The Emerging Science of the Web:

And Why it is Important

Professor Dame Wendy Hall

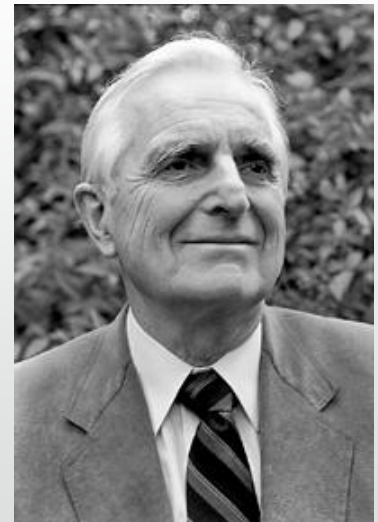23 June 2010

UNIVERSITY OF Southampton

Inspiration
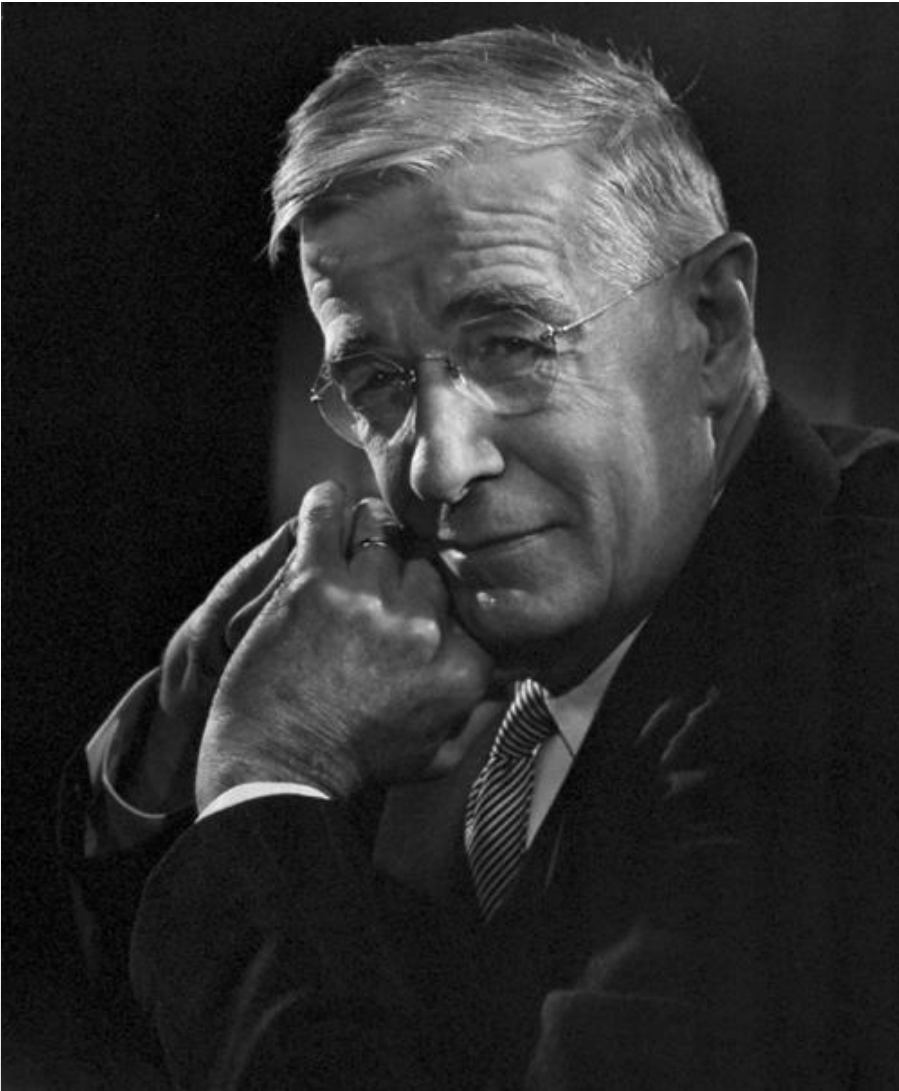
"As we may think"
Vannevar Bush
Atlantic Monthly
July 1945

Everything is deeply intertwingled

Ted Nelson
and Doug Engelbart

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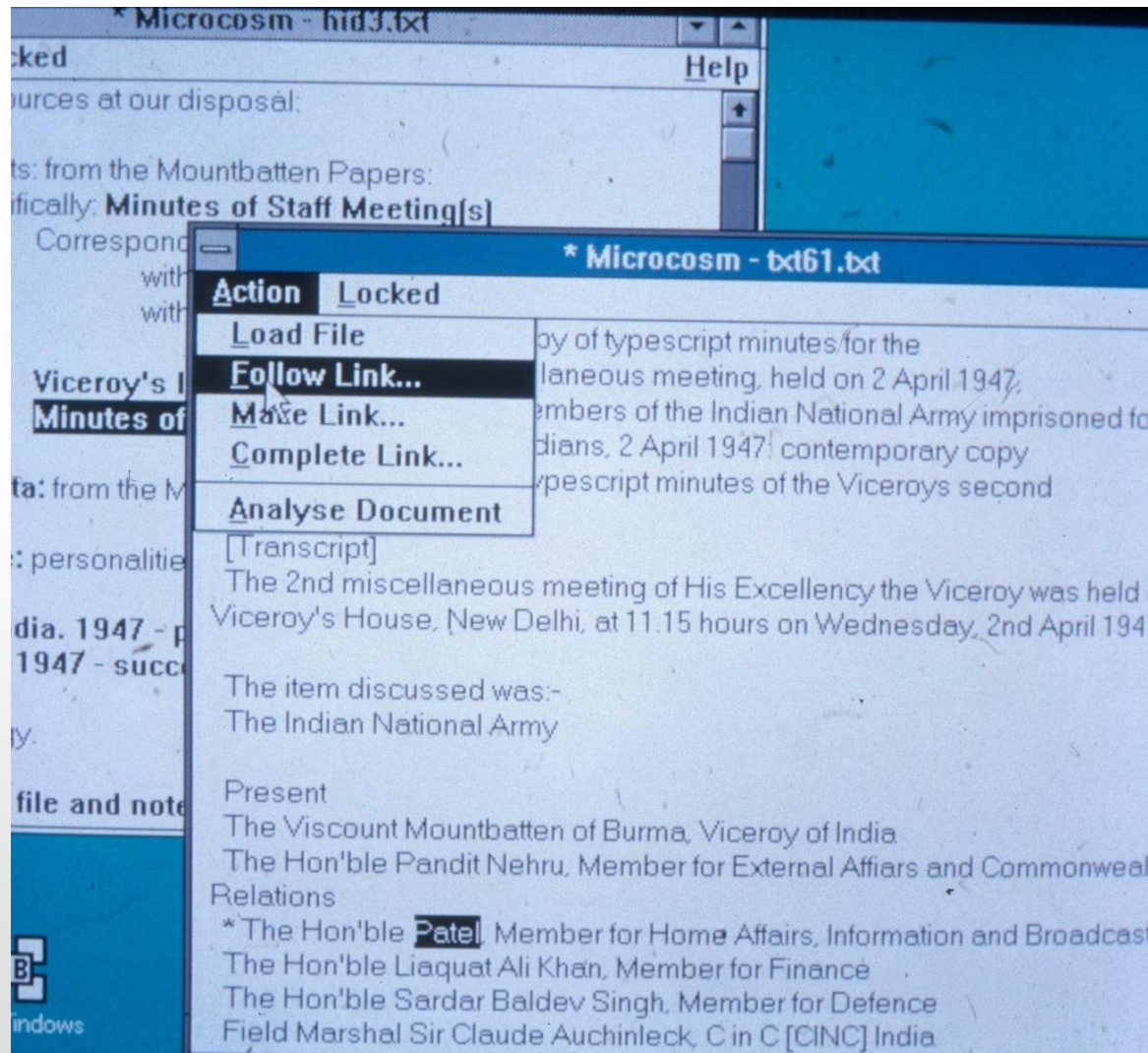
What do India, the Earl Mountbatten of Burma, the University of Southampton and my research career have in common?
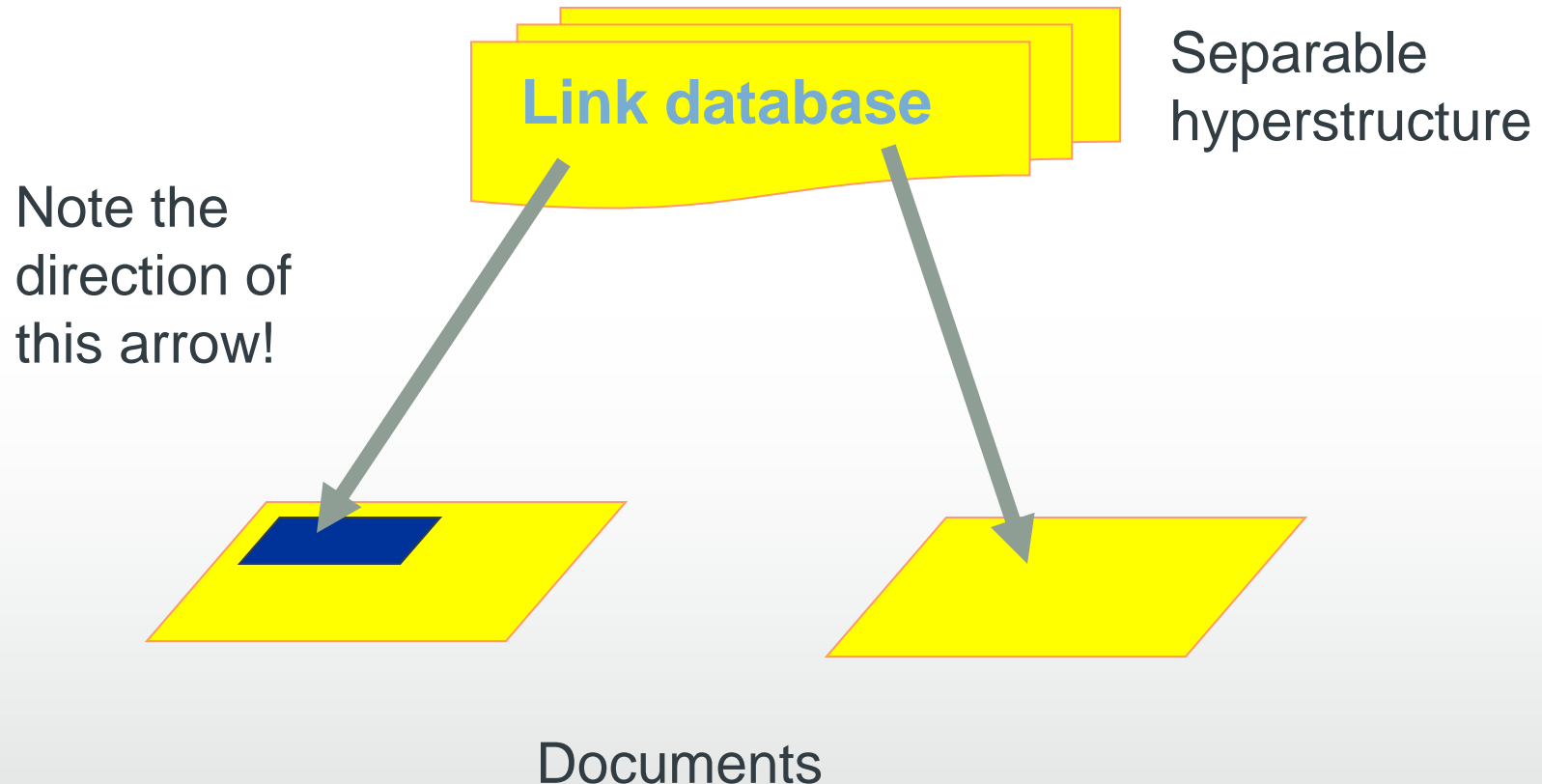
University of Southampton

The Mountbatten archive
moved to Southampton
In 1987

Microcosm: Mountbatten archive application

Linkbases in Microcosm

Link database

Separable hyperstructure

Note the direction of this arrow!

Documents

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"
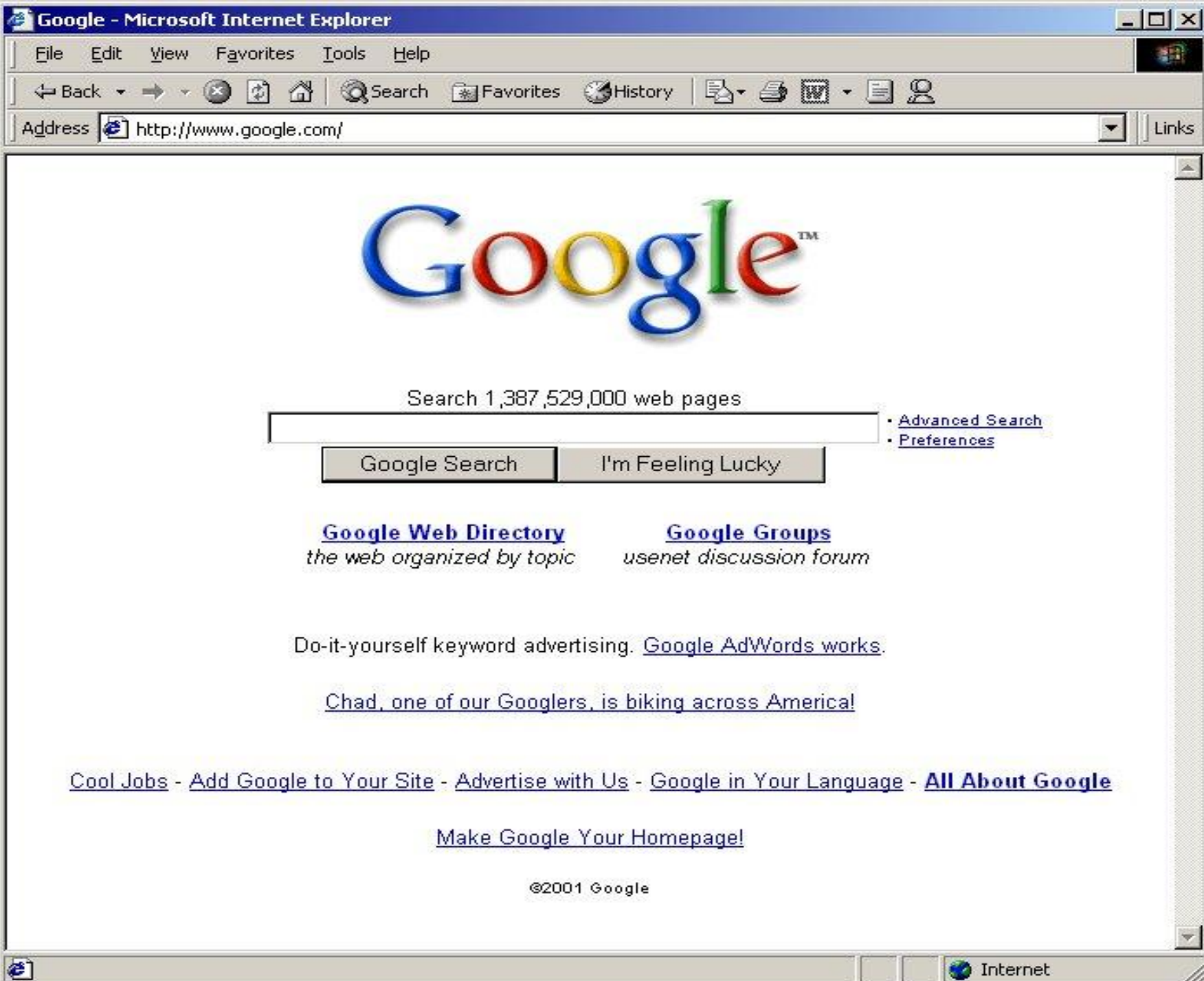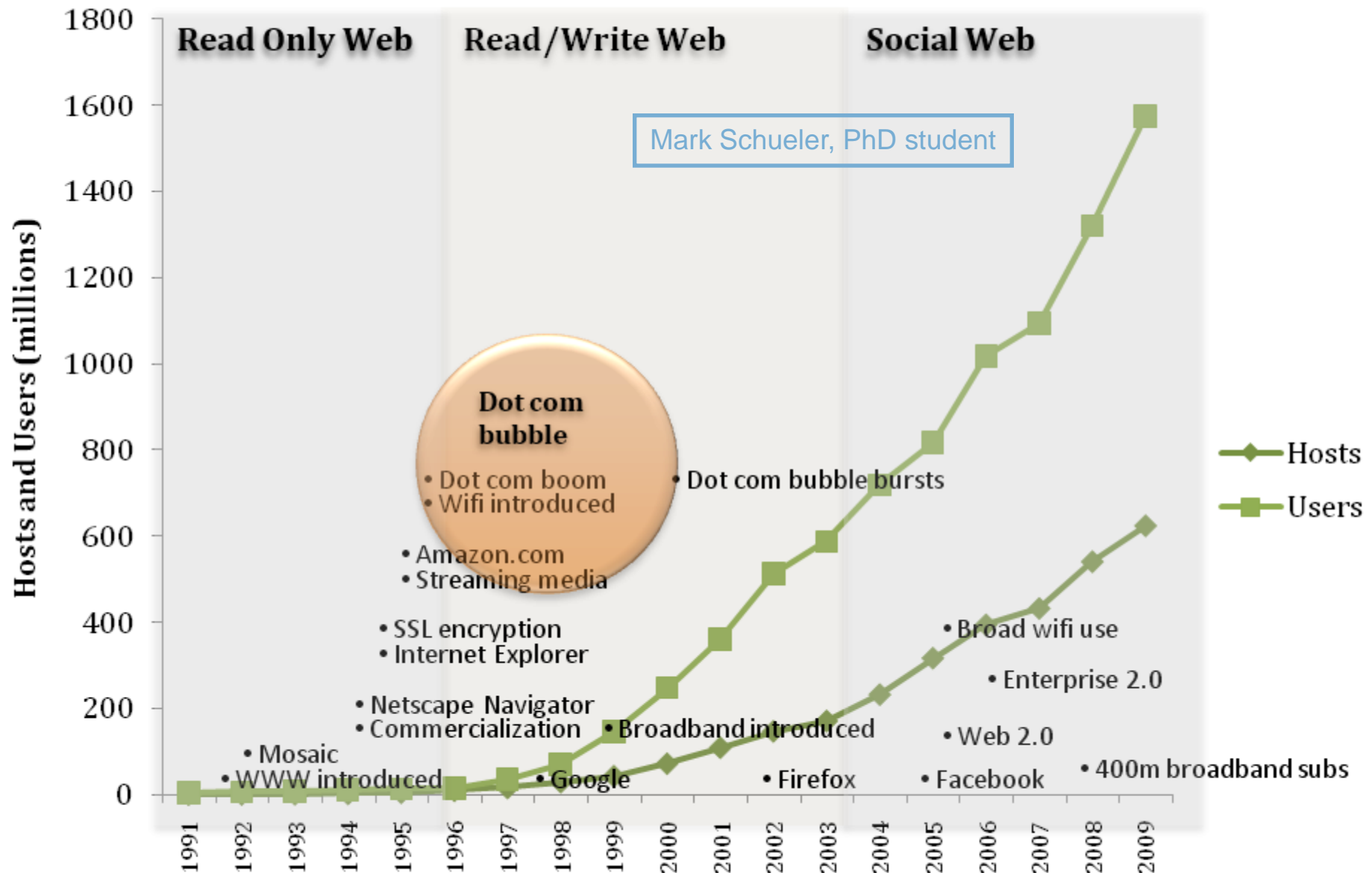
ECHT'90

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, free and universal

- But we lost (for a time) conceptual and contextual linking, and link descriptions – the Web is a strangely linkless world

- Missing links – search engines fill the gap

Internet Growth - Usage Phases - Tech Events

Mark Schueler, PhD student

Read Only Web

Read/Write Web

Social Web

Dot com bubble
- Dot com boom
- Wifi introduced

Dot com bubble bursts

- Amazon.com
- Streaming media

- SSL encryption
- Internet Explorer

- Broad wifi use

- Enterprise 2.0

- Netscape Navigator
- Commercialization

- Broadband introduced

- Web 2.0

- Mosaic
- WWW introduced

- Google

- Firefox

- Facebook

- 400m broadband subs

Hosts and Users (millions)

Hosts

Users

1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009

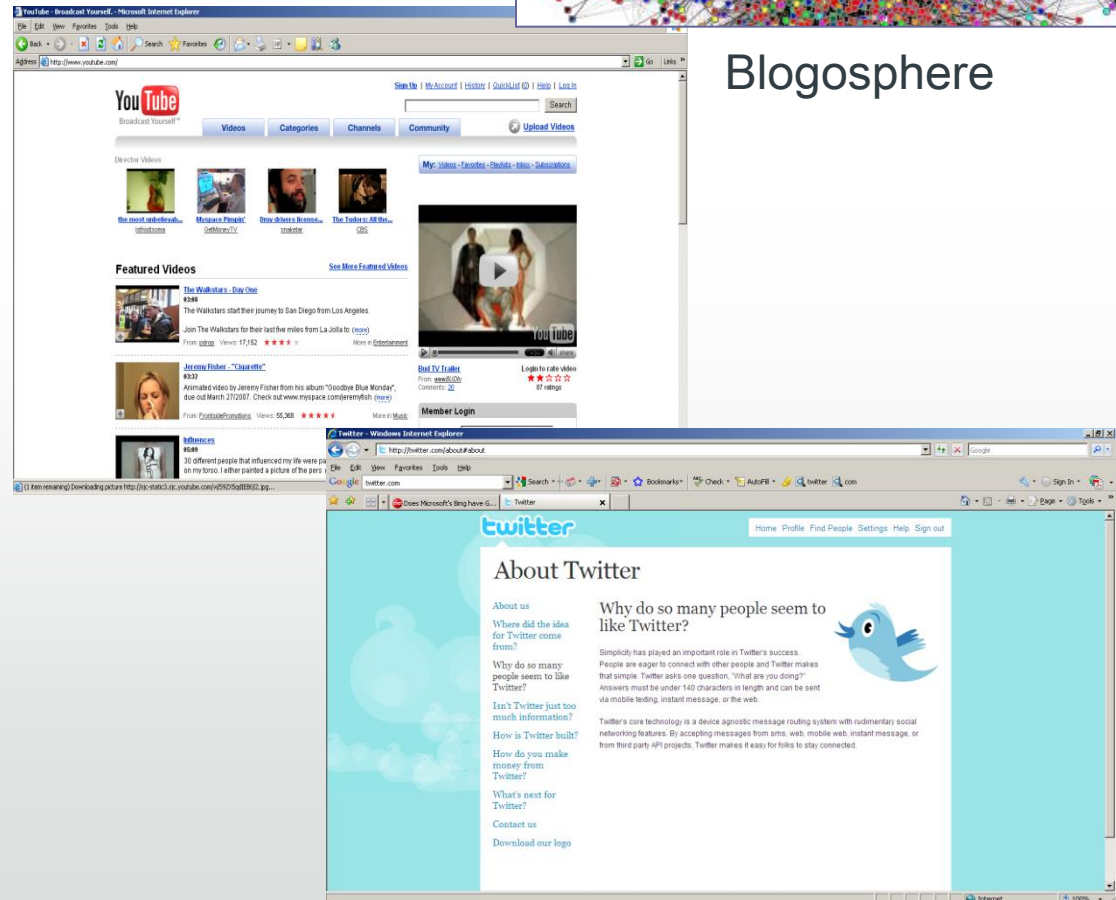**Note – events shown relate to the time axis only.**

Web 2.0

- Wiki's

- Blogs

- Flickr
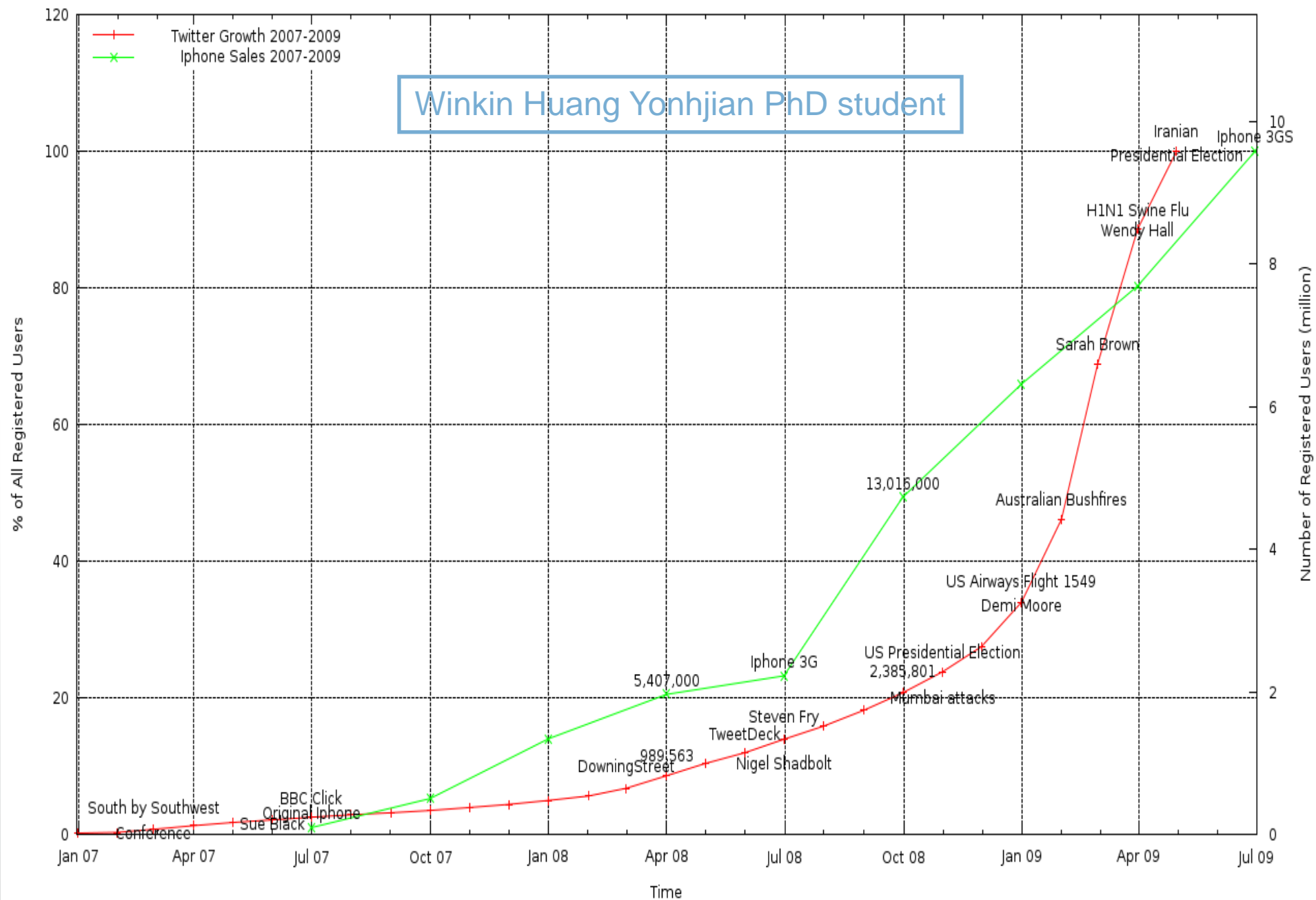
- YouTube

- MySpace

- Facebook

- Second Life

- Twitter

Blogosphere

http://en.wikipedia.org/wiki/Main_Page

Google | Wikipedia

File   Edit   View   Favorites   Tools   Help

Google | Wikipedia | Search | Bookmarks | Check | Translate | AutoFill | Wikipedia | Sign In

Wikipedia, the free ency...   realtimecollisiondetection.ne...   Home | Page | Tools

Try Beta   Log in / create account

main page | discussion | view source | history

[Hide]

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

$0.9M USD ● ─────────────────── $7.5M USD

Donate Now

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

http://www.galaxyzoo.org/

File   Edit   View   Favorites   Tools   Help

Google Galaxy Zoo   Search ▾   Sidewiki ▾   Bookmarks ▾   Check ▾   Translate ▾   AutoFill ▾   Galaxy   Zoo   Sign In ▾

Galaxy Zoo 2

Page ▾   Tools ▾

GALAXY ZOO 2

GALAXY ZOO
UNDERSTANDING COSMIC MERGERS

By colliding galaxies on your computer, you can help us trace the past, present and future of **Galaxy Zoo mergers**

We're hungry to share data and get answers to questions

How Do Galaxies Merge?

HELP US SOLVE THE PUZZLE ⊘

Home   The Story So Far   The Science   How To Take Part   Classify Galaxies   Forum   Zoo Media   Blog   FAQ   Contact Us

Welcome to Galaxy Zoo, where you can help astronomers explore the Universe

New, more detailed images added - see here for details

The Galaxy Zoo files contain almost a quarter of a million galaxies which

Classifier Log In

Click here to log in

☒ Register
☒ Forgotten Password?

Done   Internet   100%

The Semantic Web

RDF triples: Subject, Predicate, Object
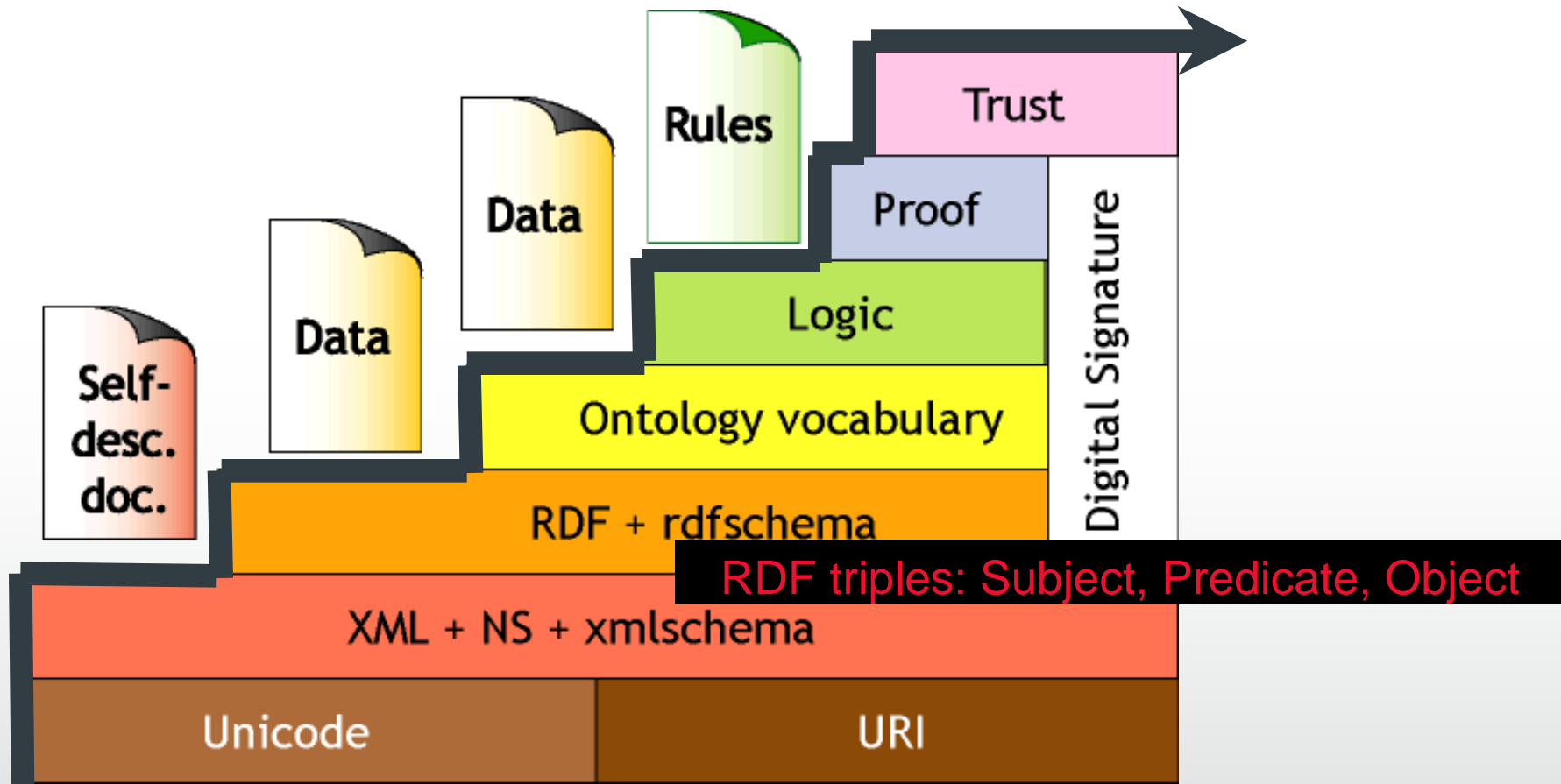
Semantic Web LayerCake **(Berners-Lee, 99;Swartz-Hendler, 2001)**

Sign In | Register

TED Ideas worth spreading

Themes	TED Conferences	About TED
Speakers	TEDx Events NEW	TED Blog
Talks	TED Prize	
Translations	TED Fellows	Q Search

TED Community

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 »

ROLEX

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.

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.

The Semantic Web Revisited

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

23

Linked Data on the Web: May 2007

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

As of May 2007
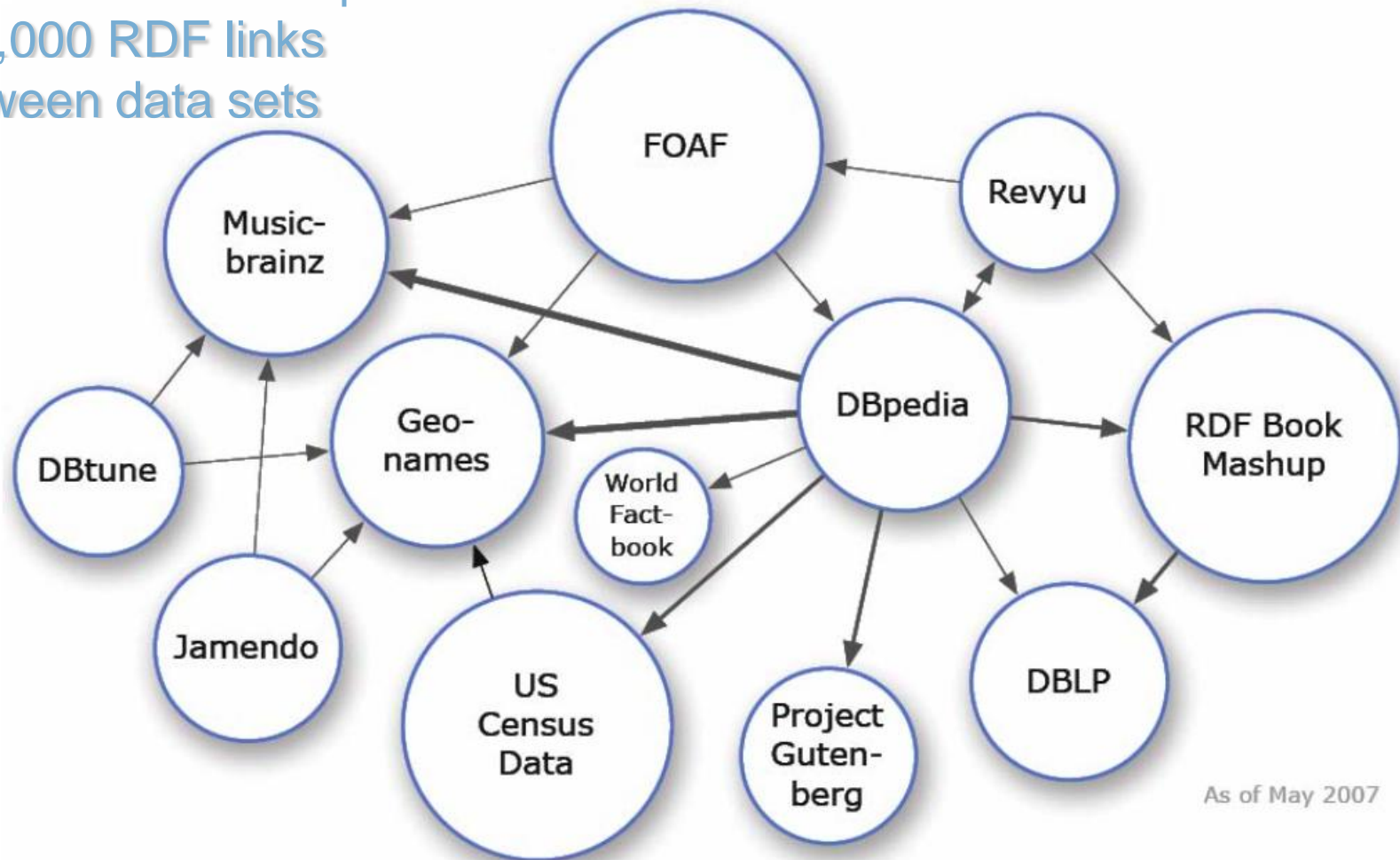
Linked Data on the Web: April 2008

23 billion
RDF Triples
3 million
RDF links
between
data sets
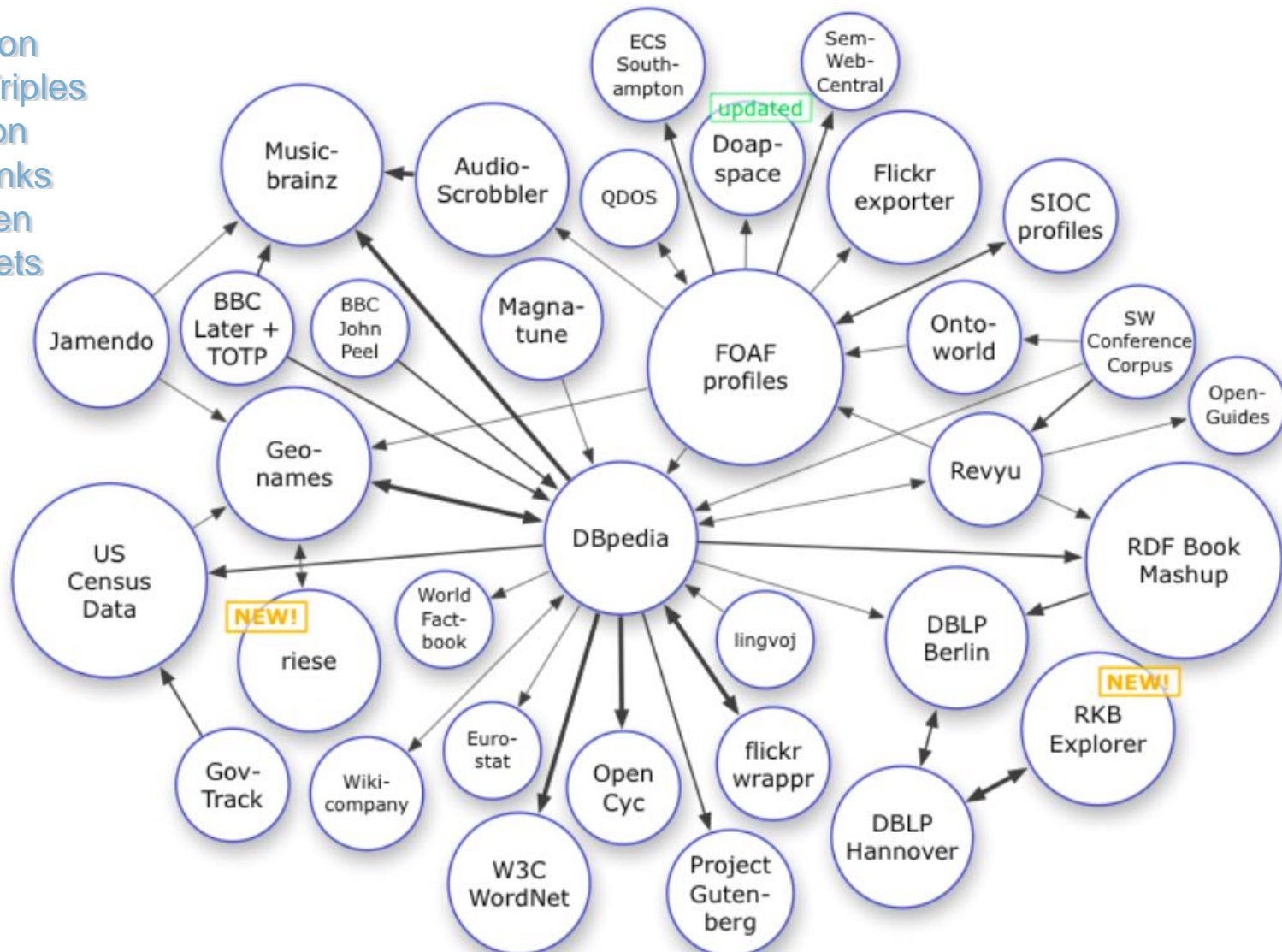
LOD Datasets on the Web: March 2009
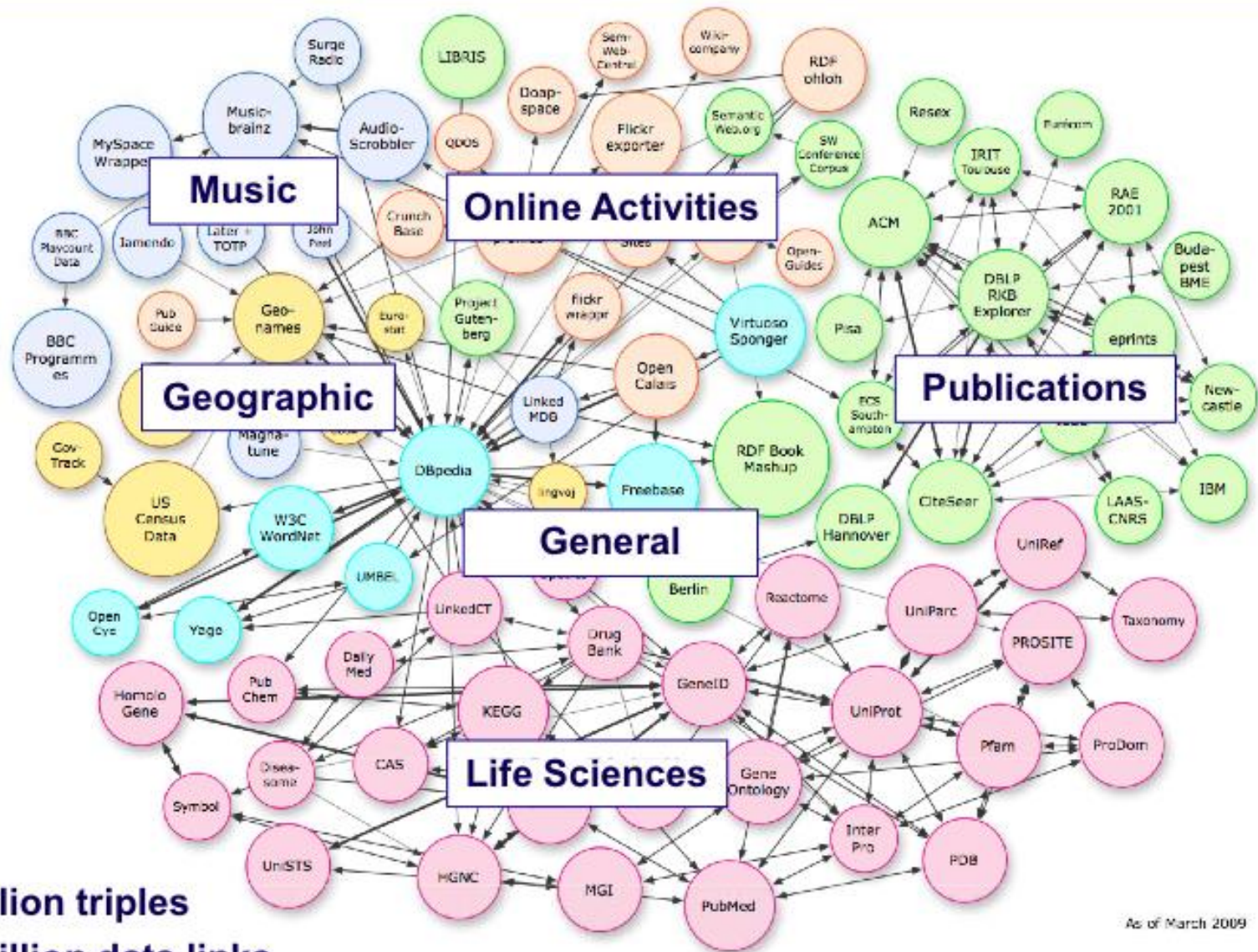
4.5 billion triples
180 million data links

As of March 2009

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.

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 then 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 :)

Google

University Thessaloniki Web Science

Search ▾ ⋅⋅ 🔶 ▾ ⊕ ▾ | 🔷 Share ▾ 🔷 ▾ | 💬 Sidewiki ▾ | ABC Check ▾ | àa Translate ▾ | 🔰 AutoFill ▾ 🖋 » | 🔧 ▾ ⬤ Sign In ▾ | X 🔷 Convert ▾ 🔷 Select

⭐ Favorites | 👍 🔷 Suggested Sites ▾ 🔷 Web Slice Gallery ▾ 🔷 Microsoft ▾ M Best of the Web M Channel Guide M Internet Start M Microsoft

⬛ ▾ BBC BBC NEWS | UK | 🔷 Modify profile - ECS EPr... | Why Study the Web? So... | Walter Tichy - Google S... | DATA Data.gov ✕ | 🏠 ▾ 🔷 ▾ 🖂 🖶 ▾ Page ▾ Safety ▾ Tools ▾ 🔷 »

DATA.GOV
EMPOWERING PEOPLE

HOME | **DATA** ▾ | **TOOLS** | **COMMUNITY** | **METRICS** | **DIALOGUE**

A GLOBAL MOVEMENT TO DEMOCRATIZE DATA

VIEW MORE ▸

Most Popular Datasets

1. Worldwide M1+ Earthquakes, Past 7 Days
2. U.S. Overseas Loans and Grants (Greenbook)
3. Latest Volumes of Foreign Relations of the...
4. OSHA Data Initiative - Establishment...
5. IT Dashboard - Federal IT Spending (major...

SEARCH OUR CATALOGS

🔍 Search our catalogs.. | **SEARCH ▸**

⟨ ❚❚ ⟩

APPS

COMMUNITY

Data.gov is leading the way in democratizing public sector data and driving innovation. The data is being surfaced from many locations making the Government data stores available to researchers to perform their own analysis. Developers are finding good uses for the datasets, providing interesting and useful

SEMANTIC WEB

As the Web of linked documents evolves to include the Web of linked data, we're working to maximize the potential of Semantic Web technologies to

🌐 Internet | Protected Mode: On | 🔷 ▾ 🔍 100% ▾

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?
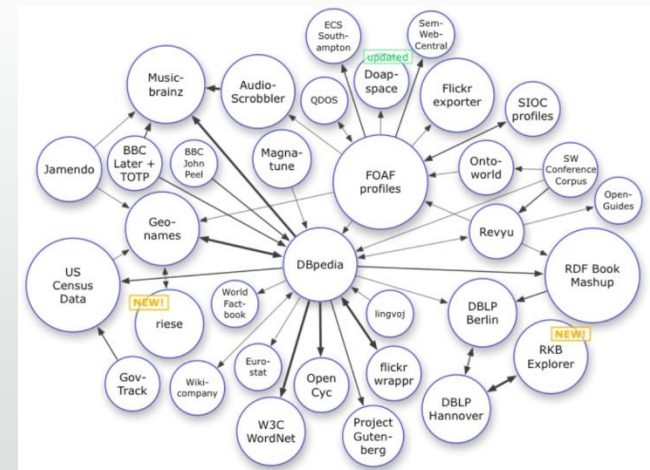
Introduction
our motivation

- 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

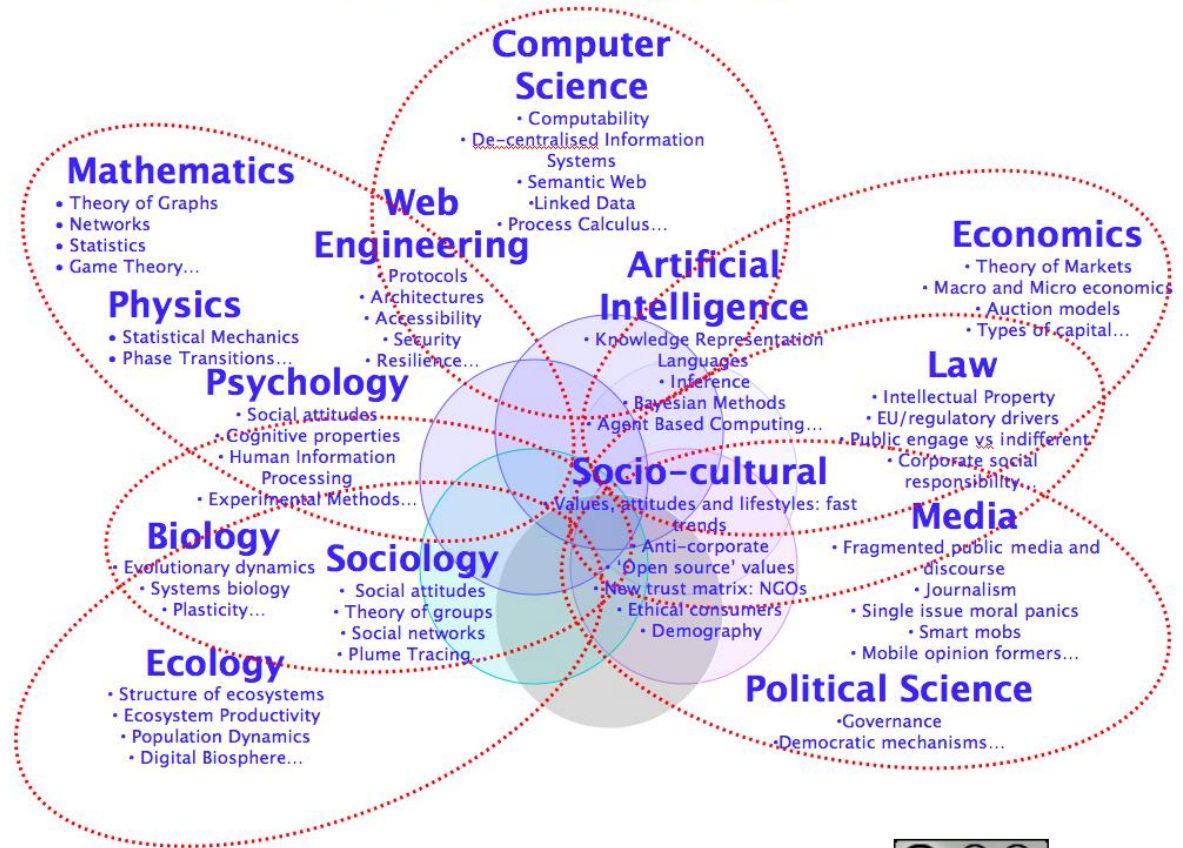- Research
- Thought leadership
- Education

research / thought leadership / insight

WSRI
web science research initiative

Web Science

is {inter|multi|trans}disciplinary

Web Science: Components

research / thought leadership / insight

Web Science
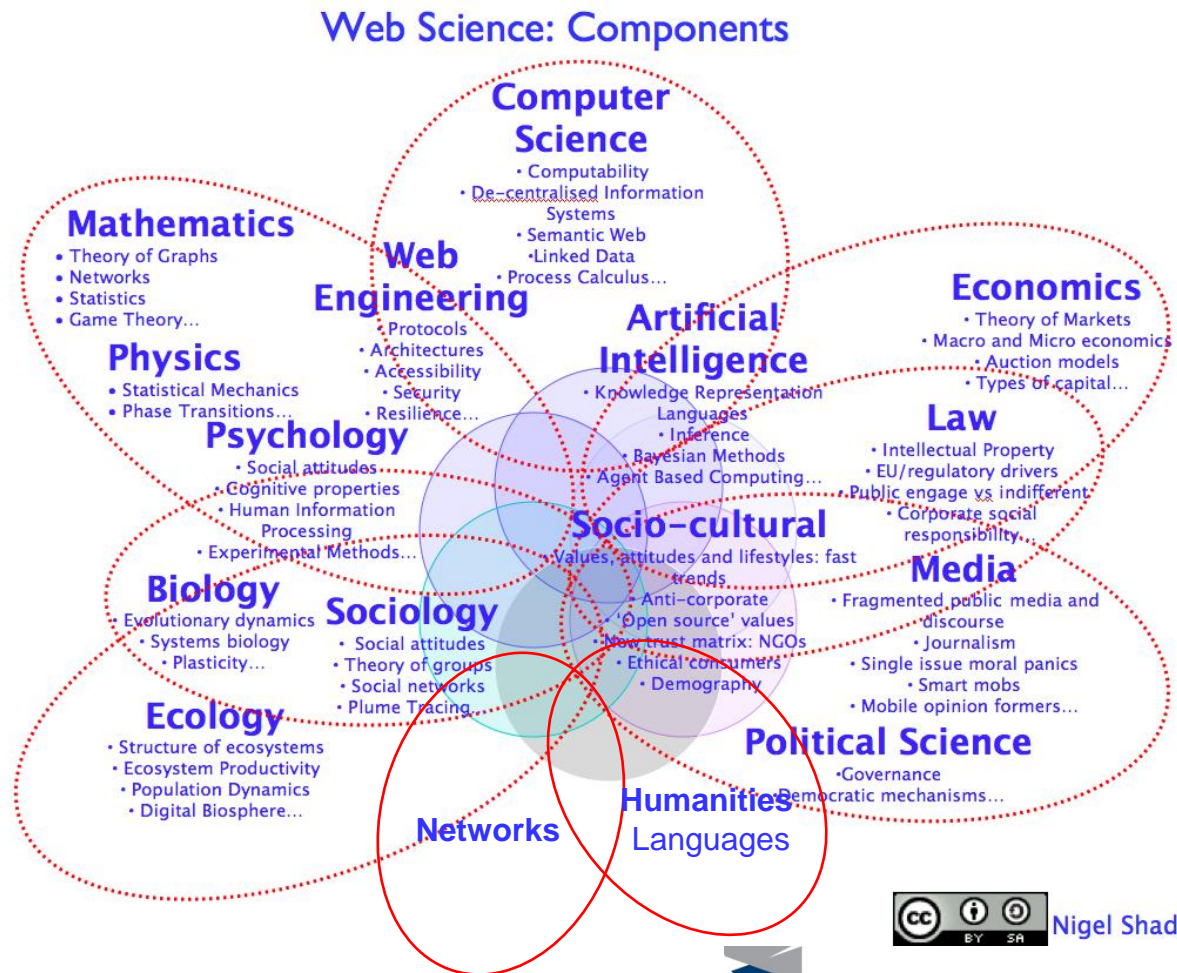## is about additionality

Not the union of the disciplines

But more than their intersection

Web Science: Components

Computer Science
· Computability
· De-centralised Information Systems
· Semantic Web
· Linked Data
· Process Calculus...

Mathematics
· Theory of Graphs
· Networks
· Statistics
· Game Theory...

Web Engineering
· Protocols
· Architectures
· Accessibility
· Security
· Resilience...

Physics
· Statistical Mechanics
· Phase Transitions...

Artificial Intelligence
· Knowledge Representation Languages
· Inference
· Bayesian Methods
· Agent Based Computing...

Economics
· Theory of Markets
· Macro and Micro economics
· Auction models
· Types of capital...

Law
· Intellectual Property
· EU/regulatory drivers
· Public engage vs indifferent
· Corporate social responsibility...

Psychology
· Social attitudes
· Cognitive properties
· Human Information Processing
· Experimental Methods...

Socio-cultural
Values, attitudes and lifestyles: fast trends
· Anti-corporate
· 'Open source' values
· New trust matrix: NGOs
· Ethical consumers'
· Demography

Media
· Fragmented public media and discourse
· Journalism
· Single issue moral panics
· Smart mobs
· Mobile opinion formers...

Biology
· Evolutionary dynamics
· Systems biology
· Plasticity...

Sociology
· Social attitudes
· Theory of groups
· Social networks
· Plume Tracing...

Ecology
· Structure of ecosystems
· Ecosystem Productivity
· Population Dynamics
· Digital Biosphere...

Political Science
· Governance
· Democratic mechanisms...

Humanities
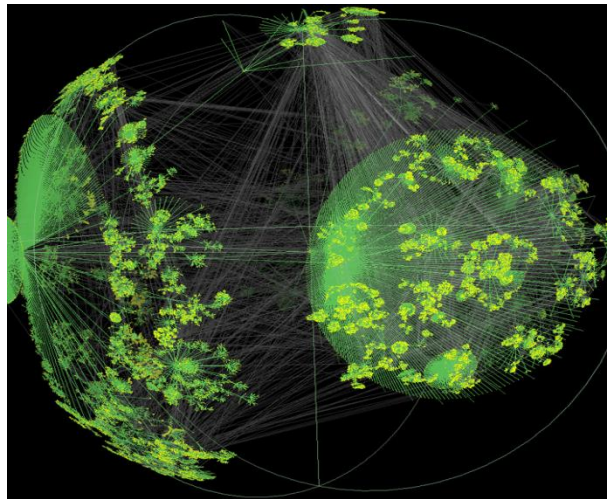Languages

Networks

Nigel Shadbolt

research / thought leadership / insight
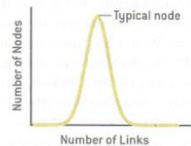
WSRI
web science research initiative

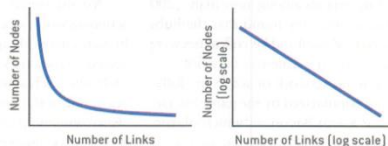# Web Science - Examples
Web Structure

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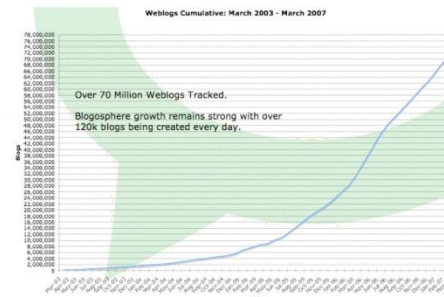
Small worlds
The average distance (or diameter) is much smaller than the order of the graph.

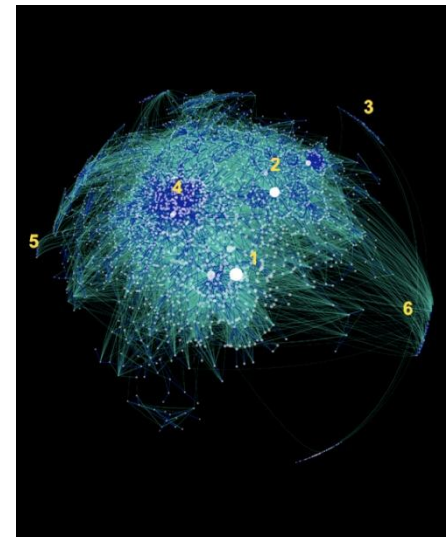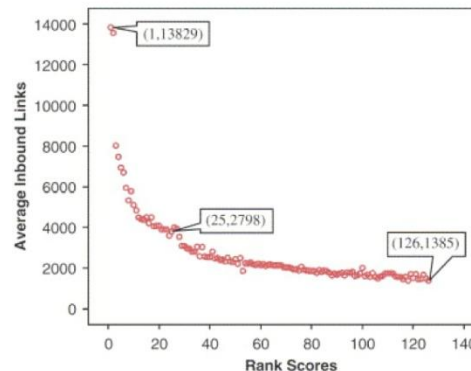research / thought leadership / insight

WSRI
web science research initiative

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

Weblogs Cumulative: March 2003 - March 2007

Over 70 Million Weblogs Tracked.

Blogosphere growth remains strong with over 120k blogs being created every day.

Top Blog sites

(1,13829)

(25,2798)

(126,1385)

Web Science - Examples
Wikipedia - Collective Intelligence

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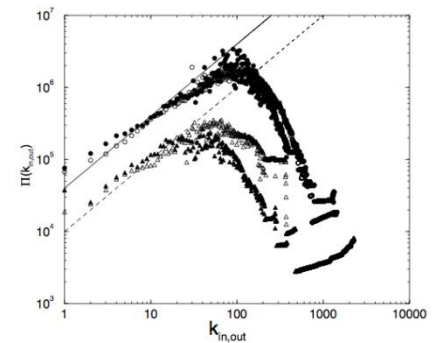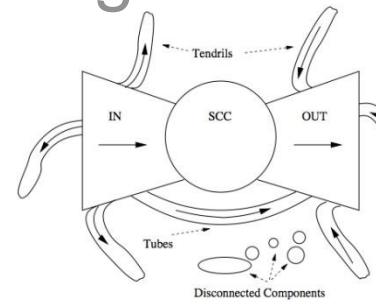- Why do people contribute?
- What lessons does it offer?

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
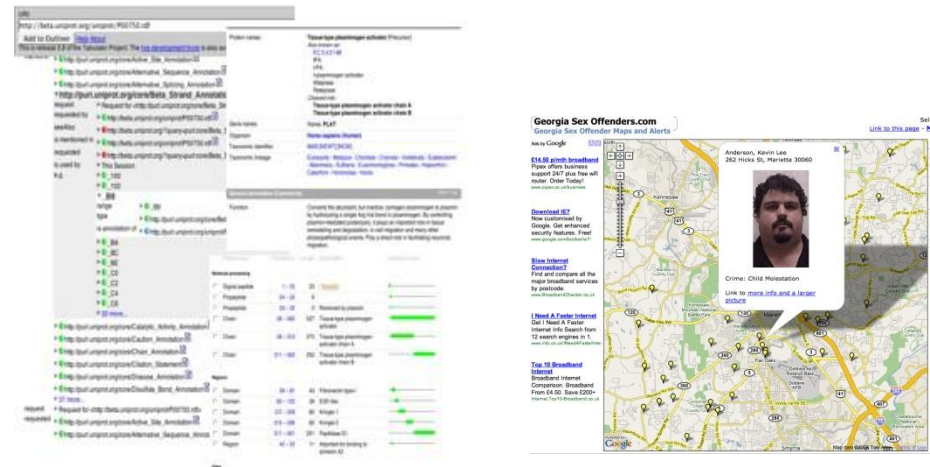- Role of the Semantic Web
- Unanticipated reuse

Figure 8 Browsing the Structured Data Web for Proteomics

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*

S ince 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 improve computing significantly. We science was launched as a forma discipline in November 2006 when the two of us and our col leagues at the Massachusetts In stitute of Technology and th University of Southampton i England announced the begin ning of a Web Science Research Initiative. Lead ing researchers from 16 of the world's top uni versities have since expanded on that effort.

This new discipline will model the Web' structure, articulate the architectural principle that have fueled its phenomenal growth, and dis cover how online human interactions are drive by and can change social conventions. It will elu cidate the principles that can ensure that the net work continues to grow productively and settl complex issues such as privacy protection and in tellectual-property rights. To achieve these ends Web science will draw on mathematics, physics computer science, psychology, ecology, sociolo gy, law, political science, economics, and more.

Of course, we cannot predict what this na scent endeavor might reveal. Yet Web scienc has already generated crucial insights, som presented here. Ultimately, the pursuit aims t answer fundamental questions: What evolu tionary patterns have driven the Web's growth Could they burn out? How do tipping point arise, and can that be altered?

Insights Already

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

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

COMMUNICATIONS
OF THE ACM

ACM.ACM.ORG
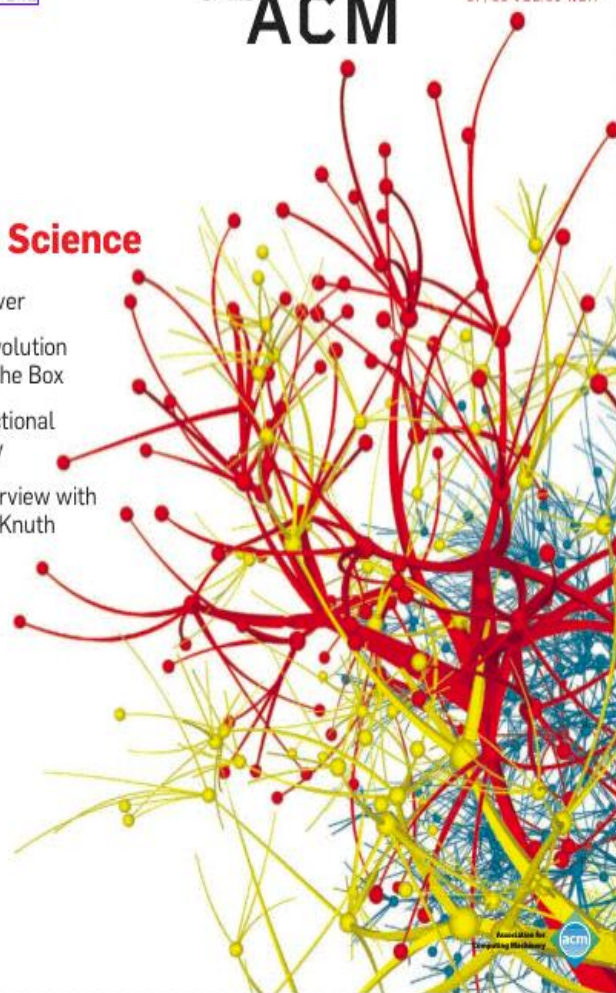
07/08 VOL.51 NO.7

Web Science

XML Fever

The Revolution
Inside the Box

Transactional
Memory

An Interview with
Donald Knuth

Research
Education
Thought Leadership

www.webscience.org

WST Outreach and Thought Leadership

- Publications e.g. Foundations and Trends in Web Science
- Impact on research agenda of funding agencies
- Summer Graduate Schools - OII July 2008, RPI July 2009, Koblenz July 2010
- Conferences
  - Web Science 2009, Athens, 18-20 March 2009
  - Web Science 2010, Raleigh Durham, 26-27 April 2010 (co-located with WWW2010)
  - Web Science 2011, Koblenz 15-17 June 2011
- Research talks and workshops all over the world
- Curriculum development
- Sponsors Forum

research / thought leadershi education

webscience
TRUST

WSTNet announced at WebSci10 in April Founding Laboratories

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

research / thought leadership education

webscience TRUST

Web Science Doctoral Training Centre

Aim – to create a cohort of web scientists

    (a) Develop appropriate research skills,

    (b) Understand /use different disciplines

    (c) Create a coherent community.

    80 students over next 8 years

    50 fully funded by RCUK Digital Economy Programme

    4 year scholarships (1+3)

> **PhD students are the life blood
> of a world-class research lab
> They are the key to innovation**

Digital Economy
Transforming Business and Society

WEB SCIENCE
T R A I N I N G

The future is mobile
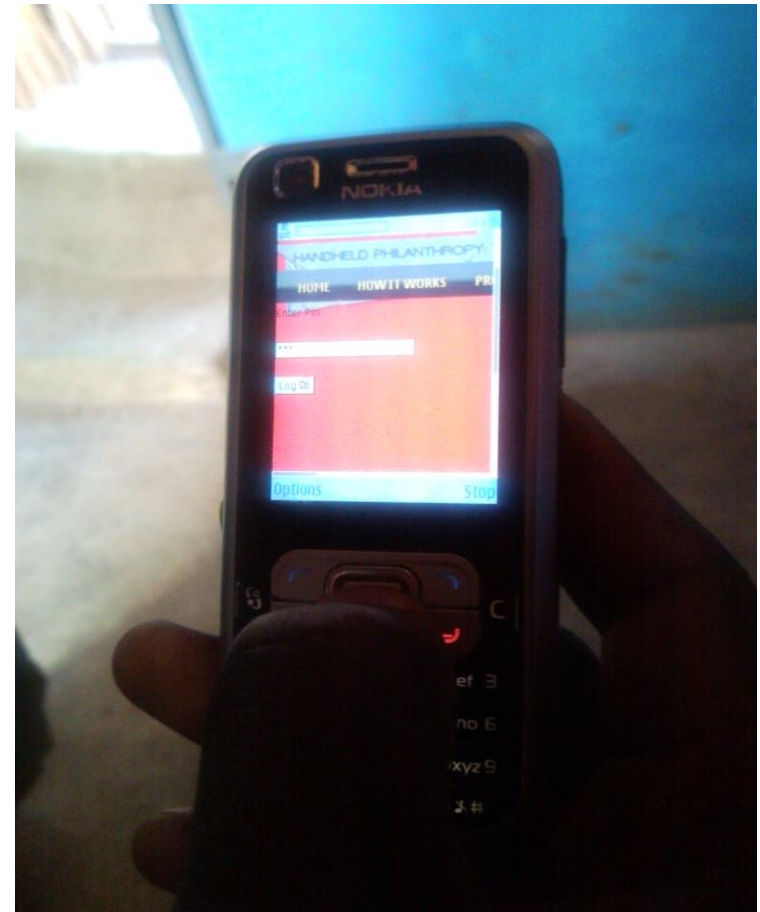
Handheld Philanthropy

research / **thought leadership** / insight

Web Science
why this matters

- the Web matters
- an essential part of humanity but less than 25% 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?
  - could we kill it?
  - 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
- For more information see
  www.webscience.org

research / thought leadershi education