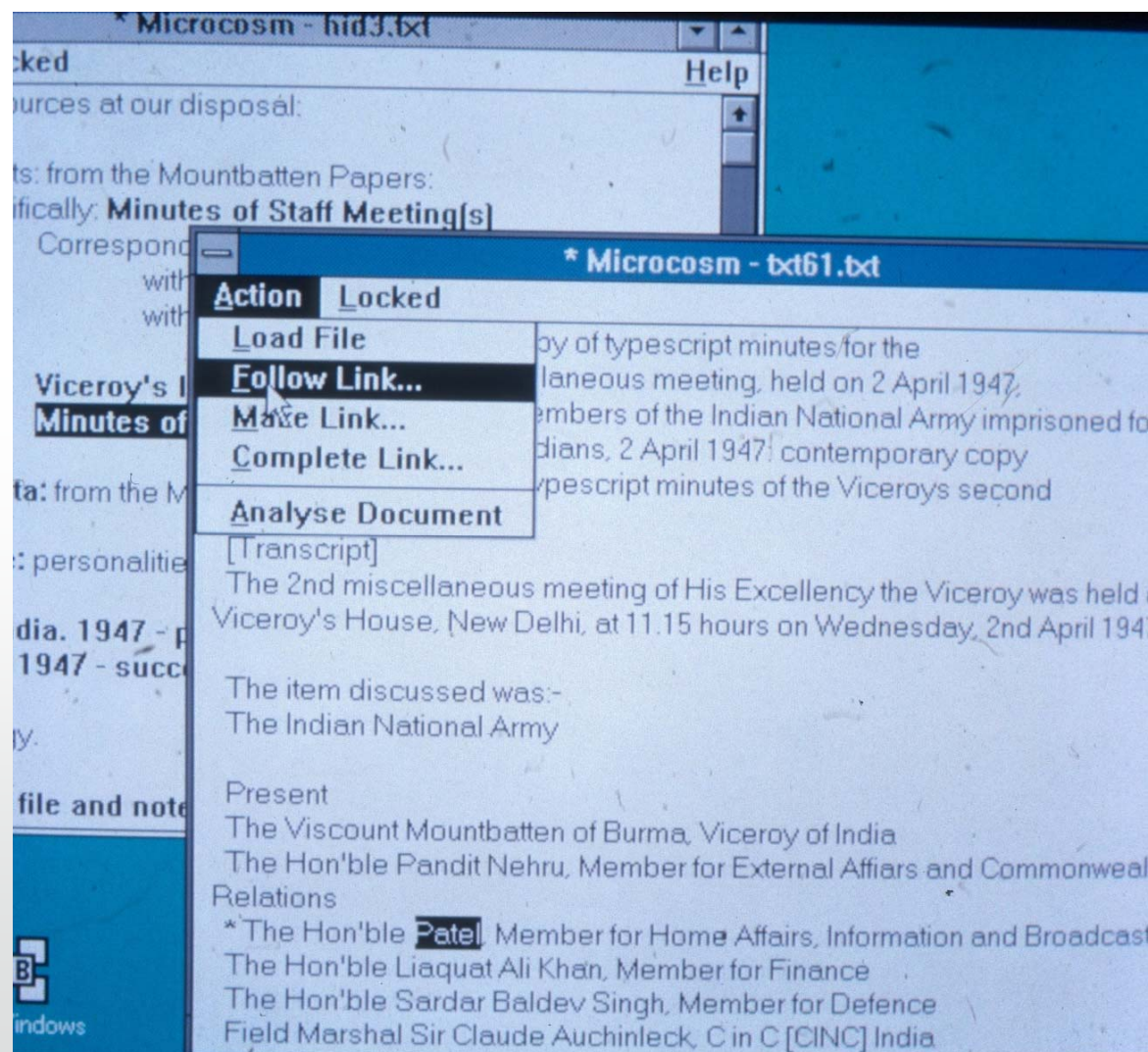
What do India, the Earl Mountbatten of Burma, the University of Southampton and my research career have in common?



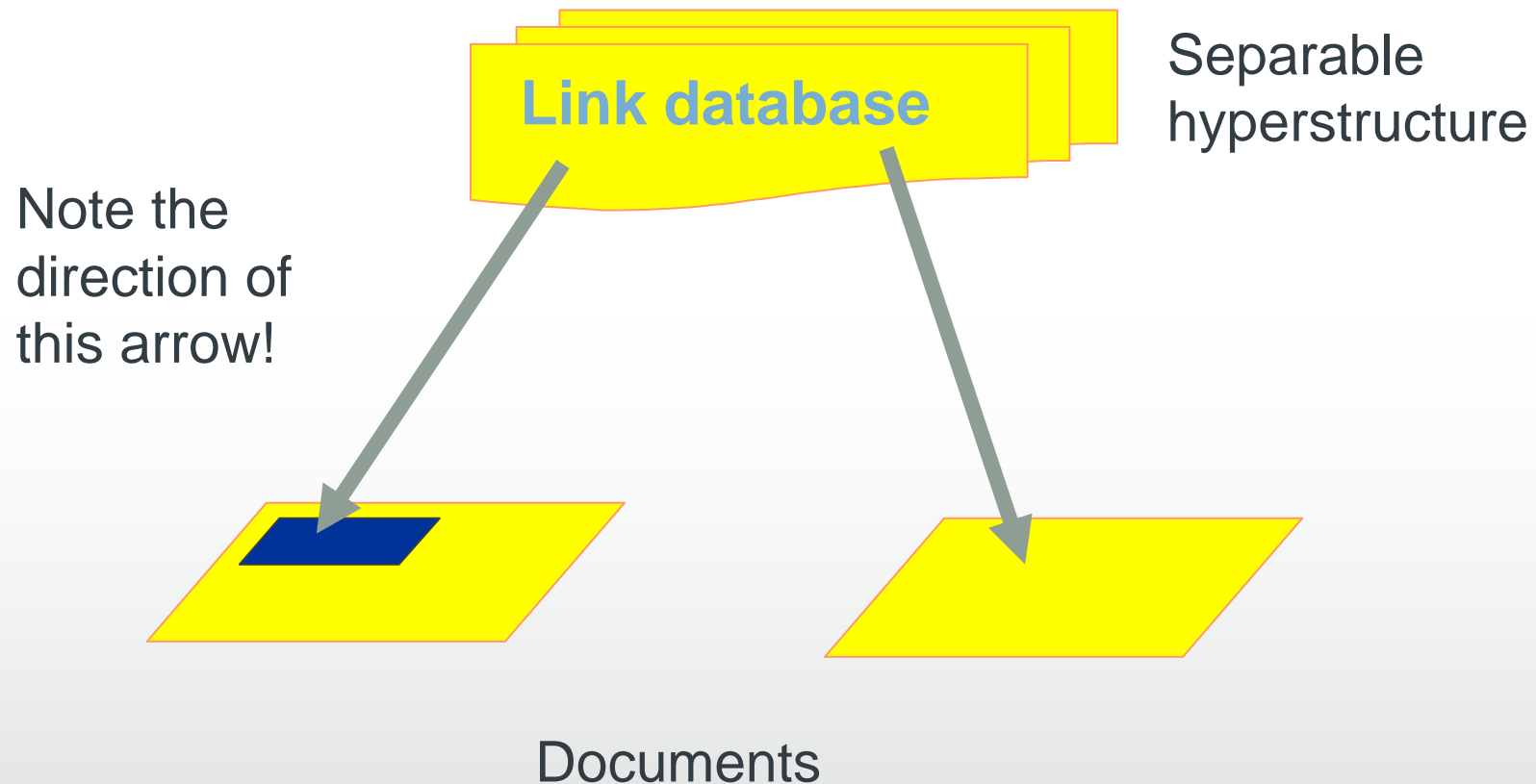
The Mountbatten archive  
moved to Southampton  
In 1987



## Microcosm: Mountbatten archive application



# Linkbases





# Links in Microcosm

- source, destination, description
- source: object | concept | context
- We generated links based on metadata description of documents in docuverse and “it all falls out”



ECHT'90



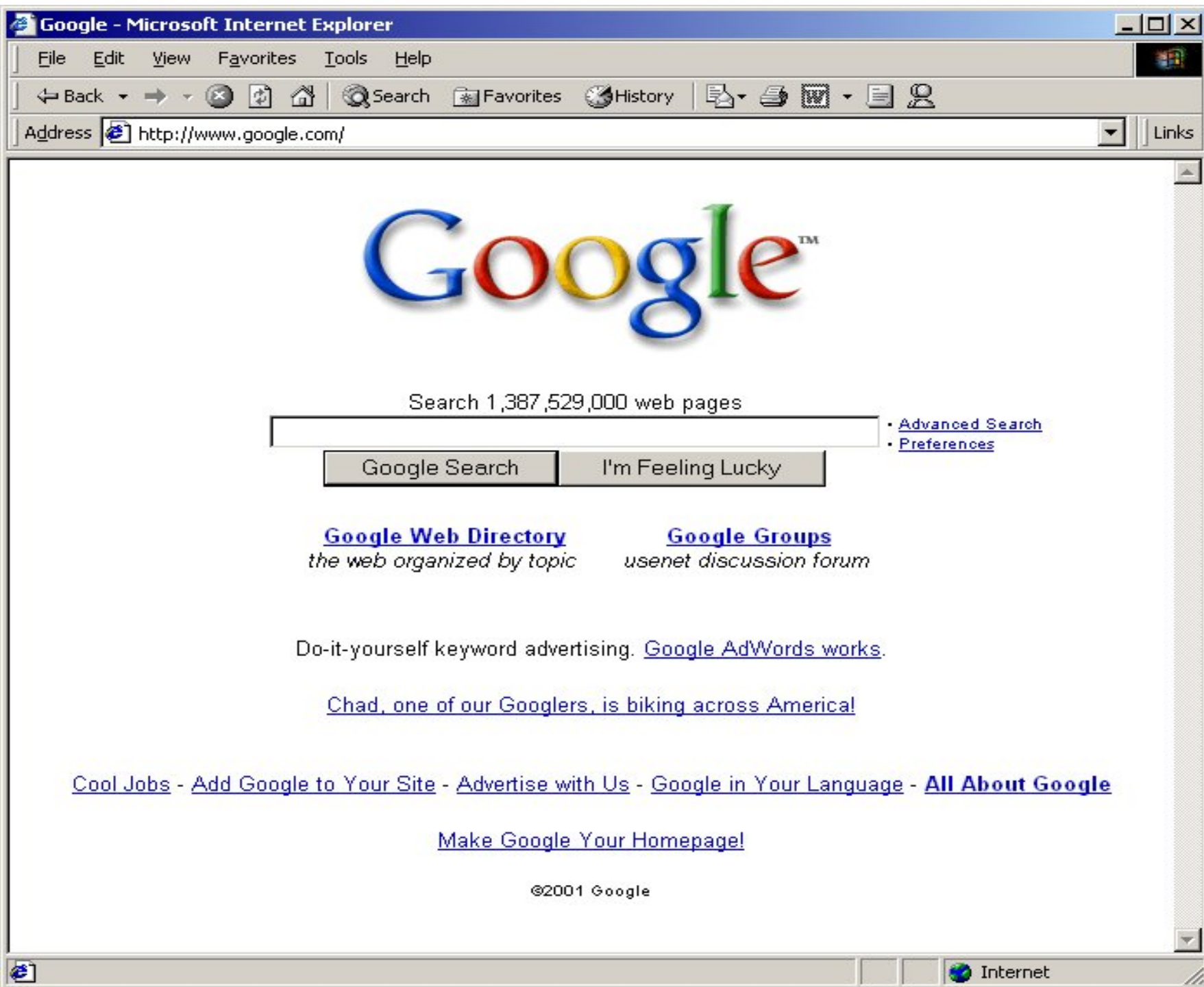


ACM Hypertext'91

# Lessons learnt

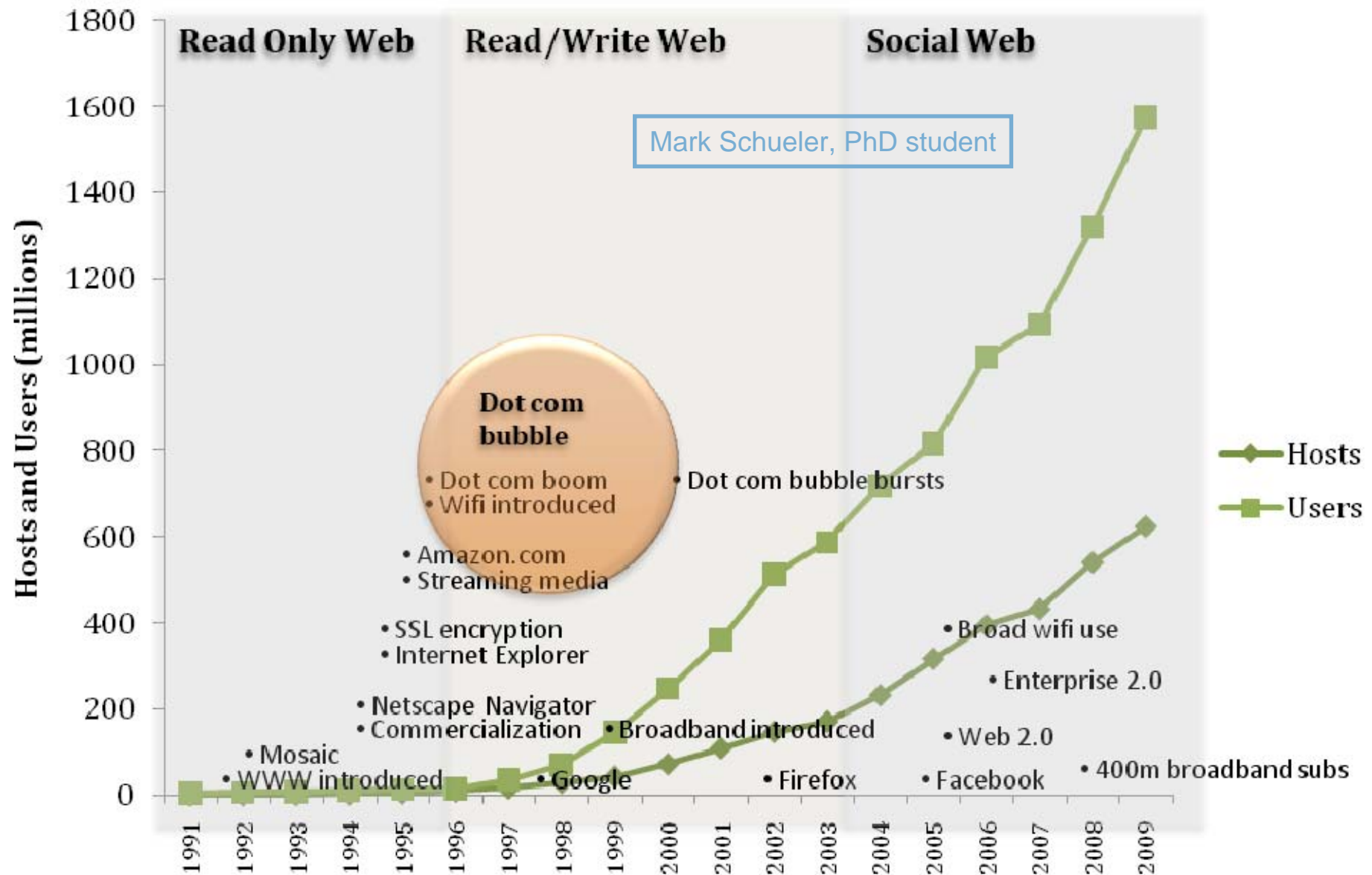
- Big is beautiful: the network is everything
- Scruffy works: let the links fail to make it scale
- Democracy rules: open and free
- But we lost (for a time) conceptual and contextual linking – the Web is a strangely linkless world
- Missing links – search engines fill the gap







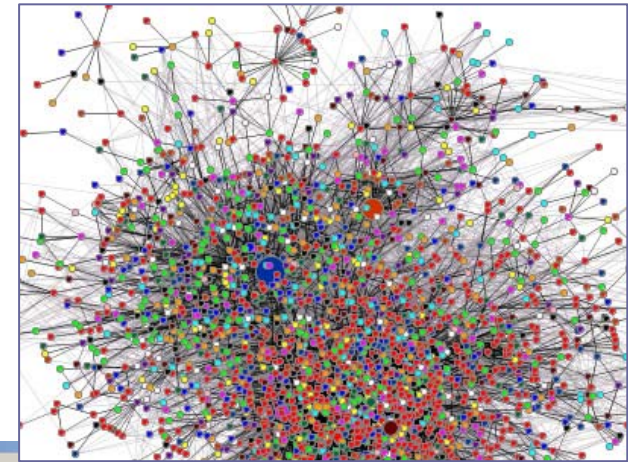
# Internet Growth - Usage Phases - Tech Events



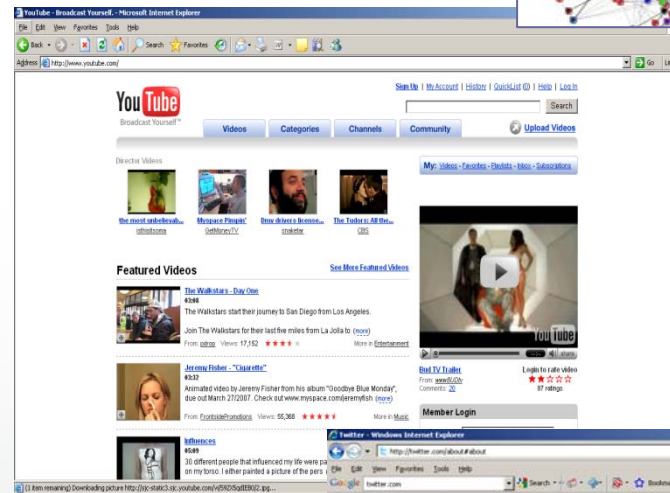
Note - events shown relate to the time axis only.

# Web 2.0

- Wiki's
- Blogs
- Flickr
- YouTube
- MySpace
- Facebook
- Second Life
- Twitter



Blogosphere



Wikipedia, the free encyclopedia - Windows Internet Explorer

W

http://en.wikipedia.org/wiki/Main\_Page

Google

File Edit View Favorites Tools Help

Google

Wikipedia

Search

+

Bookmarks

Check

Translate

AutoFill

Wikipedia


Sign In

Wikipedia, the free ency...

realtimecollisiondetection.ne...

Try Beta

Log in / create account



**WIKIPEDIA**  
The Free Encyclopedia

main page

discussion

view source

history

Wikipedia is there when you need it — now it needs you.

\$0.9M USD

\$7.5M USD

Donate Now

navigation

Main page

Contents

Featured content

Current events

Random article

search

Go

Search

interaction

About Wikipedia

Community portal

Recent changes

Contact Wikipedia

Donate to Wikipedia

Help

toolbox

What links here

Related changes

Upload file

Special pages

Printable version

Permanent link

Welcome to Wikipedia,

the free encyclopedia that anyone can edit.

3,102,741 articles in English

Arts

Biography

Geography

History

Mathematics

Science

Society


Technology

All portals

Overview · Editing · Questions · Help

Contents · Categories · Featured content · A–Z index

Today's featured article



**Grover Cleveland** (1837–1908) was both the 22nd and 24th President of the United States. Cleveland is the only President to serve two non-consecutive terms (1885–1889 and 1893–1897) and therefore is the only individual to be counted twice in the numbering of the presidents. He was the winner of the popular vote for President three times—in 1884, 1888, and 1892—and was the only Democrat elected to the Presidency in the era of Republican political domination that lasted from 1860 to 1912. Cleveland's admirers praise him for his honesty, independence, integrity, and commitment to the principles of classical liberalism. As a leader of the Bourbon Democrats, he opposed imperialism, taxes, subsidies and inflationary policies, but as a reformer he also worked against corruption, patronage, and bossism. Critics complained that he had little imagination and seemed overwhelmed by the nation's economic disasters—depressions and strikes—in his second term. Even so, his reputation for honesty and good character survived the troubles of his second term. [\(more...\)](#)

Recently featured: [Werner Mölders](#) – [Electron](#) – [Han Dynasty](#)

In the news

At least 104 people are killed in a [mine explosion](#) in Heilongjiang, People's Republic of China.

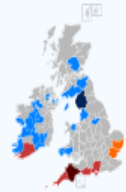
[Continued flooding](#) affecting Great Britain and Ireland (*affected areas pictured*) results in two deaths.

[Egypt recalls](#) its ambassador to Algeria in a dispute following a qualifying match between the nations for the 2010 FIFA World Cup.

[Prime Minister of Belgium Herman Van Rompuy](#) is chosen as the first permanent President of the European Council.

[British politician Catherine Ashton](#) is selected as the first High Representative of the European Union for Foreign Affairs and Security Policy.

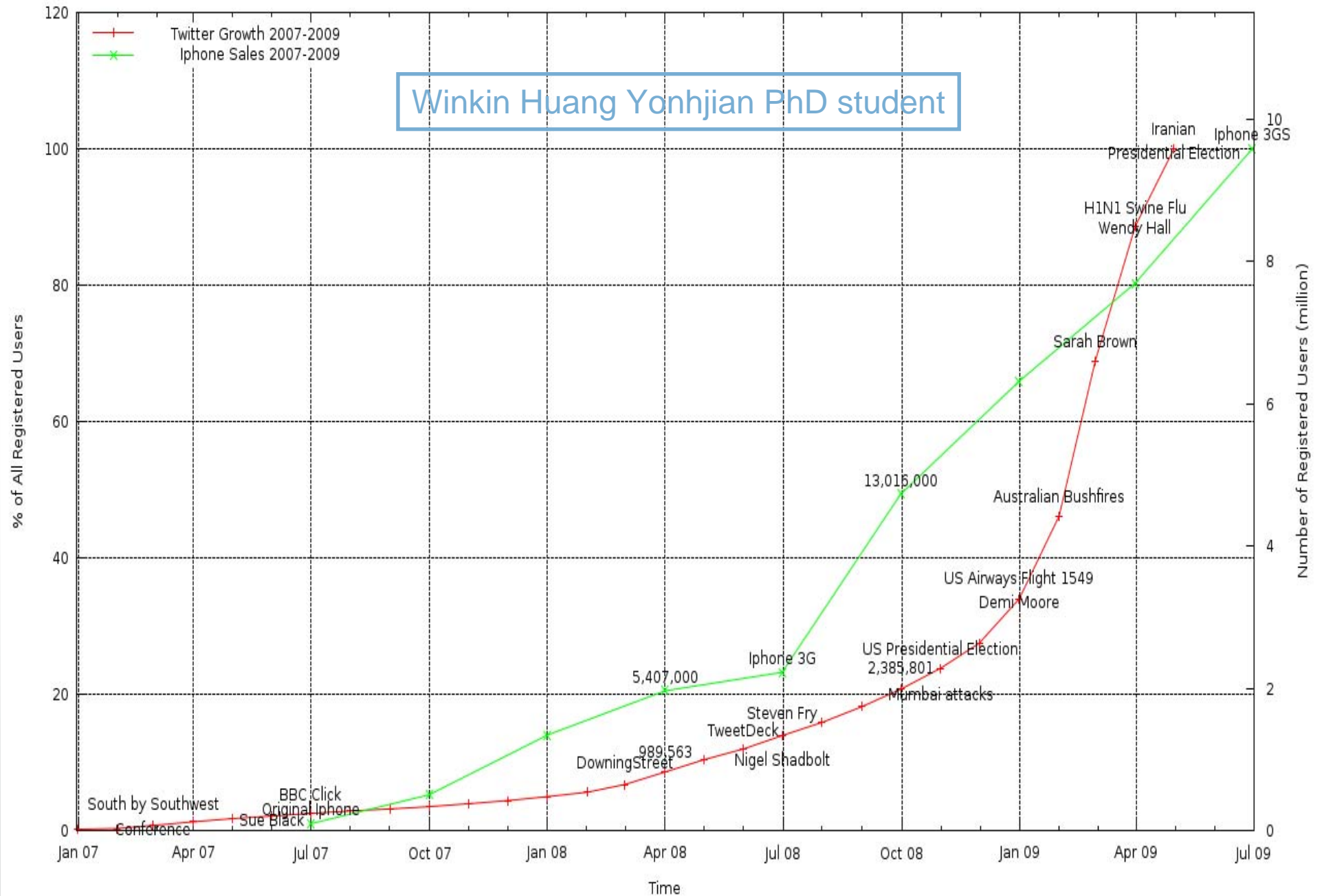
A United States district court [finds](#) the Army Corps of Engineers was negligent in maintaining [New Orleans levees](#) that contributed to flooding during Hurricane Katrina.



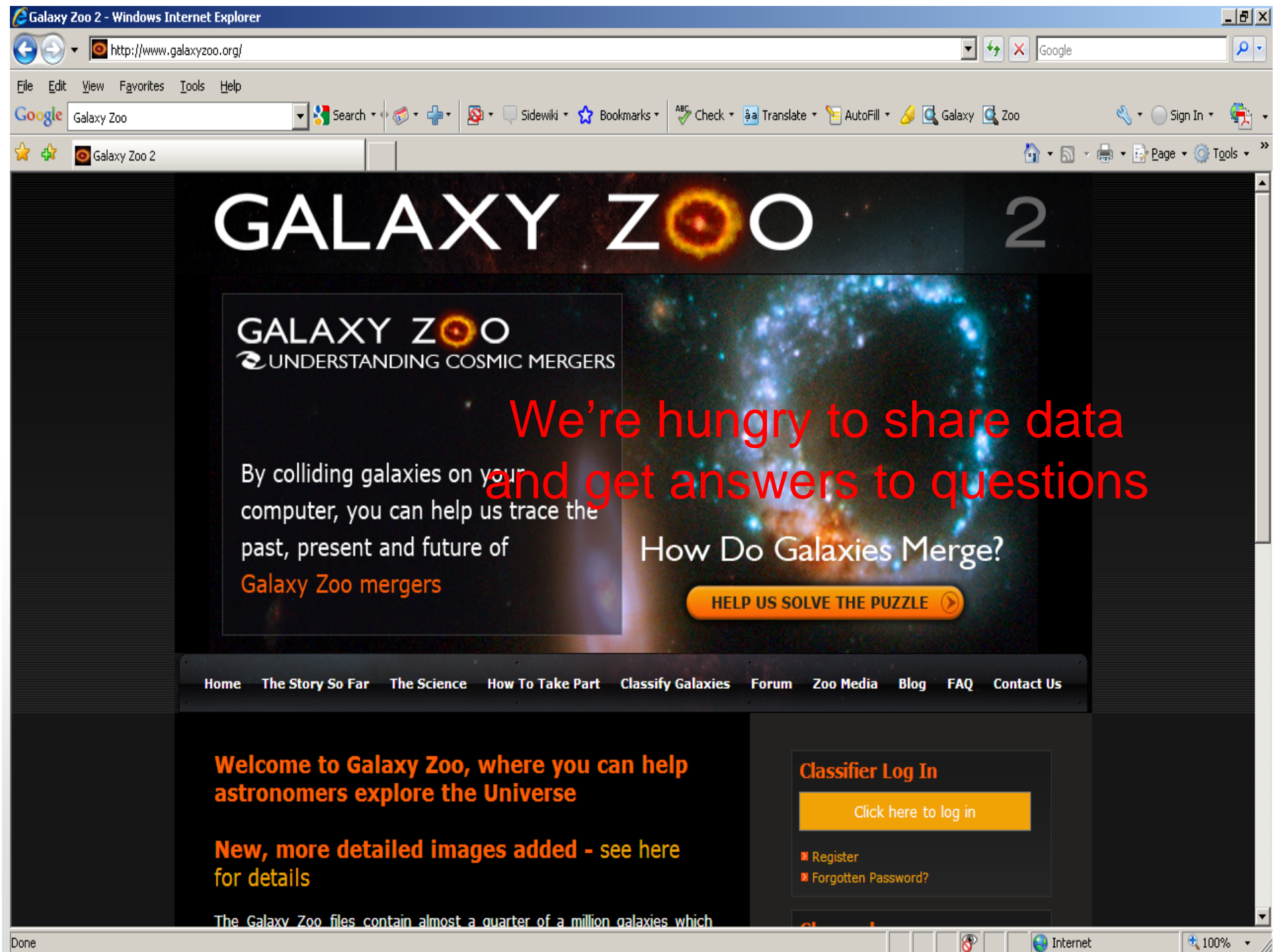
Internet

100%

Timeline of Twitter Growth





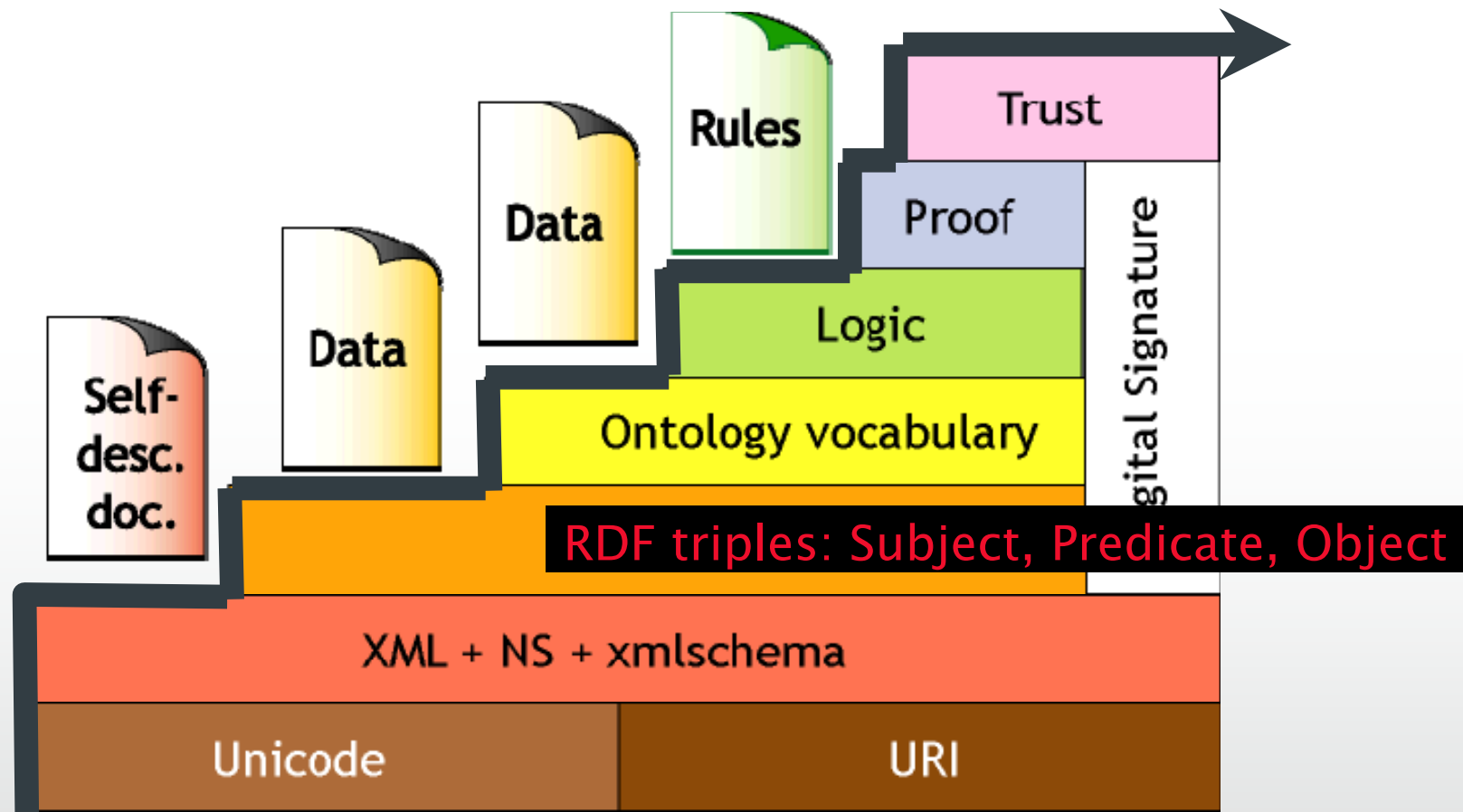


# The Semantic Web

## A Web of Data



# The Semantic Web



Semantic Web LayerCake (Berners-Lee, 99; Swartz-Hendler, 2001)

[Themes](#)[Speakers](#)[Talks](#)[Translations](#)[TED Conferences](#)[TEDx Events](#) **NEW**[TED Prize](#)[TED Fellows](#)[TED Community](#)[About TED](#)[TED Blog](#)

Search

## Talks Tim Berners-Lee on the next Web

Filmed Feb 2009; Posted Mar 2009

[Open interactive transcript »](#)

### About this talk

20 years ago, Tim Berners-Lee invented the World Wide Web. For his next project, he's building a web for open, linked data that could do for numbers what the Web did for words, pictures, video: unlock our data and reframe the way we use it together.

### About Tim Berners-Lee

Tim Berners-Lee invented the World Wide Web. He leads the World Wide Web Consortium, overseeing the Web's standards and development. [Full bio and more links](#)

### About our sponsor

A dive into the rich tradition of Rolex innovation, revealing unprecedented [watchmaking achievement »](#)



### What to watch next

# Content, Emergence and Unanticipated Reuse

The four micro principles of the Semantic Web

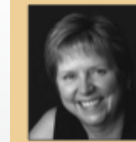
1. All entities of interest, such as information resources, real-world objects, and vocabulary terms should be identified by URI references.
2. URI references should be dereferenceable, meaning that an application can look up a URI over the HTTP protocol and retrieve RDF data about the identified resource.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



**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](mailto: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](mailto: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](mailto:wh@ecs.soton.ac.uk).

## The Semantic Web Revisited

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

ECS - Prof Wendy Hall - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Reload Home Search Favorites RSS Print Mail W Yellow Blue People

Address <http://www.ecs.soton.ac.uk/people/wh> Go Links >>

**ECS**  
Electronics & Computer Science  
University of Southampton

Site Search Enter keywords here go

Home UG Admissions Research PG Admissions People Alumni Contact Intranet

University of Southampton > ECS > People > Prof Wendy Hall

## Prof Wendy Hall

School of Electronics and Computer Science  
University of Southampton  
Southampton  
SO17 1BJ  
United Kingdom

**Position:** Academic staff in [Intelligence, Agents, Multimedia Group](#)  
**Extension:** 22748 22388  
**Telephone:** +44 (0)23 8059 2748 +44 (0)23 8059 2388  
**Homepage:** <http://www.ecs.soton.ac.uk/~wh/>  
**Email:** [wh@ecs.soton.ac.uk](mailto:wh@ecs.soton.ac.uk)  
**Interests:** [semantic web](#)  
**VCard:** <http://vcard.ecs.soton.ac.uk/person/1650>

### Seminars Given and Upcoming

- [Web Science Briefing](#) (Wednesday 20th September, 2006)

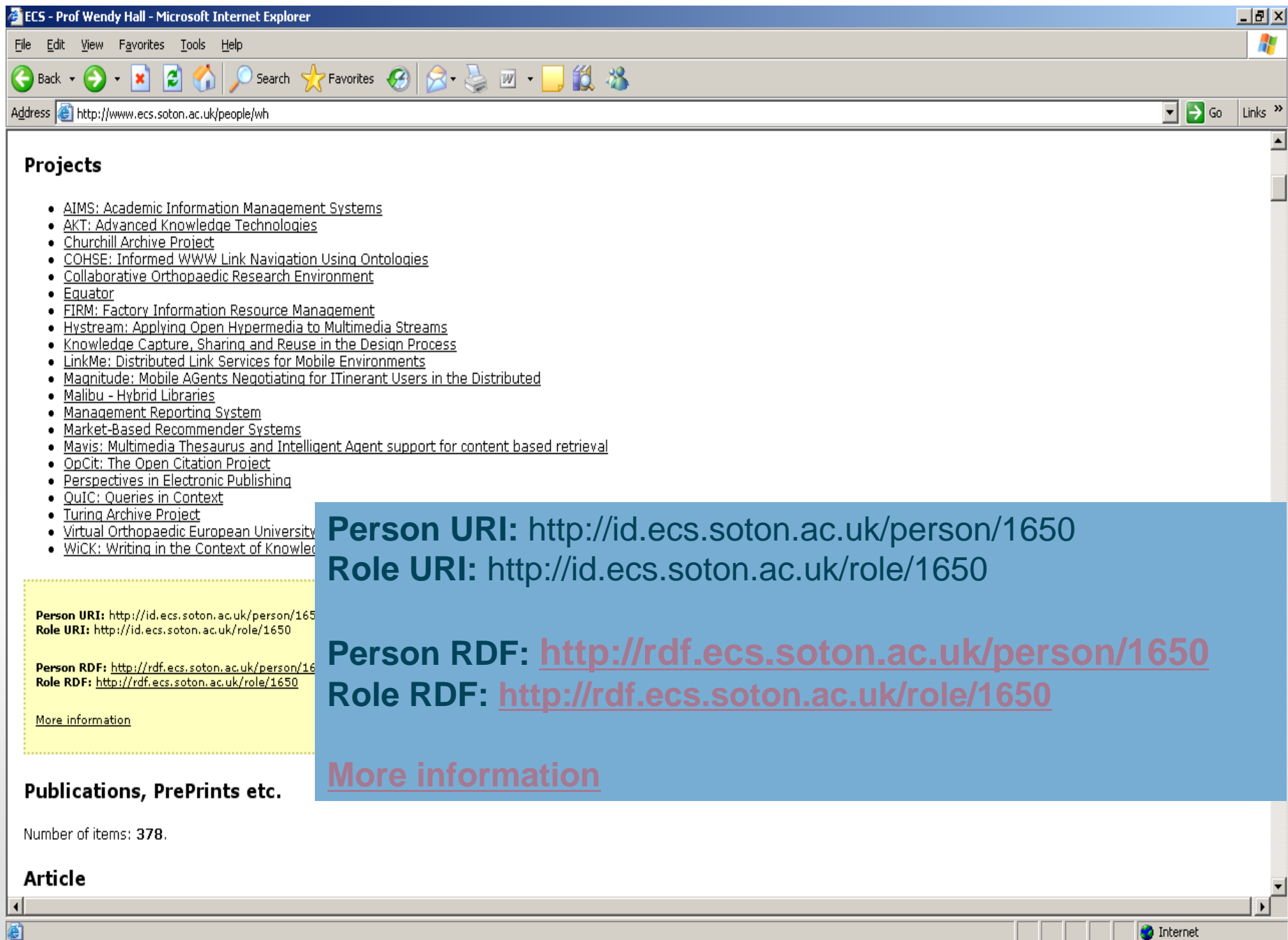
### Memberships

- Strategy Committee
- Finance Committee
- Strategy Committee
- Head of school
- Policy and Resource Committee

### Projects

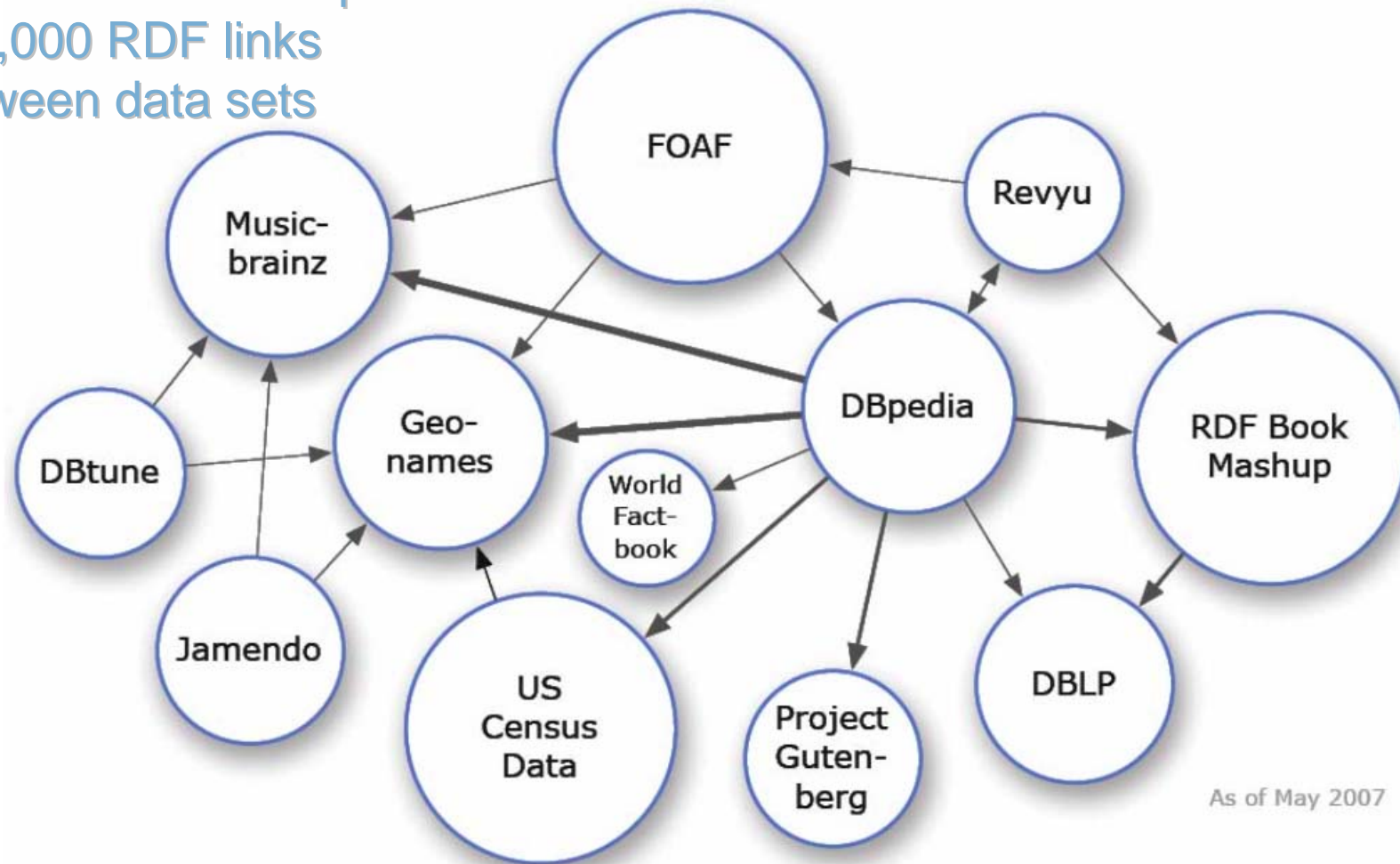
- [AIMS: Academic Information Management Systems](#)
- [AKT: Advanced Knowledge Technologies](#)

Internet



# Linked Data on the Web: May 2007

500 Million RDF Triples  
120,000 RDF links  
between data sets

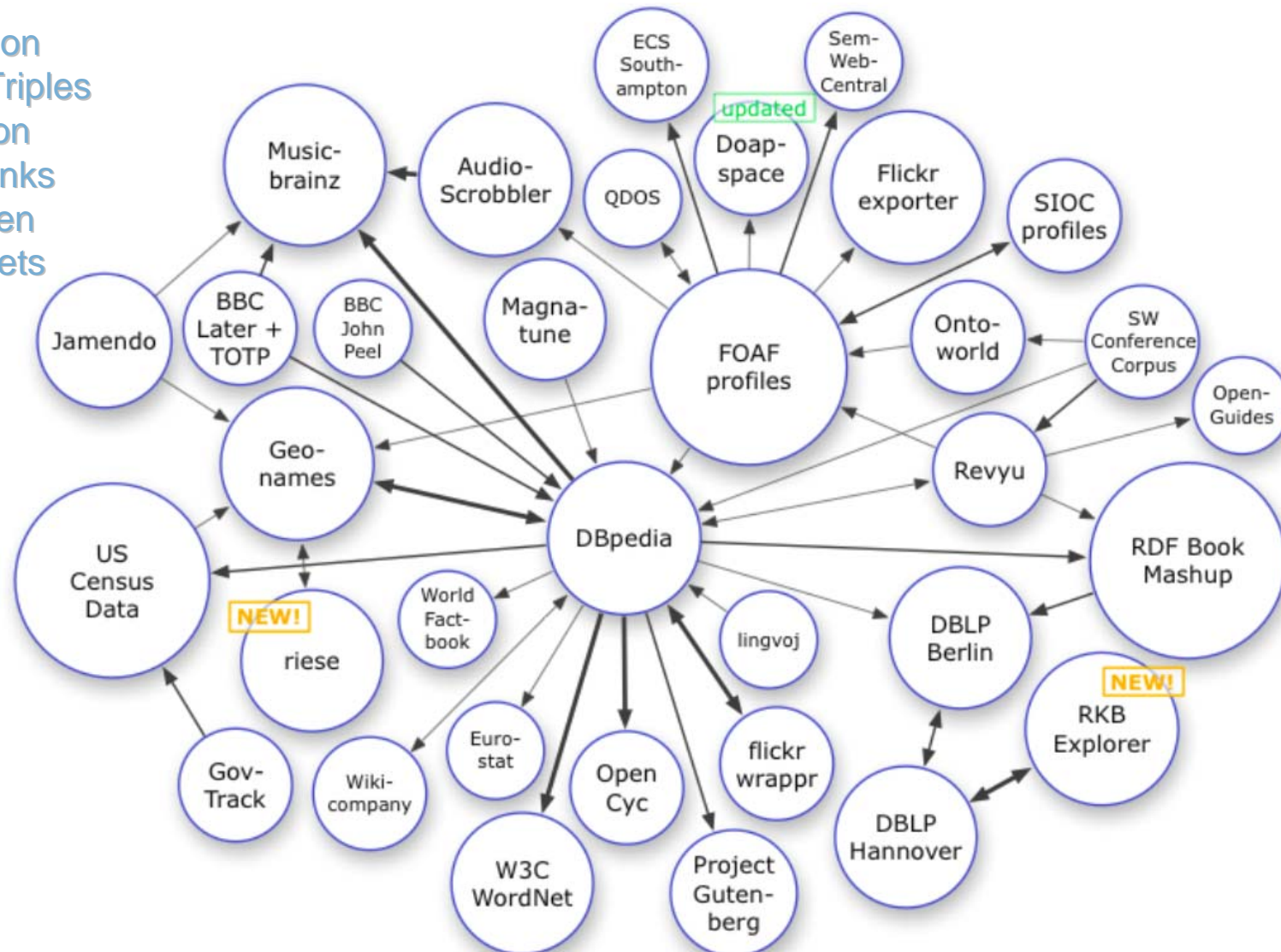


As of May 2007

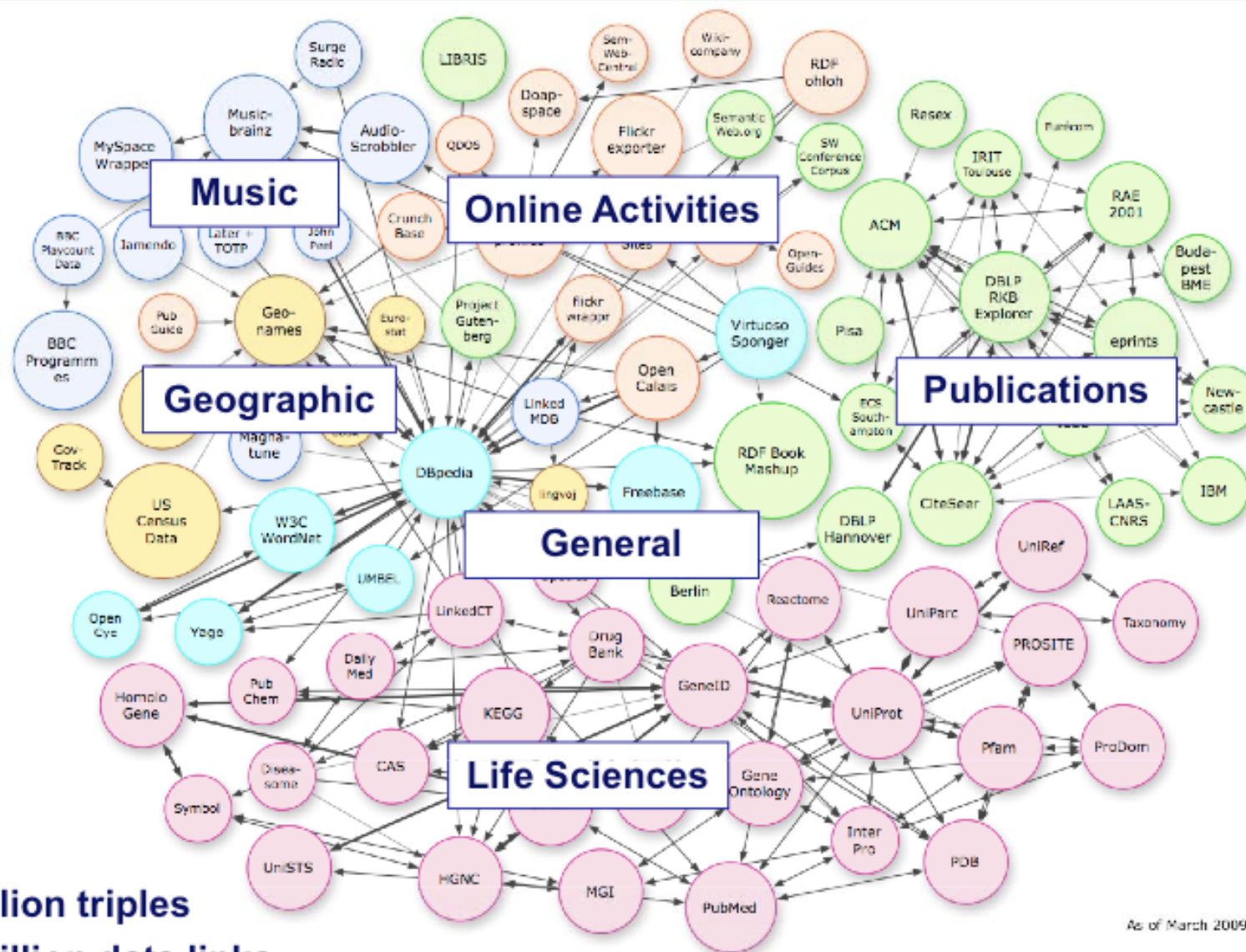


# Linked Data on the Web: April 2008

23 billion  
RDF Triples  
3 million  
RDF links  
between  
data sets



# LOD Datasets on the Web: March 2009



**4.5 billion triples**  
**180 million data links**

As of March 2009

Semantic Wikipedia Query: Tennis players from Moscow - Windows Internet Explorer

<http://wikipedia.aksw.org/index.php?qid=13>

Google

File Edit View Favorites Tools Help

Google

Go

19 blocked

Check

AutoLink

AutoFill

Send to

Logitech

Y!

Web Search

Anti-Spy

Bookmarks

Mail

My Yahoo!

Answers

W.wiki.dbpedia.org : Online Ac...

Semantic Wikipedia Quer...

Home

Print

Page

Tools

Query Wikipedia

This semantic database contains over 10 million statements extracted from the English Wikipedia.

search for queries

[Most popular](#)
[Upcoming](#)

[Tennis players from Moscow](#)
[People influenced by Friedrich Nietzsche](#)
[Sitcoms set in NYC](#)
[Soccer player with tricot nr. 11, playing for a club having a stadium with >40.000 seats, born in a country with >10M inhabitants](#)
[Pictures of Clint Eastwood's Movies](#)
[Films longer than 5 hours](#)
[Space Missions](#)
[Film music composer born 1965](#)
[People being 1.80m tall](#)
[What connects Innsbruck and Leipzig](#)
[List of Web browser software](#)
[Mayors of US cities higher than 1000m](#)
[Pictures of American guitarists](#)
[Scientists and their doctoral advisors](#)
[Battles in Saxony](#)

<< 1 >>

More information: at dbpedia.org and in



Tennis players from Moscow

[Modify this query or create your own!](#)

Results: 10

Click on a column header to sort results on this page.

11 results found in 0.005s

Nr.	?person	?image	~Moscow	~.*tennis_players\$
1	<a href="#">Igor Andreev</a>	 	<a href="#">Moscow, Russia</a>	<a href="#">Category:Russian tennis players</a>

Done, but with errors on page.

Internet

100%





Thanks to Hugh Glaser and Ian Millard

www.rkbexplorer.com/explorer/

people research areas publications **projects** search

recently viewed reset help

### Resilience for Survivability in IST

High-density mammos disc in blue wavelen...  
 Predictably Dependable Computing Systems  
 NEXT TTA - High-Confidence Architecture ...  
 Network of Excellence in distributed and...  
 Assaying Means of Design Expression f...  
 Theory and applications of continuous in...  
 Basic Research on Advanced Distributed C...  
 Network of Excellence in Distributed C...  
 New chromate-free passivation treatments...  
 Malicious- and Accidental-Fault Toleranc...  
 Design of an Open Dependable Accompanying Measure System Deper...  
 Dependable Systems Of Systems  
 The esprit network of excellence in dist...  
 Time-triggered architecture  
 Design for validation

Maximize

21 / 20

### Detail

**Title:**  
Resilience for Survivability in IST

**Other Names:**  
ReSIST Resilience for Survivability in IST

**Funding:**  
4500000

**Project Leader:**  
Jean-Claude Laprie

**Leading Organisation:**  
Centre National de la recherche Scientifique

### People

Jean-Claude Laprie

Aad van Moorsel

Abdelmajid Khellil

Achour Mostefaou

Adnan Noon Mian

Afriz Jaffri

Agnan de Bonneval

Alberto Pasquini

### Research Areas

Information processing, information systems

Development Failure

Dependability And Security Analysis

Dependability And Security Provision

### Publications

Revised version in Predictably dependable computing systems

ASSERT Automated proof based system and software engineering for real-time applications

Guide de la sûreté de fonctionnement

### Projects

Malicious- and Accidental-Fault Tolerance for Internet Applications

The esprit network of excellence in distributed computing systems architectures

Design for validation

Network of Excellence in

http://www.rkbexplorer.com/sameAs/				
http://www.rkbexplorer.com/sameAs/?uri=http%3A%2F%2Fsouthampton.rkbexplorer.com%2Fid%2Fperson-~ Google				
i-Metro eprints3 iPhone SA Single (74) And Benny Uni DH ECS UoS ReSIST MacOS X Football P SWC GPS Stately Homes SS Add To Twine AKT Shopping >>				
http://www.rkbexplorer.com/sameAs/?uri=http%3A%2F%2Fsouthampton.rkbexplorer.com%2Fid%2Fperson-~ 9cdf8d665f2fcc20fc5...				
Wendy Hall	akt:has-affiliation	http://acm.rkbexplorer.com/id/org-e9f6460843767ee5386b100d61bfd7d	acm-periodicals.rdf >>	
Wendy Hall	akt:has-affiliation	http://acm.rkbexplorer.com/id/org-fb53de9717078ad0b24ca8b5b5a40b1f	acm-proceedings.rdf >>	
Wendy Hall	akt:has-affiliation	University of Southampton	southampton-people.rdf	
Wendy Hall	akt:has-appellation	akt:Prof	southampton-people.rdf	
Wendy Hall	has email address	wh@ecs.soton.ac.uk [Email]	southampton-people.rdf	
Wendy Hall	akt:has-postal-address	School of Electronics and Computer Science, University of Southampton, SO17 1BJ, United Kingdom	southampton-people.rdf	
Wendy Hall	akt:has-telephone-number	+44 (0)23 8059 2388	southampton-people.rdf	
Wendy Hall	kisti:engNameOfPerson	Wendy Hall [Explore]	datatypeproperties.ttl >>	
Wendy Hall	kisti:engNameOfPerson	Wendy Hall [Explore]	datatypeproperties.ttl >>	
Wendy Hall"@e	rdf:type	academic of the University of Southampton	Wendy_Hall >>	
Wendy Hall"@e	rdf:type	commander of the Order of the British Empire	Wendy_Hall >>	
Wendy Hall"@e	rdf:type	scientist	Wendy_Hall >>	
Wendy Hall"@e	rdf:type	http://dbpedia.org/class/yago/FellowsOfTheBritishComputerSociety	Wendy_Hall >>	
Wendy Hall"@e	rdf:type	fellow of the Institution of Engineering and Technology	Wendy_Hall >>	
Wendy Hall"@e	rdf:type	fellow of the Royal Academy of Engineering	Wendy_Hall >>	
Wendy Hall"@e	rdf:type	person	Wendy_Hall >>	
Wendy Hall"@e	rdf:type	scientist	Wendy_Hall >>	
Wendy Hall	rdf:type	akt:Academic	southampton-people.rdf	
Wendy Hall	rdf:type	akt:Affiliated-Person	acm-bookchapters.rdf >	
Wendy Hall	rdf:type	akt:Affiliated-Person	acm-periodicals.rdf >>	
Wendy Hall	rdf:type	akt:Affiliated-Person	acm-proceedings.rdf >>	
Wendy Hall	rdf:type	akt:Affiliated-Person	cs_metadata_1.rdf >>	
Wendy Hall	rdf:type	akt:Affiliated-Person	cs_metadata_2.rdf >>	
Wendy Hall	rdf:type	akt:Affiliated-Person	cs_metadata_2.rdf >>	
Wendy Hallz	rdf:type	akt:Affiliated-Person	cs_metadata_2.rdf >>	
Wendy Hall	rdf:type	akt:Affiliated-Person	cs_metadata_3.rdf >>	
And Wendy Hall	rdf:type	akt:Affiliated-Person	cs_metadata_3.rdf >>	
Wendy Hall	rdf:type	akt:Affiliated-Person	cs_metadata_3.rdf >>	
Wendy Hall	rdf:type	akt:Affiliated-Person	cs_metadata_4.rdf >>	
Wendy Hall	rdf:type	akt:Affiliated-Person	cs_metadata_4.rdf >>	
Wendy Hall	rdf:type	akt:Affiliated-Person	cs_metadata_4.rdf >>	



http://www.rkbexplorer.com/sameAs/			
http://www.rkbexplorer.com/sameAs/?uri=http%3A%2F%2Fsouthampton.rkbexplorer.com%2Fid%2Fperson-...			
i-MetrO eprints3 iPhone SA Single (74) And Benny Uni DH ECS UoS ReSIST MacOS X Football P SWC GPS Stately Homes SS Add To Twine AKT Shopping			
http://www.rkbexplorer.com/sameAs/?uri=http%3A%2F%2Fsouthampton.rkbexplorer.com%2Fid%2Fperson-...			
Subject	Property	Object	Source
Association for Computing Machinery"@p	leader name	Wendy Hall"@e	Association_for_Compu
Association for Computing Machinery"@p	leader name	Wendy Hall"@e	Association_for_Compu
http://dbpedia.org/resource/1970_Wimbledon_Championships_-_Women%27s_Singles	rd1 team06	Wendy Hall"@e	1970_Wimbledon_Charr _Women%27s_Singles
Wimbledon 1969 (vrouwen)"@n	rd1 team12	Wendy Hall"@e	1969_Wimbledon_Charr _Women%27s_Singles
Wimbledon 1969 (vrouwen)"@n	rd1 team12	Wendy Hall"@e	1969_Wimbledon_Charr _Women%27s_Singles
Wimbledon 1968 (vrouwen)"@n	rd1 team14	Wendy Hall"@e	1968_Wimbledon_Charr _Women%27s_Singles
Wimbledon 1968 (vrouwen)"@n	rd1 team14	Wendy Hall"@e	1968_Wimbledon_Charr _Women%27s_Singles
1967 Wimbledon Championships - Women's Singles"@e	http://dbpedia.org/property/rd2Team02	Wendy Hall"@e	1967_Wimbledon_Charr _Women%27s_Singles
1967 Wimbledon Championships - Women's Singles"@e	http://dbpedia.org/property/rd2Team02	Wendy Hall"@e	1967_Wimbledon_Charr _Women%27s_Singles
A Portable Multimedia Communicator Scheme	akt:edited-by	Wendy Hall	southampton-publication
Multimedia Technologies and Future Applications	akt:edited-by	Wendy Hall	southampton-publication
Proceedings of the Fourth ACM International Multimedia Conference, Multimedia '96	akt:edited-by	Wendy Hall	southampton-publication
Media integration issues in open hypermedia systems	akt:edited-by	Wendy Hall	southampton-publication
Extending the role of the digital library	akt:has-author	Wendy Hall	acm-proceedings.rdf >>
WICKEd	akt:has-author	Wendy Hall	acm-proceedings.rdf >>
When open hypermedia meets peer-to-peer computing	akt:has-author	Wendy Hall	acm-proceedings.rdf >>
Assisting artifact retrieval in software engineering projects	akt:has-author	Wendy Hall	acm-proceedings.rdf >>
The case for explicit knowledge in documents	akt:has-author	Wendy Hall	acm-proceedings.rdf >>
An implementation of Prolog for the Inmos T800 transputer	akt:has-author	Wendy Hall	acm-proceedings.rdf >>
Applications of hypermedia in education	akt:has-author	Wendy Hall	acm-periodicals.rdf >>
Applications of hypermedia in education	akt:has-author	Wendy Hall	acm-proceedings.rdf >>
MICROCOSM: an open model for hypermedia with dynamic linking	akt:has-author	Wendy Hall	acm-bookchapters.rdf >>



# Co-Reference

- Repositories have people, publications, etc. from other institutions who also have records there and elsewhere
- And vice versa
- Co-Reference is a Big Problem (the ACM has done a lot of work on this)
  - Everything is a URI (not title, name, number...)
  - Identifying multiple URIs for one resource
  - Rejecting incorrectly conflated resources
  - Publishing
  - Using
- Coldstart
  - A serious problem
  - Nothing is linked to anything
  - Not even (reliably) within most repositories



# Where is it Taught?



# Tipping points?

**David Peterson**

David Peterson has been a web developer since the early years - 1995. He works in the [steamy tropics](#) and is trying to kick start the Semantic Web down under. Not only that, but he is an Advisory Committee representative to the W3C, wow. His wonderful family, making lovely photographs and searching for the perfect espresso keeps him happy.

Subscribe

**SitePoint Newsletters**

Get all the latest tips by signing up to all of SitePoint's informative newsletters.

- ☒ SitePoint TechTimes
- ☒ SitePoint Tribune
- ☒ SitePoint Design View
- ☒ SitePoint Market Watch
- ☒ Community Crier

Subscribe!

☒ HTML ☐ Plain

- [Our Privacy Policy](#)
- [Newsletter Archives](#)

**Does your system cater for SMS alerts?**

SMS enable your system alerts and enjoy the mobility of push technology, simple set-up and rapid integration.

SMS alerts can **reach over 712 mobile networks** using a range of API connections including HTTP/S, FTP, SMPP, SMTP, XML, Com Object & SOAP connections.

**FIND OUT HOW TO SMS ENABLE YOUR SYSTEM IN 5 EASY STEPS:**

[Download the free developer guide now.](#)

Clickatell

Any message, anywhere.

Download FREE Developer Guide



Blog Post

[Blogs](#) » [Web Tech](#) » [Obama's Groundbreaking use of the Semantic Web](#)
[« Microsoft SuperPreview: a New Way to Test Web Sites](#)
[Open Source Image Editors ... for Designers? »](#)

## Obama's Groundbreaking use of the Semantic Web

by [David Peterson](#)

In a revolutionary move, Obama's administration is set to utilise next generation web technologies to bring an unprecedented level of transparency to government. In this case it will shed light on how the roughly US \$800 billion dollar economic stimulus will be spent. The recently launched [recovery.gov](#) website (powered by nothing other than Drupal) brought with it the promise that citizens would be able to view where the money was going and how it was going to be spent.

To enable the citizen masher to do their wizardry, the administration will be opening up a veritable candy store of goodies: [Semantic Web](#), [RDF](#), [Linked Data](#), [SPARQL](#), [RDFa](#), [SIOC](#), [ATOM](#), [RESTful](#) APIs, JSON, Widgets, [Wikis](#), XForms, P2P Networks. Wow. They only forgot the lions and tigers and bears oh my... This is an unbelievable stack of technology. I didn't think the government even knew what an RSS feed was :)



## News

## Webchats

M	T	W	T	F	S	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

« Sep

## PM welcomes Sir Tim Berners-Lee to Downing Street

The Prime Minister welcomed the creator of the World Wide Web, Sir Tim Berners-Lee, and Professor of Artificial Intelligence at the University of Southampton, Nigel Shadbolt, to Downing Street this morning.



Mr Berners-Lee and Mr Shadbolt presented an update to Cabinet on their work advising the Government on how to make data more accessible to the public.

Gordon Brown has already spoken publicly about his aim of making the UK a world leader in opening up government information on the internet, an important element of [Building Britain's Future](#).

## Newsletter

Sign up to our newsletter to keep updated with the latest information from Number 10

[Click here to subscribe](#)

## Around the Web

**flickr**

## Latest Photos

[View all images](#)

# data.gov.uk

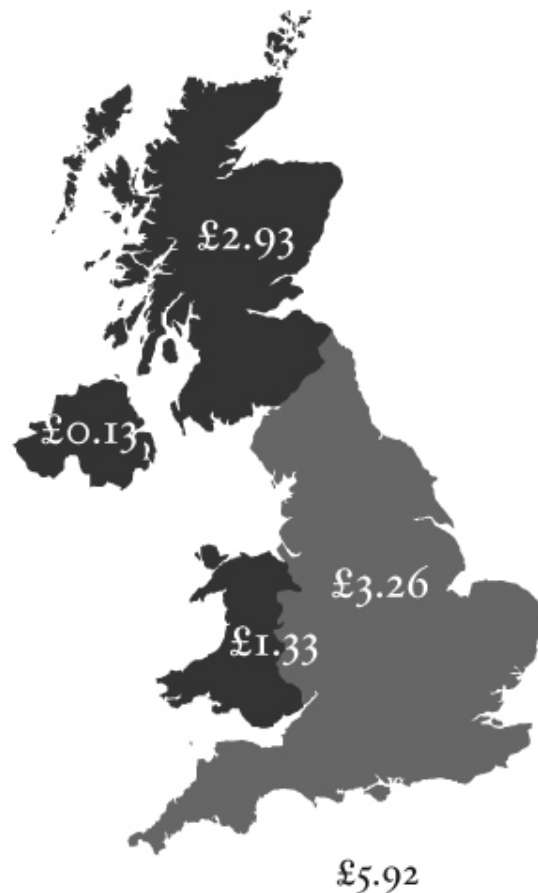
- Currently under development scrutiny and extension
  - Over 1100 datasets
  - Over 1000 registered developers
  - The beginning of linked data
- Beta service to be released January 2010
- Access to OS from April 2010
- Clear schedule of release from Gov Depts thereafter



# £ Massive Outgoings

Cost for average British taxpayer per day

## GOVERNMENTS



## ESSENTIALS

NHS



£9.05

Schools



£2.56

Defence



£3.99

## TRAVEL

Roads



£3.10

LRT



£0.14

## BANKING CHARGES

Recapitalisations



£9.23

## SECURITY

Police



£0.53

Prisons



£0.19

Afghanistan



£0.32

## GOING OUT

Arts



£0.04

Museums



£0.03

## HELPING OTHERS

Child support



£1.86

Job Seekers



£0.25

Housing benefit



£1.52

Disability benefit



£1.44

Pensions



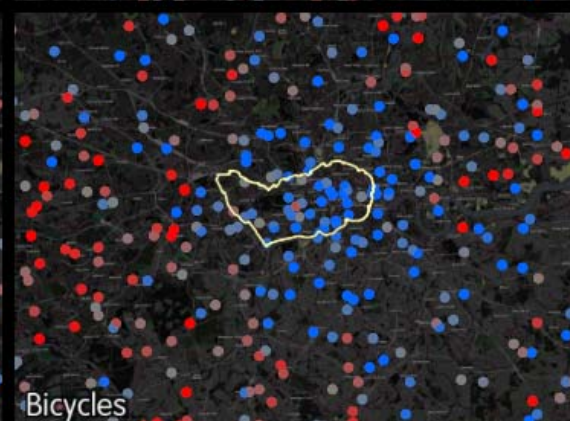
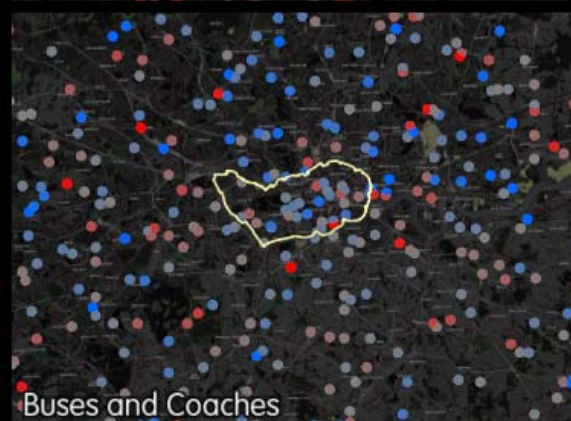
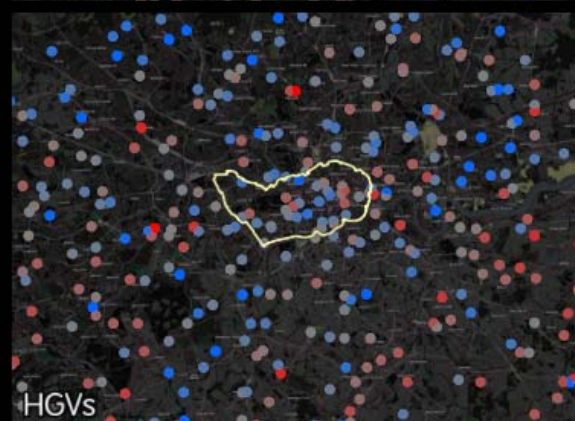
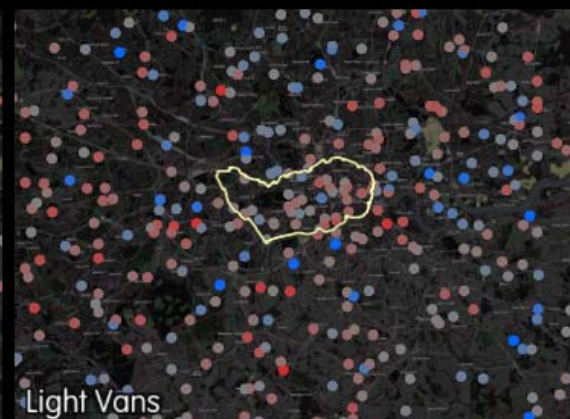
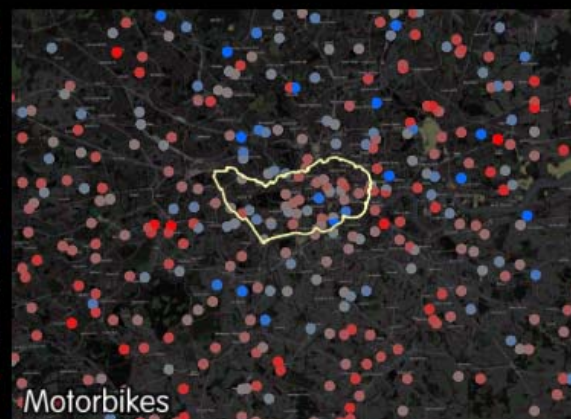
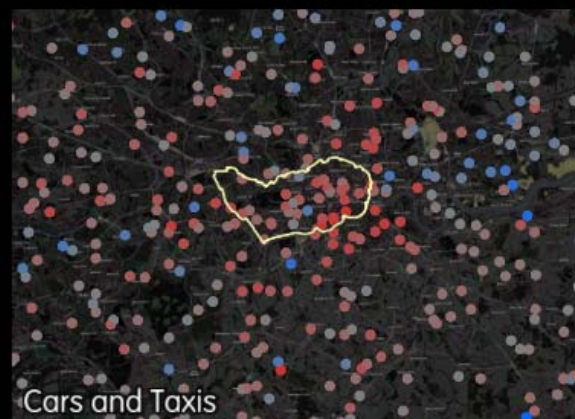
£5.55

Overseas aid



£0.43

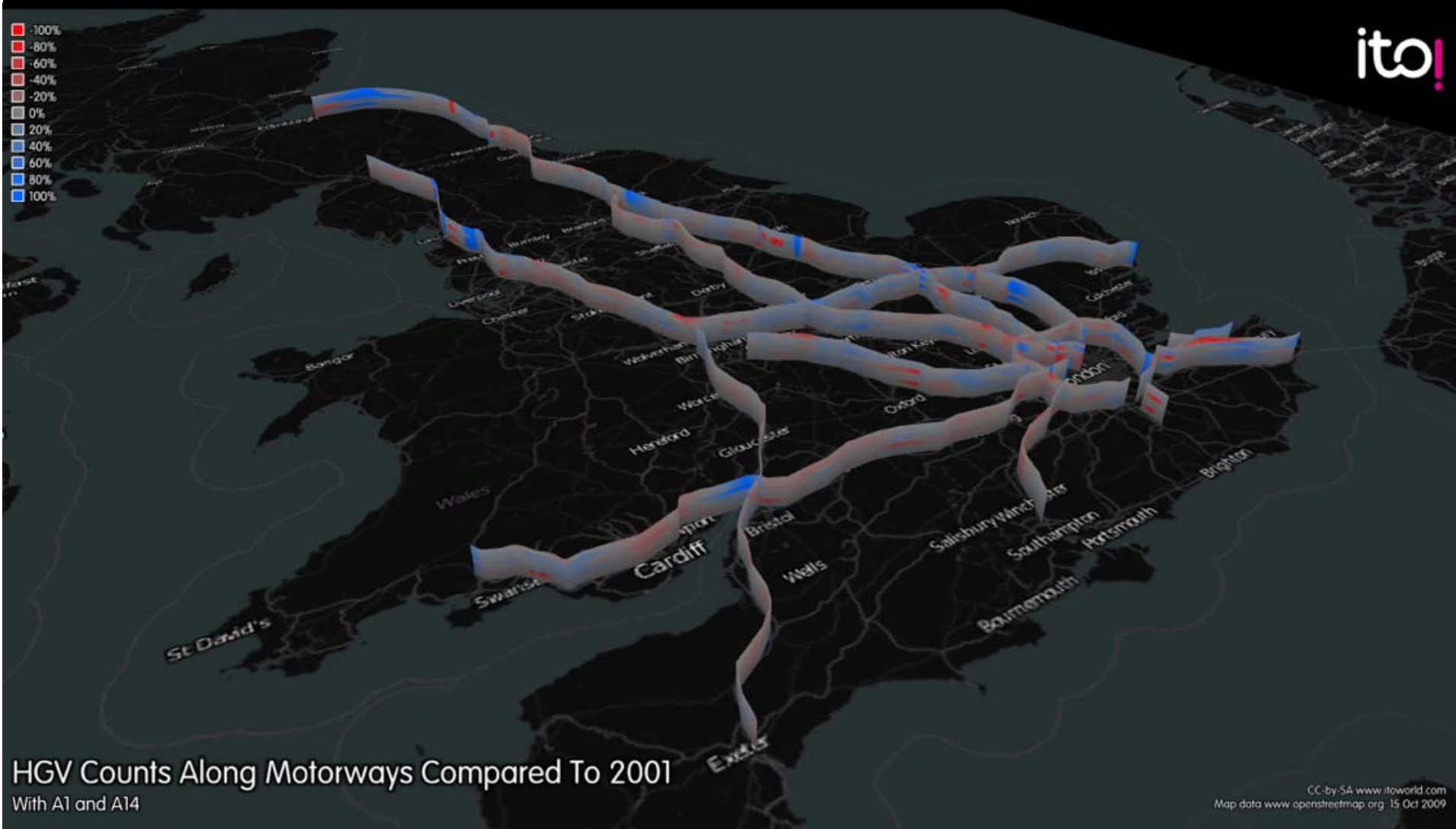
David McCandless



2008 London Counts Compared To 2001

CC-by-SA [www.itoworld.com](http://www.itoworld.com)  
Map data [www.openstreetmap.org](http://www.openstreetmap.org): 15 Oct 2009

Chris Osborne



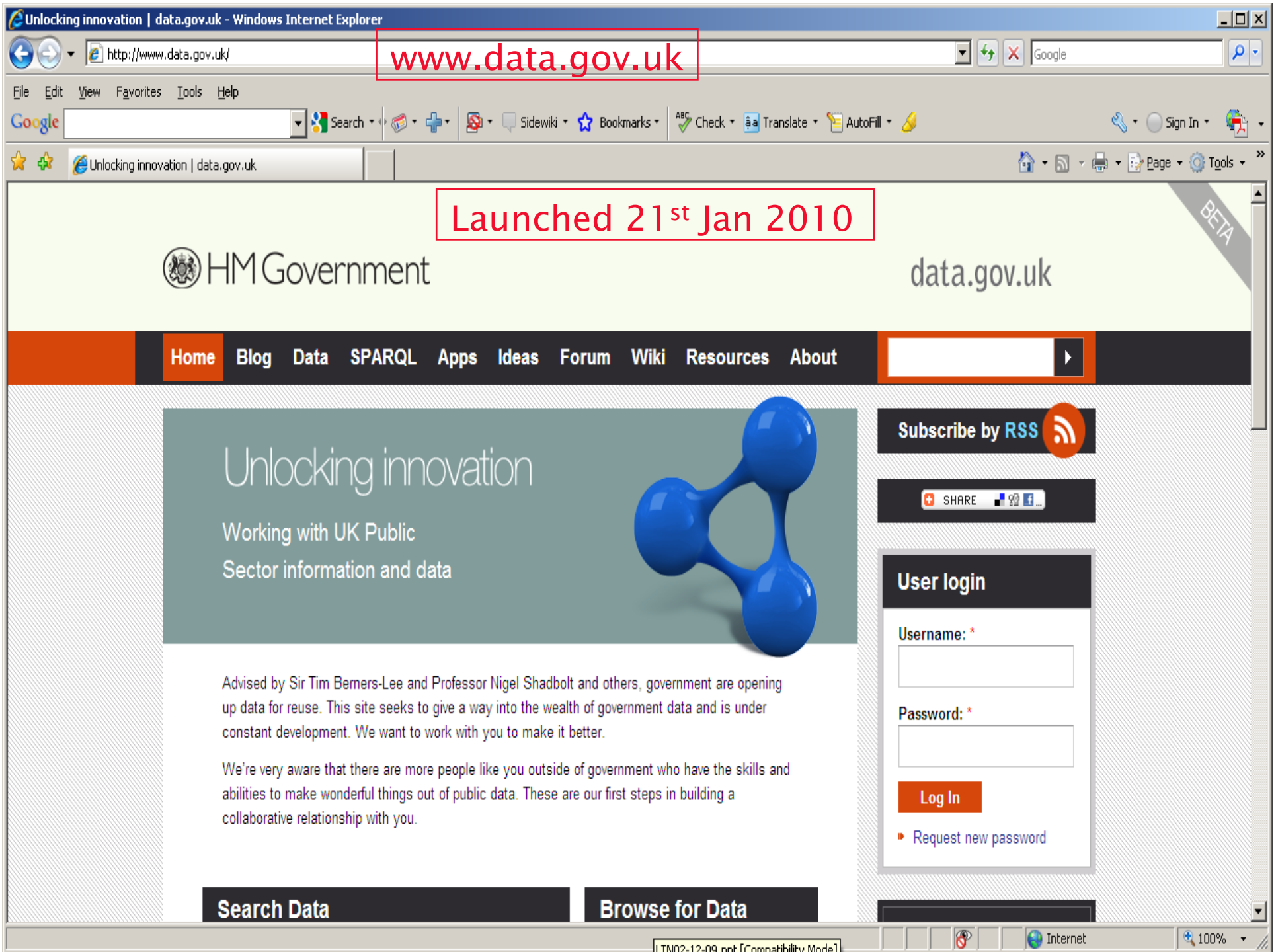
HGV Counts Along Motorways Compared To 2001  
With A1 and A14

CC-by-SA [www.itoworld.com](http://www.itoworld.com)  
Map data [www.openstreetmap.org](http://www.openstreetmap.org) 15 Oct 2009



# Challenges to free Public Data

- Cultural & Organisational
  - Data hugging
  - We have released it - sort of...
  - License impediments
- Worries about
  - Confidentiality
  - How it will be Interpreted
  - QoI
  - Lost value
  - Disrupted workflow
  - New IT procurement
- Importance of local authority and other public bodies
- Capability
  - What assets have we got (no duty to hold register of public non-personal datasets)
  - Who can do this work
  - How to share best practice
  - Minting URIs
- Policy
  - Extending the principle to all public sector bodies
  - Consistency in what is published, how it is licensed and its format
  - Range of exceptions do need to be appreciated
  - Confidentiality, accuracy, liability and reputation





Unlocking innovation | data.gov.uk - Windows Internet Explorer

http://www.data.gov.uk/

Google

File Edit View Favorites Tools Help

Google

Search

Sidewiki

Bookmarks

Check

Translate

AutoFill

Sign In

Unlocking innovation | data.gov.uk

Page Tools

Search Data

Enter keyword(s)

e.g. education, NHS, crime, transport, environment

Search

Powered by: CKAN


Browse for Data

[List all datasets](#)

[Common tags](#)

Most Recent Apps


View all apps +



Clear Books - Companies House Data

An accounting software app that allows you to efficiently search and store data from over 2 million companies through the Companies House WebCheck service.


By Nick Braithwaite  
21/01/2010 - 11:33



Scottish Index of Multiple Deprivation

Rich interactive map describing areas of concentrated deprivation across Scotland from 2009. Makes comparative use of previous data sets from 2006 and 2004,

By Dan Moss  
21/01/2010 - 11:20



mycounciltax.org.uk

mycounciltax.org.uk is an unofficial mash-up, which was built using publicly available government information. mycounciltax.org.uk makes it easier to find out how much...


By Don Riden  
21/01/2010 - 10:56

What is the Semantic Web?

Combining different data sources has never been easy but the Semantic Web will enable data to be joined easily across boundaries.

Read more

Digital Engagement Twitter stream



RT @DirDigEng:  
#datagovuk  
hearing  
@tomtaylor of  
@newspaperclub  
abt the Postcode Paper  
<http://bit.ly/35Kx9n>: data use  
doesn't have to be digital

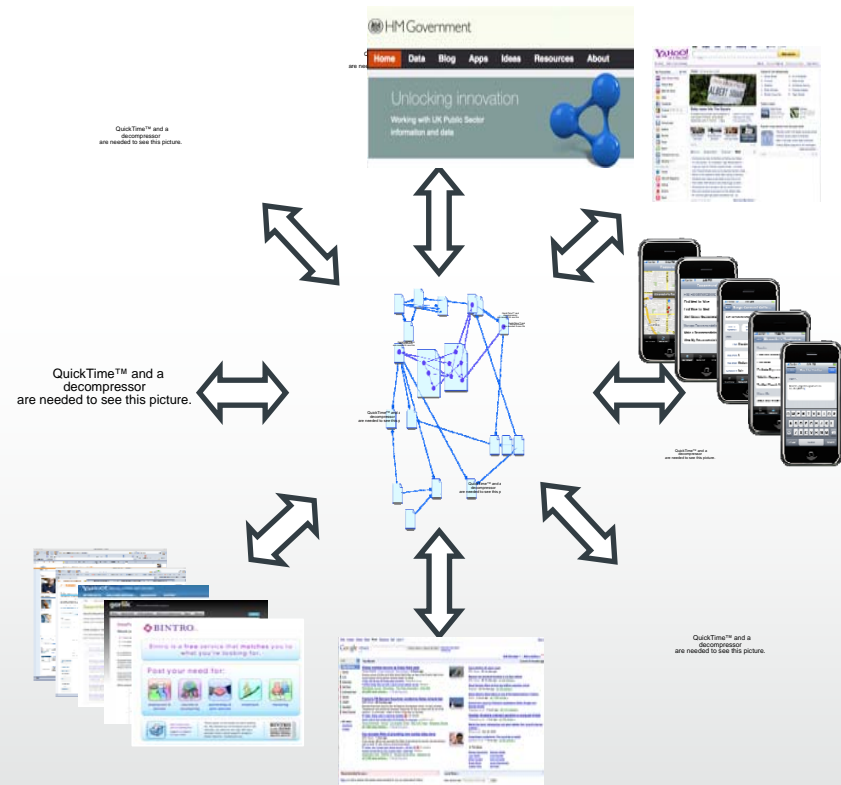
Follow us on Twitter

Internet

100%

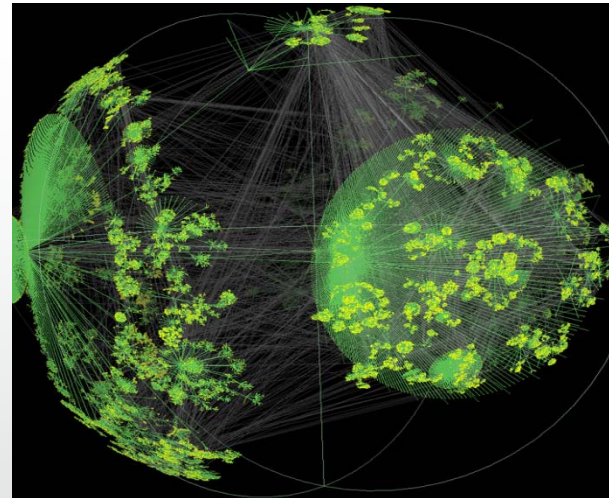
# Why does this matter to you

- Supports information interoperability
- The Sum is greater than the parts
- Exploitation of
  - open data
  - open standards
- Can apply to data intranets as well as extranets
- Publishing, news, media, manufacturing, retail, content services all classic examples where there is heterogeneous information that could be shared, linked, mapped and enriched using this approach



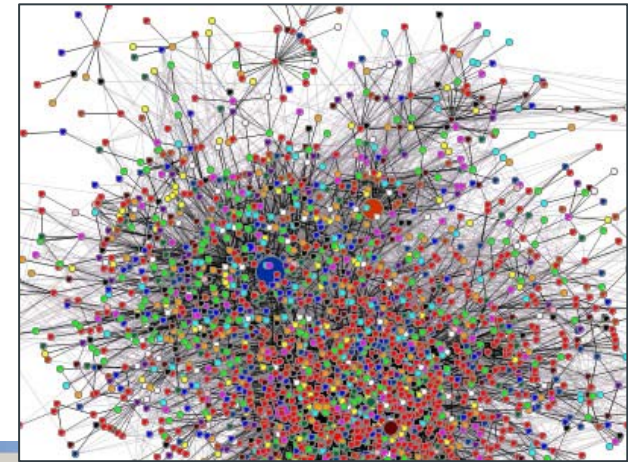
# Web 1.0

- A Web of Documents

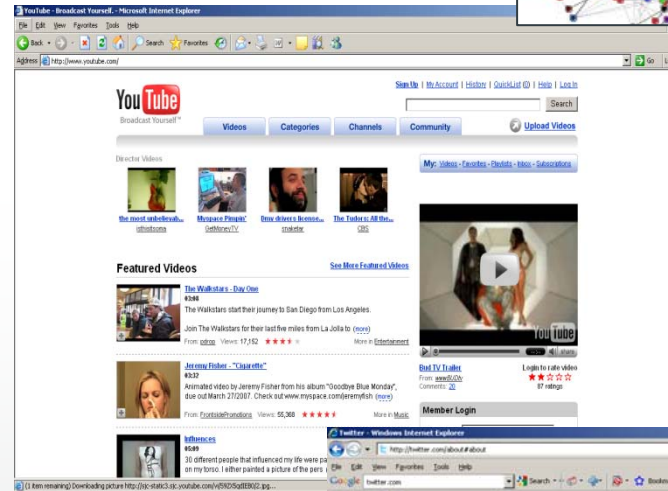


# Web 2.0

- Wiki's
- Blogs
- Flickr
- YouTube
- MySpace
- Facebook
- Second Life
- Twitter

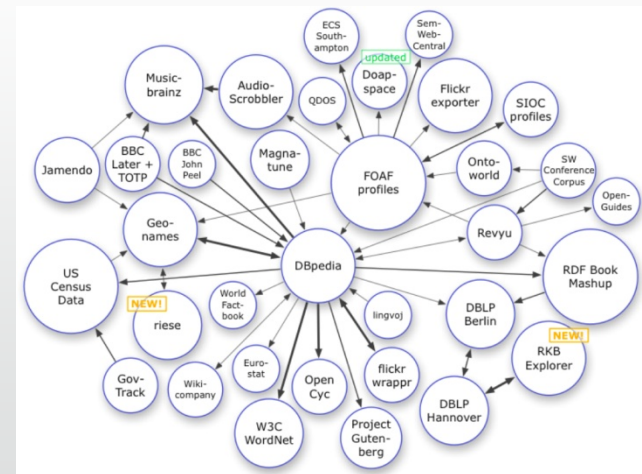


Blogosphere



# Tomorrow the Web of Linked Data

- We're eager to share data for all sorts of reasons
- Can we develop a theory of tipping points?
- What are the implications for business?
- What are social and policy implications?
- What will we do with it?
- Web 3.0?





UNIVERSITY OF  
**Southampton**



School of Electronics  
and Computer Science



**WSRI**

web science research initiative

# Introduction

## our motivation

- the Web has been transformational
- we need to understand it
- anticipate future developments
- identify opportunities and threats
- we have established a new discipline: Web Science



research / thought leadership / insight



# Web Science Research Initiative - launched in November 2006



Jim Hendler



Tim Berners-Lee



Wendy Hall



Nigel Shadbolt



Daniel Weitzner

- Research
- Thought leadership
- Education

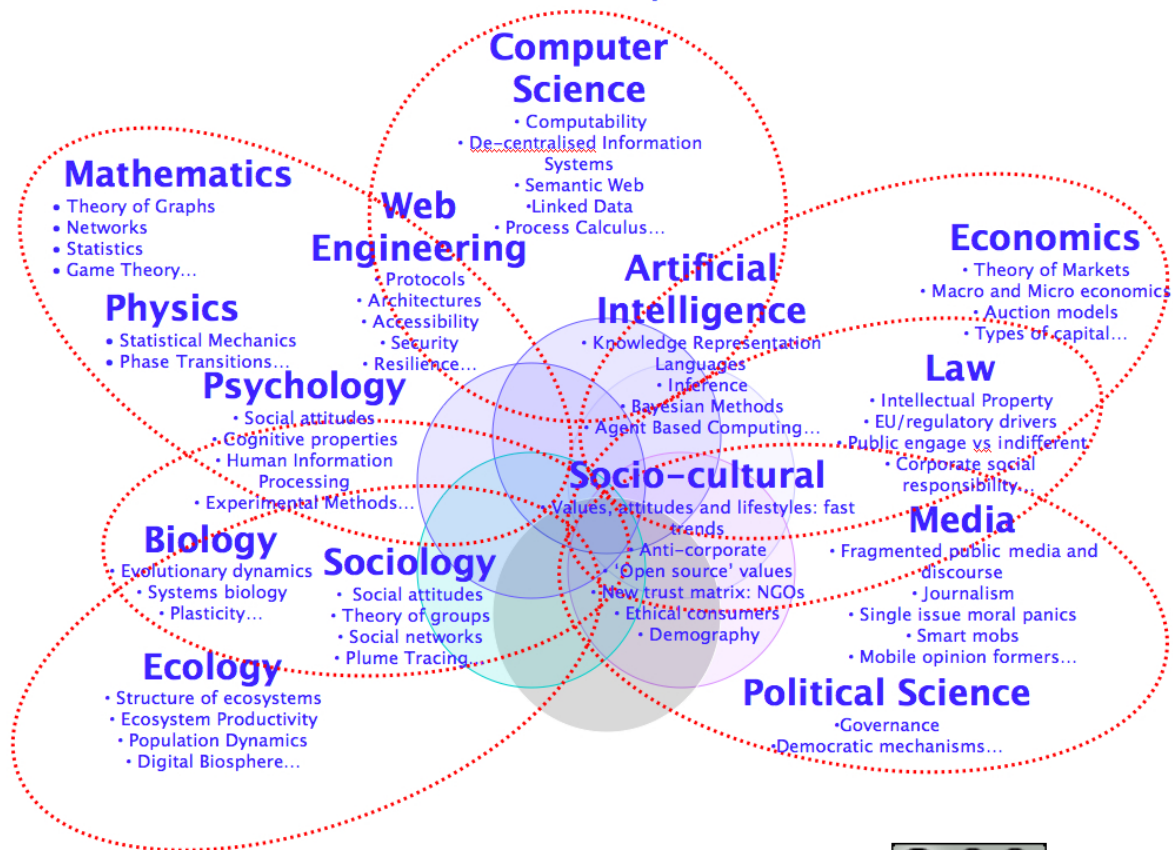
research / thought leadership / insight



# Web Science

is {inter|multi|trans}disciplinary

## Web Science: Components



Nigel Shadbolt

research / thought leadership / insight



**WSRI**  
web science research initiative

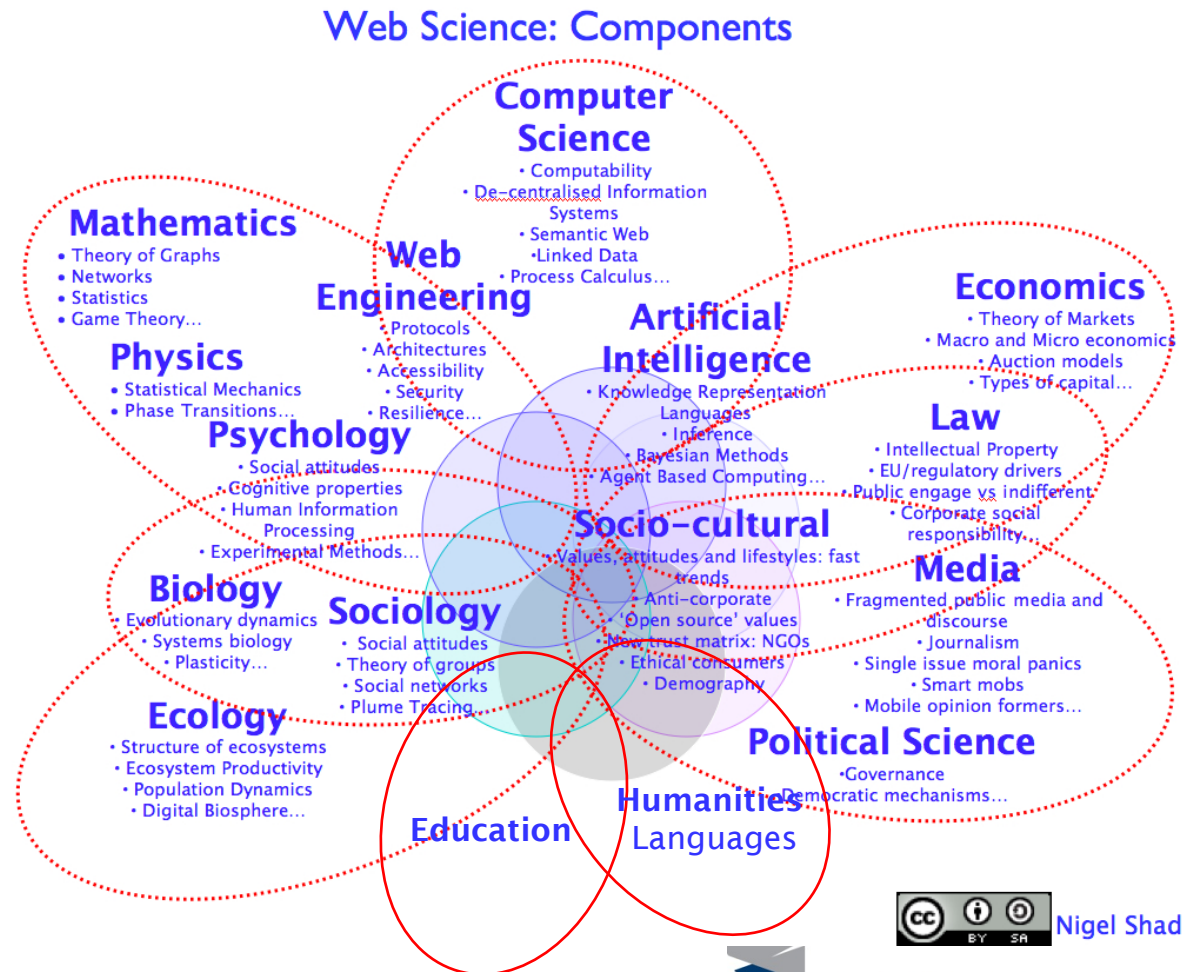


# Web Science

is about additionality

Not the union of the disciplines

But more than their intersection



Nigel Shadbolt

research / thought leadership / insight

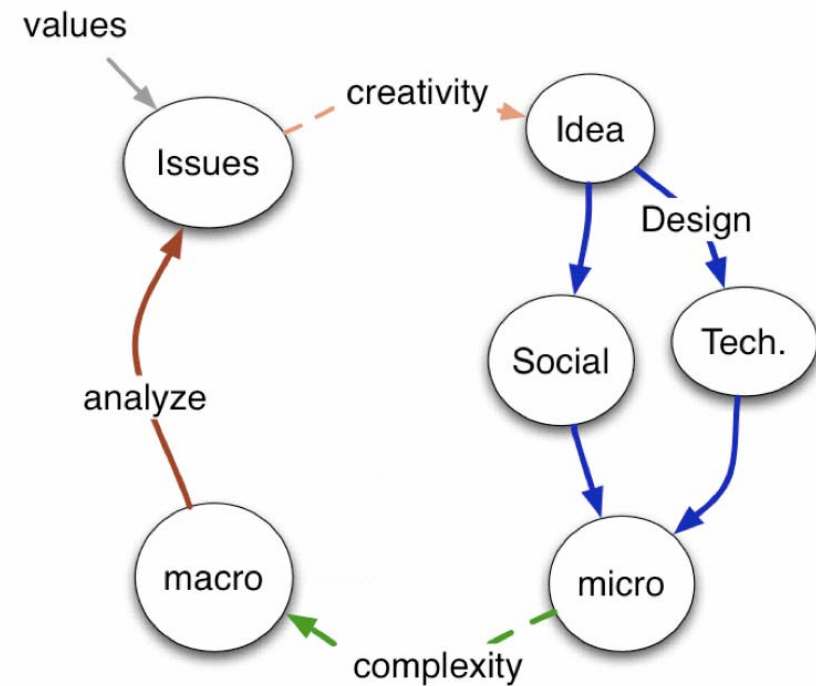


**WSRI**  
web science research initiative

# Web Science

## A Process

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



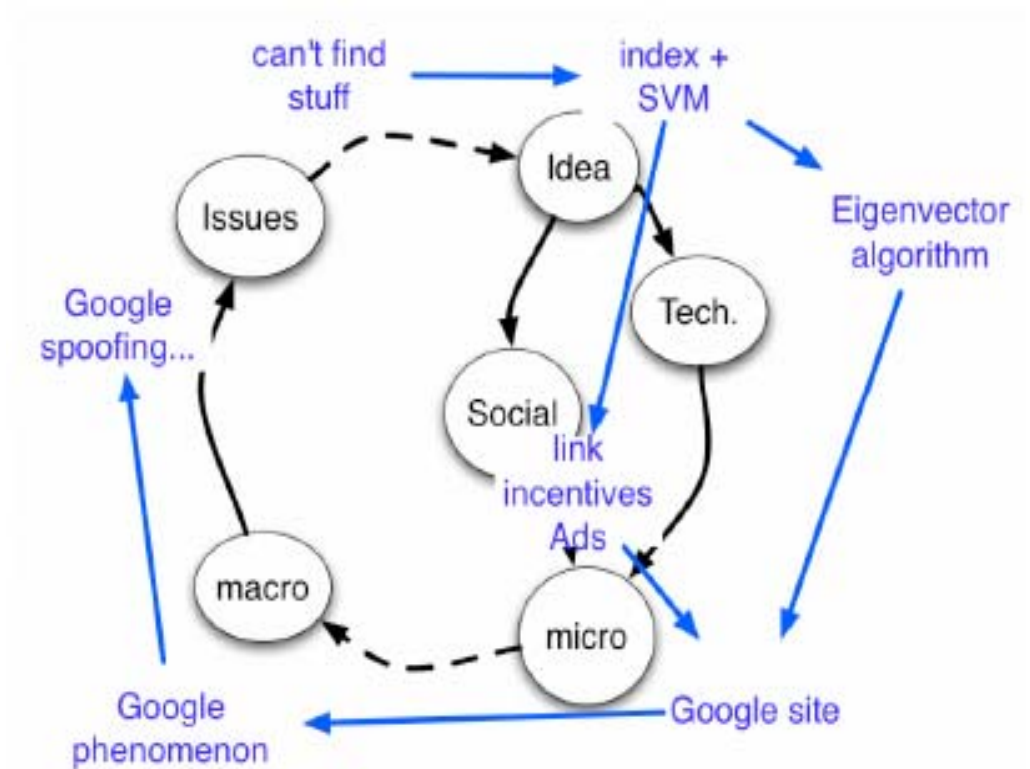
research / thought leadership / insight

# Web Science

## A Process

**We can understand after the fact, e.g the Google effect**

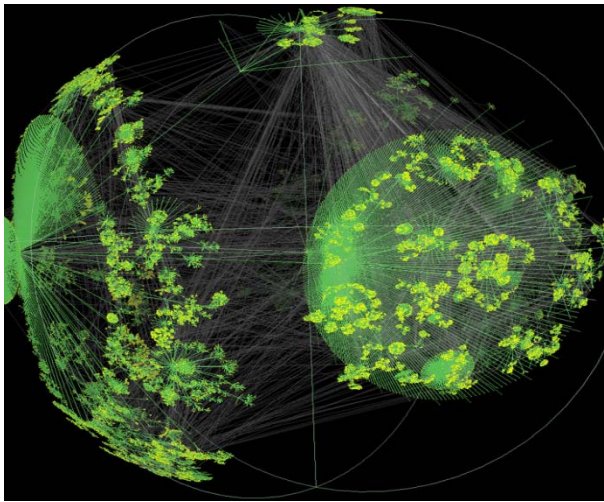
## Can we predict, anticipate before the fact?



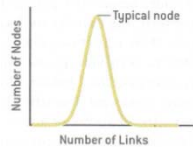
research / thought leadership / insight

# Web Science - Examples

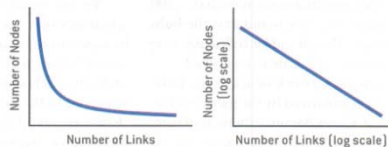
## Web Structure



Bell Curve Distribution of Node Linkages



Power Law Distribution of Node Linkages



### Scale-free

The Web has a fractal nature

### Power laws

Over the Web the numbers of links into and links out of any Web page obey a Power Law

### Small worlds

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

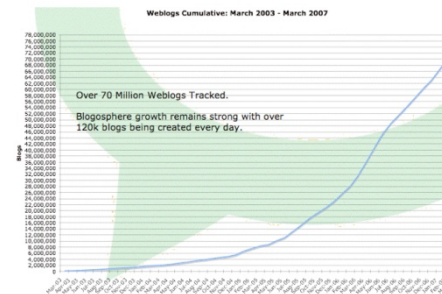
research / thought leadership / insight



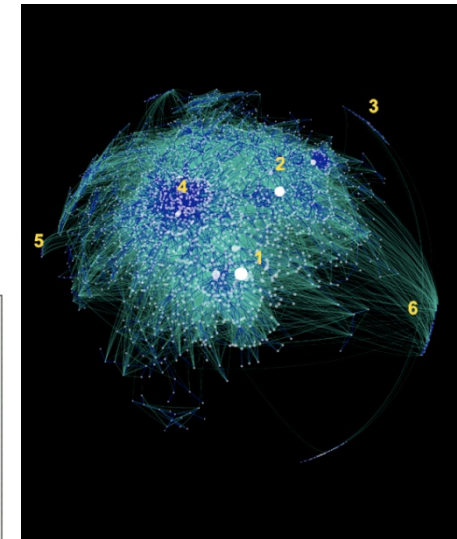
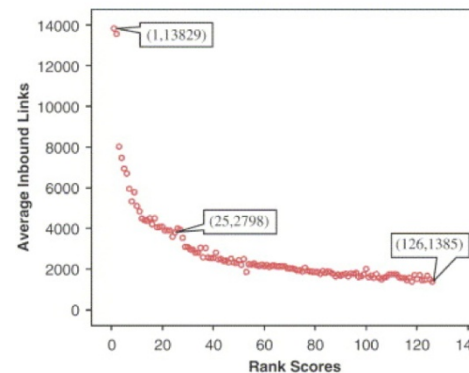
# Web Science - Examples

## The Blogosphere

- The Blogosphere
  - Why did it take off?
  - What structure does it have?
  - What drives its evolution?
- Web Science aims to understand the scientific, technical and social factors that drive the growth of the Web



Top Blog sites



research / thought leadership / insight

# Web Science - Examples

## Wikipedia - Collective Intelligence

- What is its structure?
- How stable is it?
- Why do people contribute?
- What lessons does it offer?

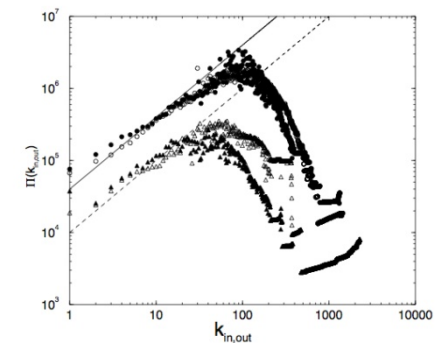
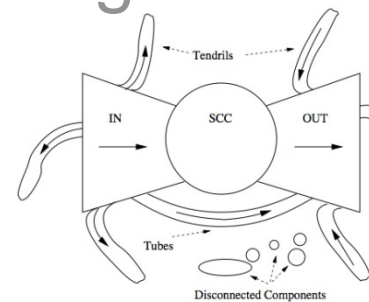


FIG. 1: The shape of the Wikipedia network

Motivation	Question example
Protective	"By writing/editing in Wikipedia I feel less lonely."
Values	"I feel it is important to help others."
Career	"I can make new contacts that might help my business or career."
Social	"People I'm close to want me to write/edit in Wikipedia."
Understanding	"Writing/editing in Wikipedia allows me to gain a new perspective on things."
Enhancement	"Writing/editing in Wikipedia makes me feel needed."
Fun	"Writing/editing in Wikipedia is fun."
Ideology	"I think information should be free."

research / thought leadership / insight

# Web Science - Examples

## Linked data

- Moving from a Web of documents to a Web of data
- Methods for linking data
- Role of the Semantic Web
- Unanticipated reuse

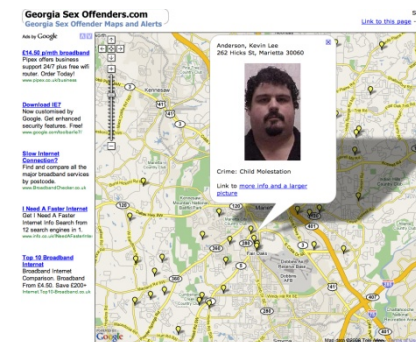


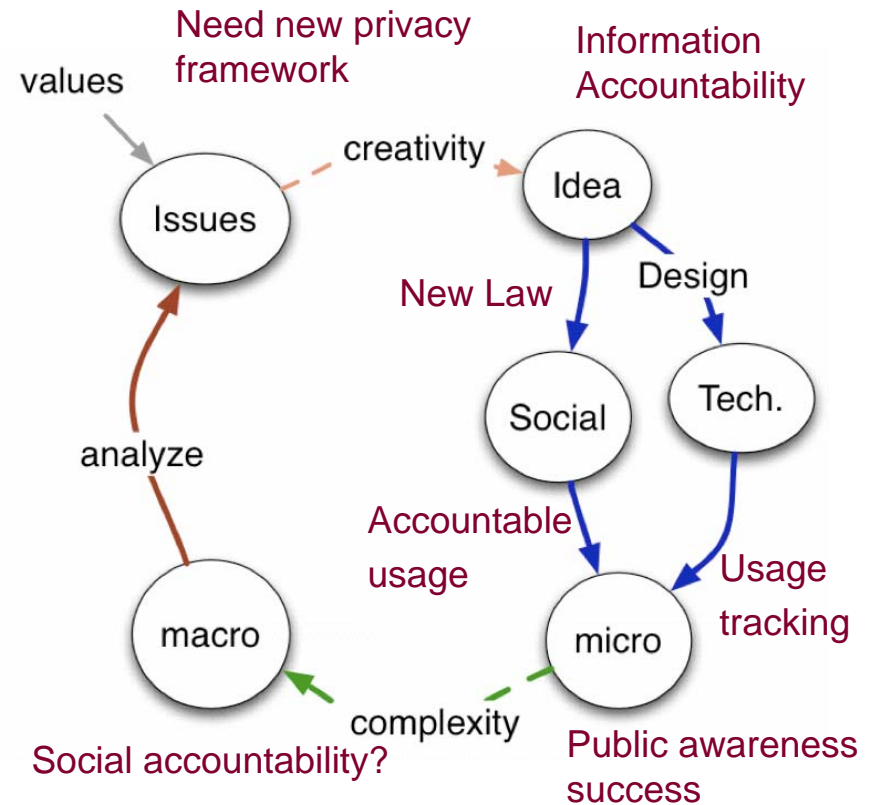
Figure 8 Browsing the Structured Data Web for Proteomics

research / thought leadership / insight

# Web Science

## Information Accountability

Issues of security, privacy, trust  
New privacy framework  
Technical architecture for usage tracking and policy explanation  
Legal rules that concentrate on permissible/impermissible use  
Analysis of technical/social interplay



research / thought leadership / insight



# Web Science

## why this matters

- the Web matters
- an essential part of humanity but only 20% of us have access at the moment
- understanding the Web is a major challenge as big as any other global cause
  - nobody owns the Web
  - what would happen if someone did?
  - imagine the consequences if the Web was turned off tomorrow
  - it has become our cultural legacy, our social heritage
  - we cannot take for granted the freedom to exchange information that is at the heart of the Web
  - it is important we protect and preserve this freedom
  - if we don't do this, who will?
- For more information see [www.webscience.org](http://www.webscience.org)



research / thought leadership / insight



[www.webscience.org](http://www.webscience.org)

# WST Outreach and Thought Leadership

- Publications e.g. Foundations and Trends in Web Science “Framework for Web Science” 2006
- Impact on research agenda of funding agencies
- Summer Graduate Schools - OII July 2008, RPI July 2009
- Conferences
  - Web Science 2009, Athens, 18-20 March 2009
  - Web Science 2010, Raleigh Durham, 26-27 April 2010 (co-located with WWW2010)
- Research talks and workshops all over the world
- Curriculum development

research / thought leadership / insight



# WSTNet - founding laboratories

- **Southampton**
- **MIT**
- **RPI**
- **Oxford Internet Institute**
- **DERI, Galway**
- **Tsinghua Graduate School at Shenzhen**
- **Koblenz**
- **VU, Amsterdam**
- **NorthWestern, Chicago**
- **UC Irvine**

research / thought leadership / insight



# 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 begin-

ning 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 their 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

# COMMUNICATIONS OF THE ACM

ACM ACM ORG

07/08 VOL. 51 NO. 7

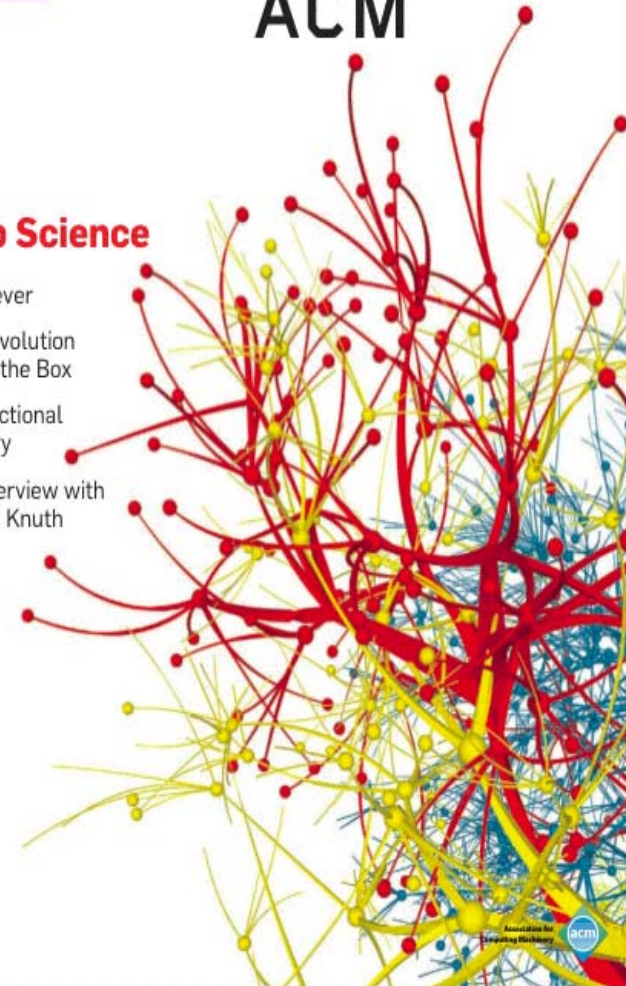
## Web Science

XML Fever

The Revolution  
Inside the Box

Transactional  
Memory

An Interview with  
Donald Knuth



Association for  
Computing Machinery

