

Web Science

**Professor Wendy Hall
University of Southampton
www.ecs.soton.ac.uk/~wh**

May 2009

Chercheurs Sans Frontières

**Professor Wendy Hall
University of Southampton
www.ecs.soton.ac.uk/~wh**

May 2009

Congratulations to ERCIM

**Professor Wendy Hall
University of Southampton
www.ecs.soton.ac.uk/~wh**

May 2009



ACM Sans Frontières

Wendy Hall
ACM President



Association for
Computing Machinery

Advancing Computing as a Science & Profession



Plans for ACM Europe

ACM has 16,000+ members in Europe

Plans to launch ACM Europe Council chaired by
Fabrizio Gagliardi

Represented on ACM Council

ACM Europe activities – research meetings,
European chapters and SIGs, help nominate
European ACM members for fellowship etc.

Similar plans for ACM India and ACM China



Association for
Computing Machinery

Advancing Computing as a Science & Profession



ACM Europe Launch event (to be confirmed)

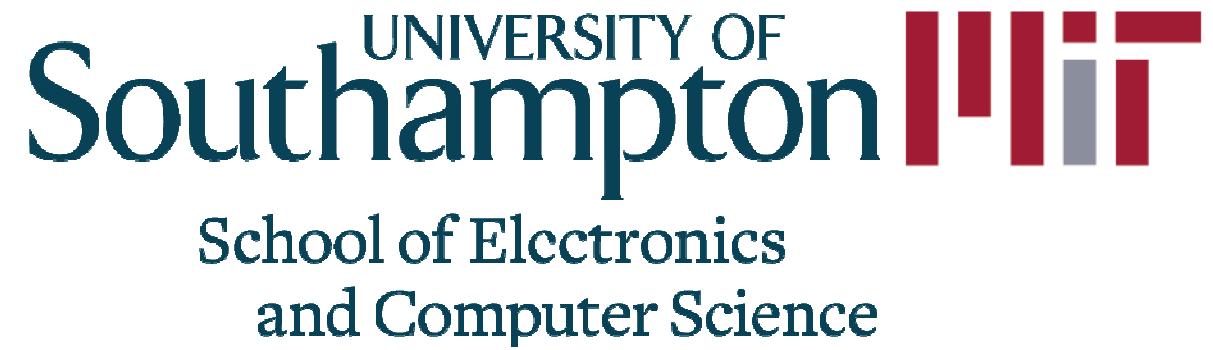
Plans for pre-dinner reception at Informatics
Europe Symposium

October 8th, Paris



Association for
Computing Machinery

Advancing Computing as a Science & Profession



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



Tim Berners-Lee



Wendy Hall



Nigel Shadbolt



Daniel Weitzner

the reputations, experience and skills of our Directors enables us to work closely alongside academia, government, industry and donors to realize our aims

research / thought leadership / insight

Introduction

our aims

promote and encourage multidisciplinary collaborative research
to study the development of the Web

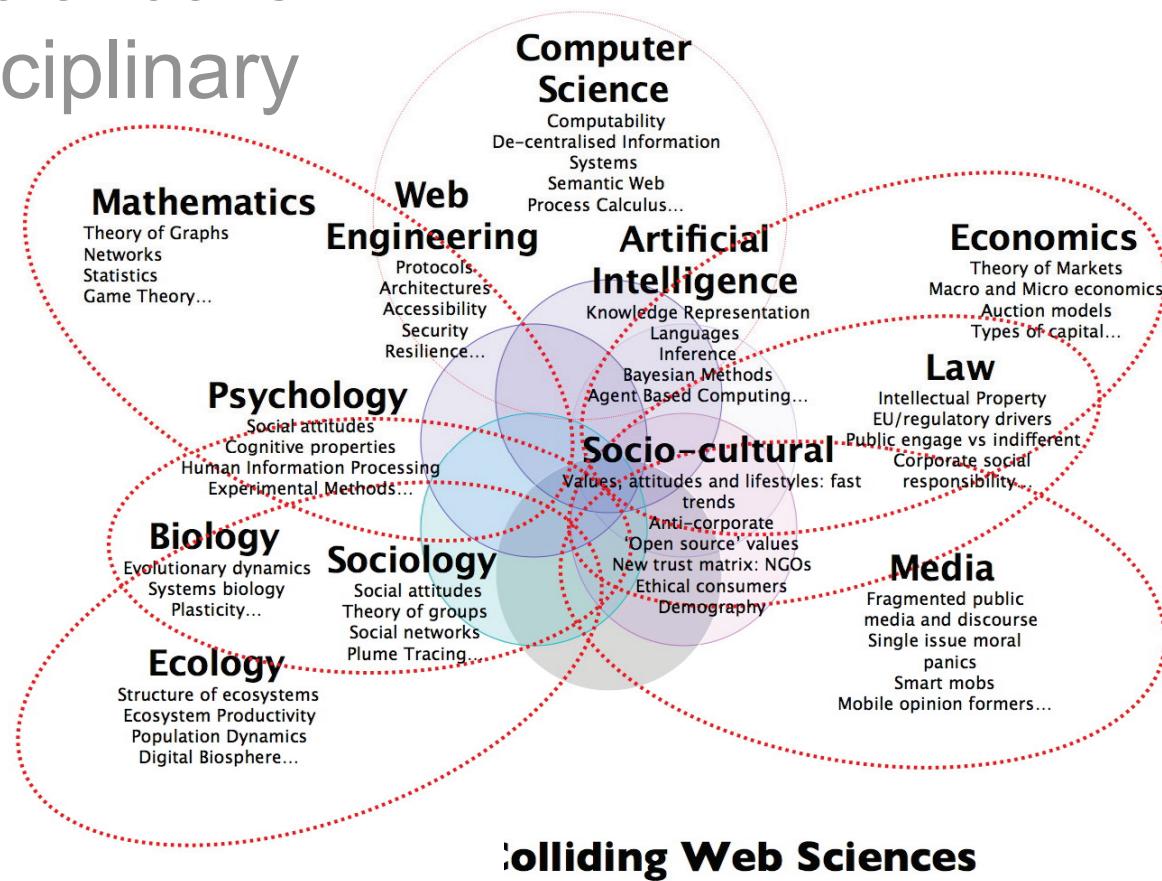
provide a global forum to enable academia, government and
industry to understand the scientific, technical and social
factors that drive the growth of the Web and enable
innovation

devise curricula for the new discipline of Web Science so as to
train future generations of Web Scientists

research / thought leadership / insight



Web Science is interdisciplinary



research / thought leadership / insight

Web Science is about emergence

- simple principles and protocols (e.g. links) create complex structures - the Web
- simple behaviors (e.g. Blogging) create complex phenomena - the Blogosphere
- we will anticipate new principles and behaviors

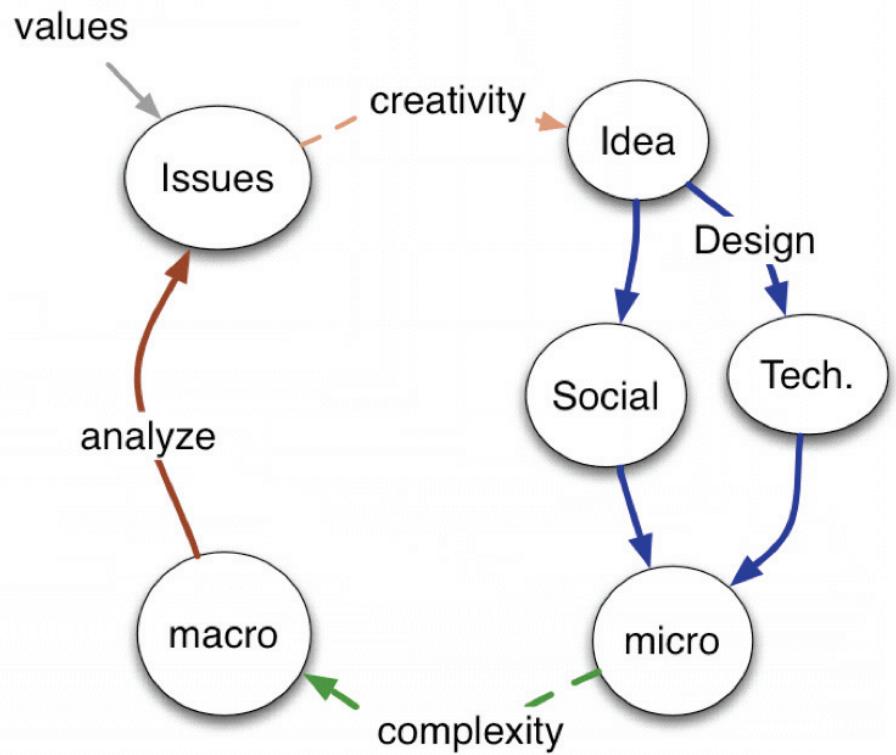
research / thought leadership / insight



Web Science

A Process

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



research / thought leadership / insight

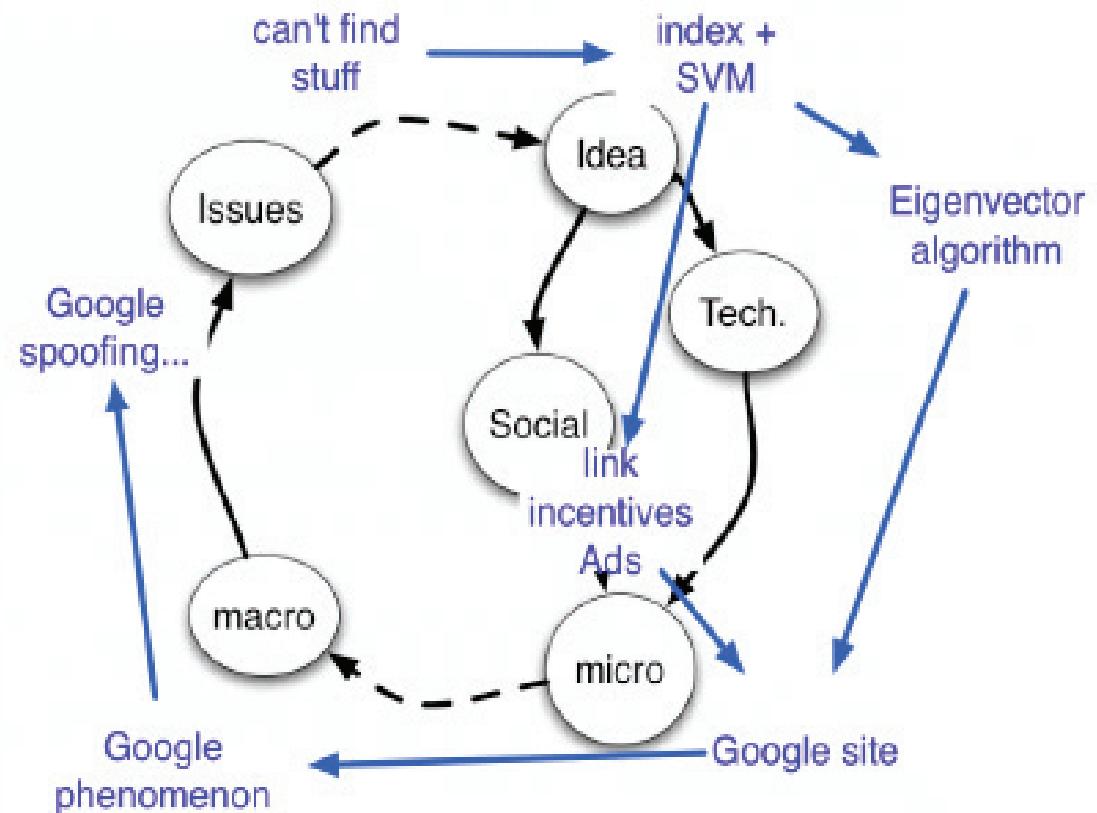
Web Science

A Process

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

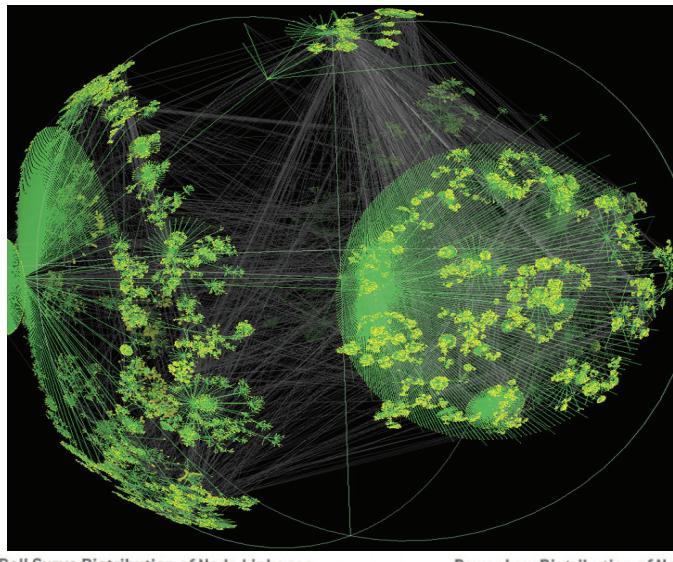
Can we predict,
anticipate before the fact?

research / thought leadership / insight

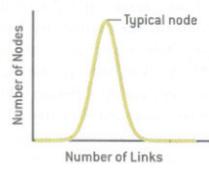


Web Science - Examples

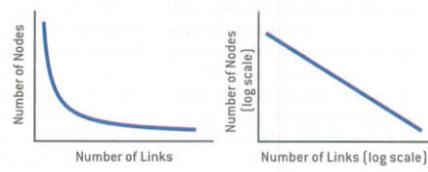
Web Structure



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

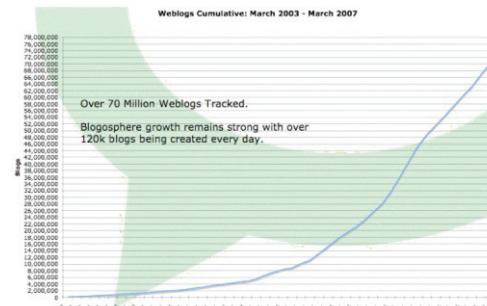
The average distance (or diameter) is much smaller than the order of the graph.

research / thought leadership / insight

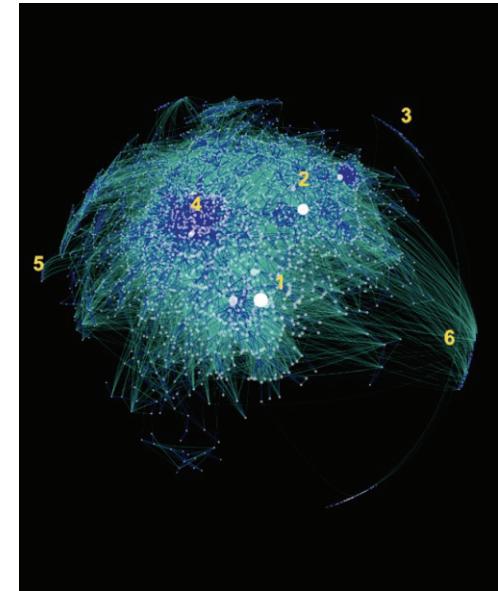
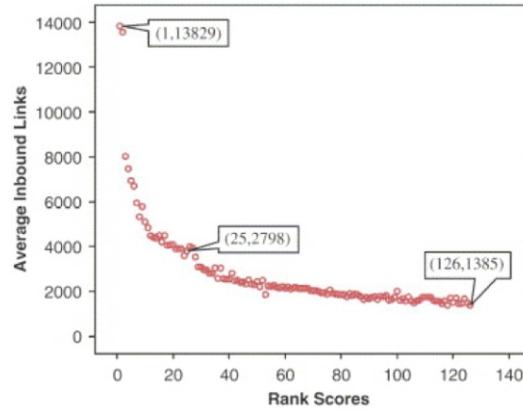
Web Science - Examples

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



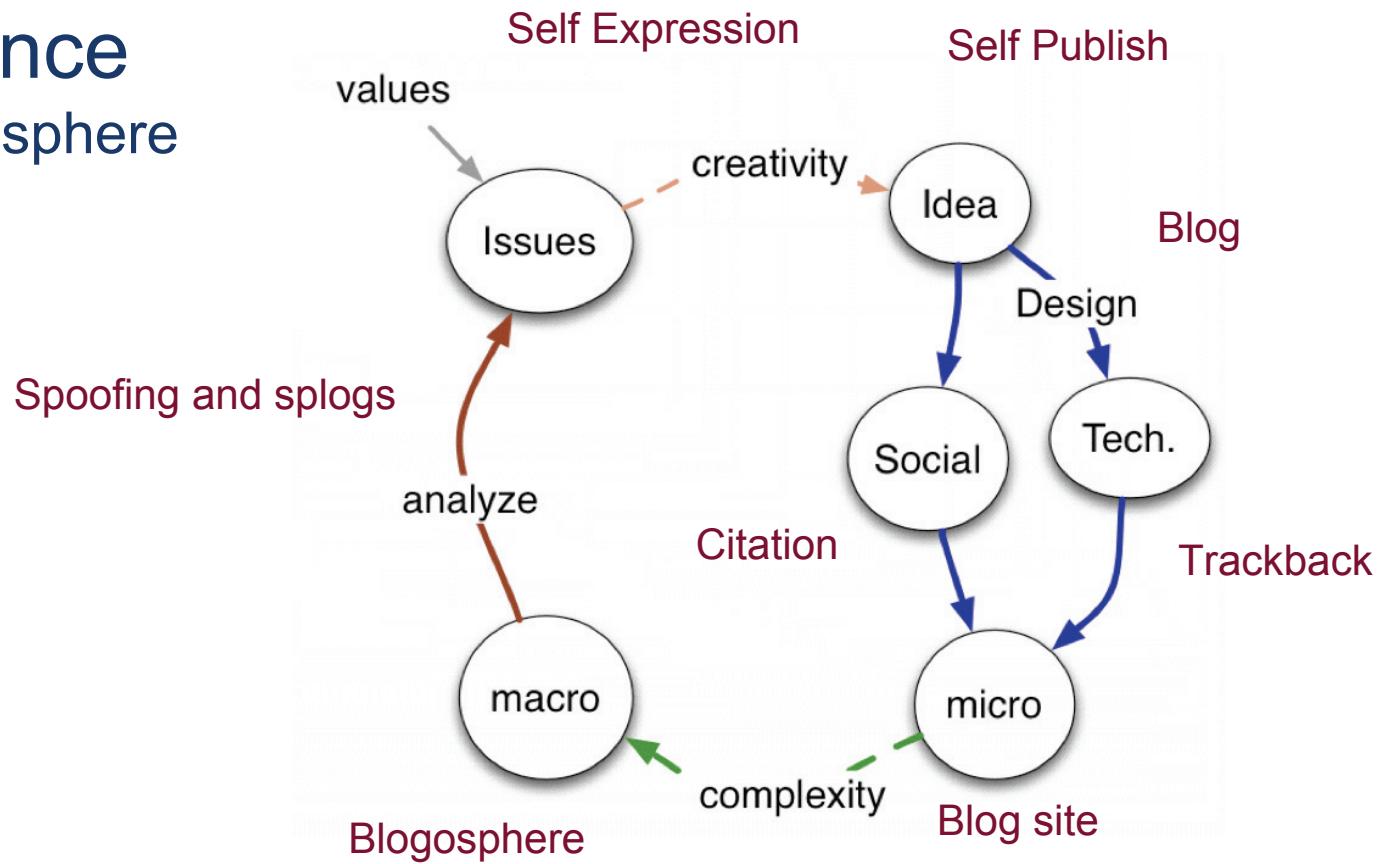
Top Blog sites



research / thought leadership / insight

Web Science

The Blogosphere



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?

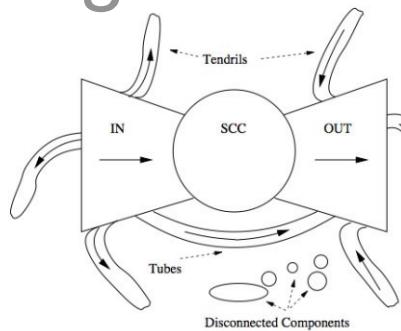
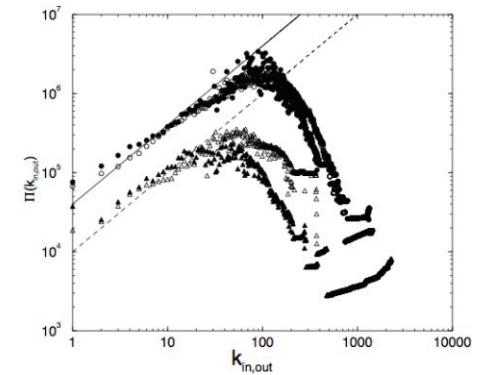


FIG. 1: The shape of the Wikipedia network



| 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

the social web

- social networking sites
- (inter-linked) blogs + comments + aggregators
- community-edited news sites, participatory journalism
- content-sharing sites
- discussion forums, newsgroups
- wikis, Wikipedia
- services that allow sharing of bookmarks/favorites
- ...and **mashups** of the above services

“democratic”, participatory, conversational

Material from Ciro Cattuto

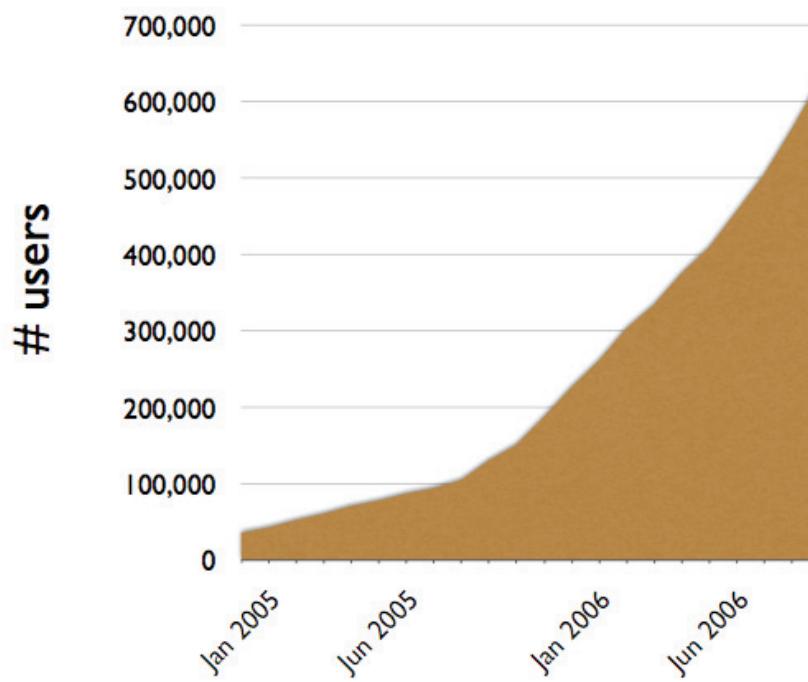
<http://isiosf.isi.it/~cattuto>

research / thought leadership / insight



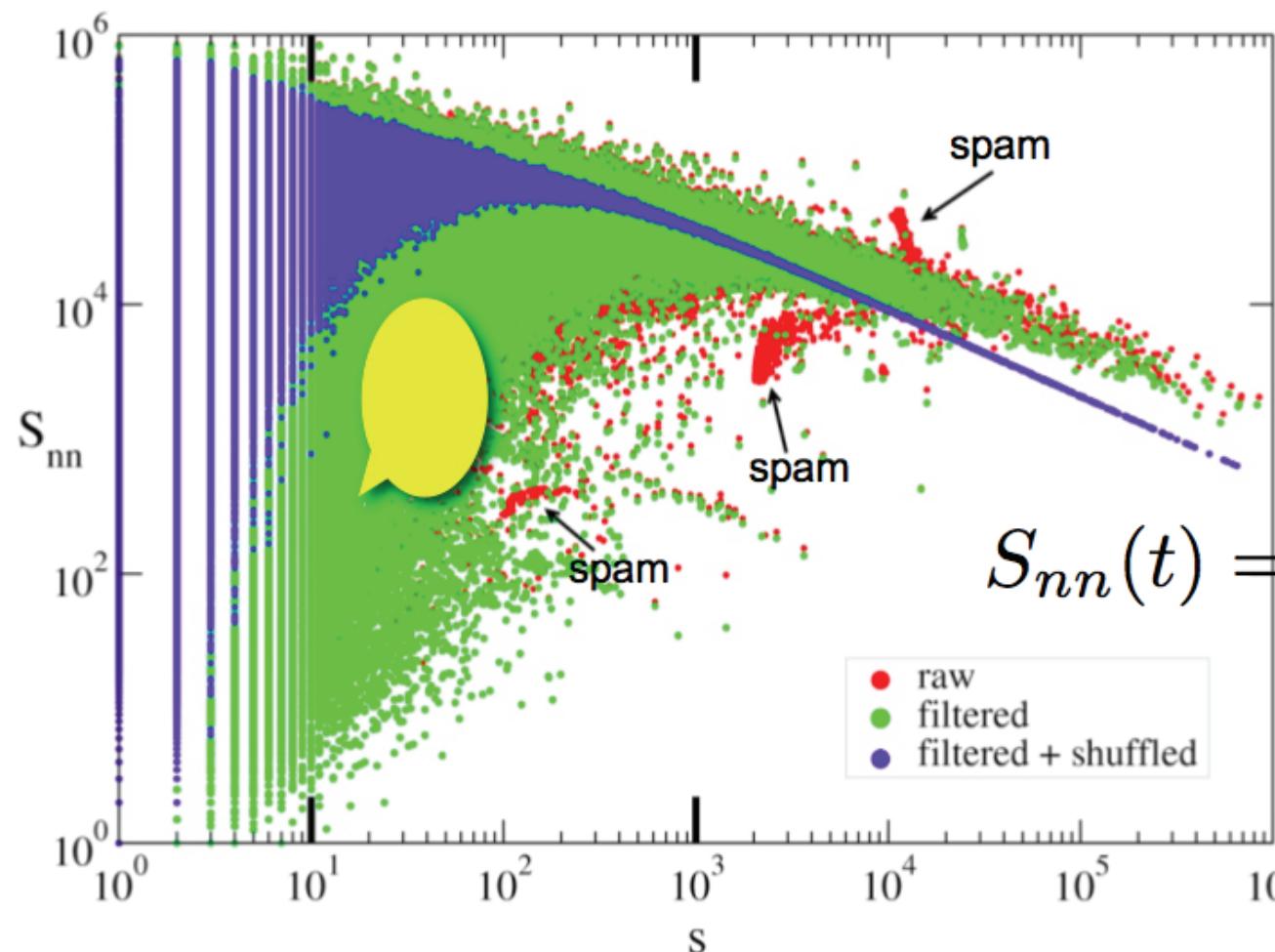
experimental data

- large-scale *del.icio.us* snapshot from distributed crawl
- from beginning of 2004 up to november 2006
- full hypergraph structure + timestamps



~ 650,000 users
~ $2 \cdot 10^7$ resources
~ $5 \cdot 10^7$ posts
~ $3 \cdot 10^6$ *distinct* tags

networks of tag co-occurrence



Material from Ciro Cattuto

CC et al., AI Communications **20**, 245 (2007)

<http://isiosf.isi.it/~cattuto>

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

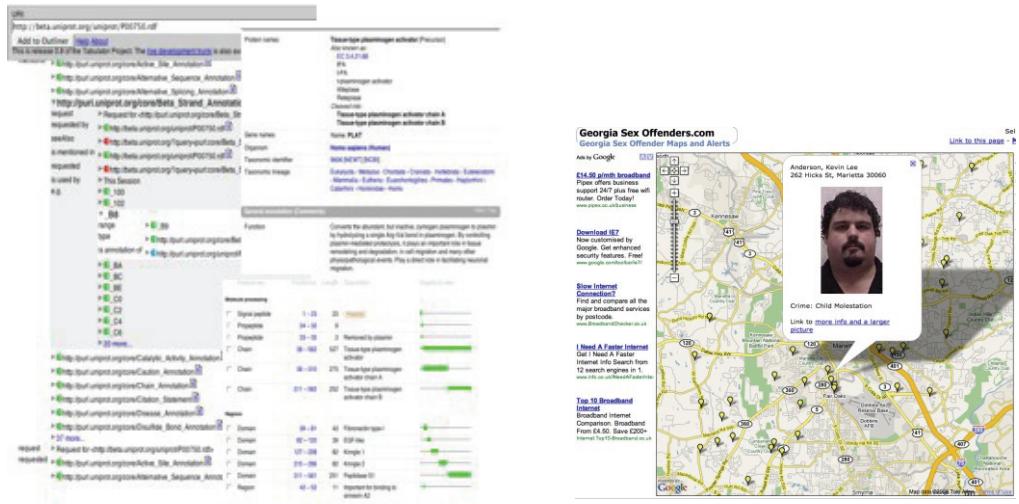


Figure 8 Browsing the Structured Data Web for Proteomics

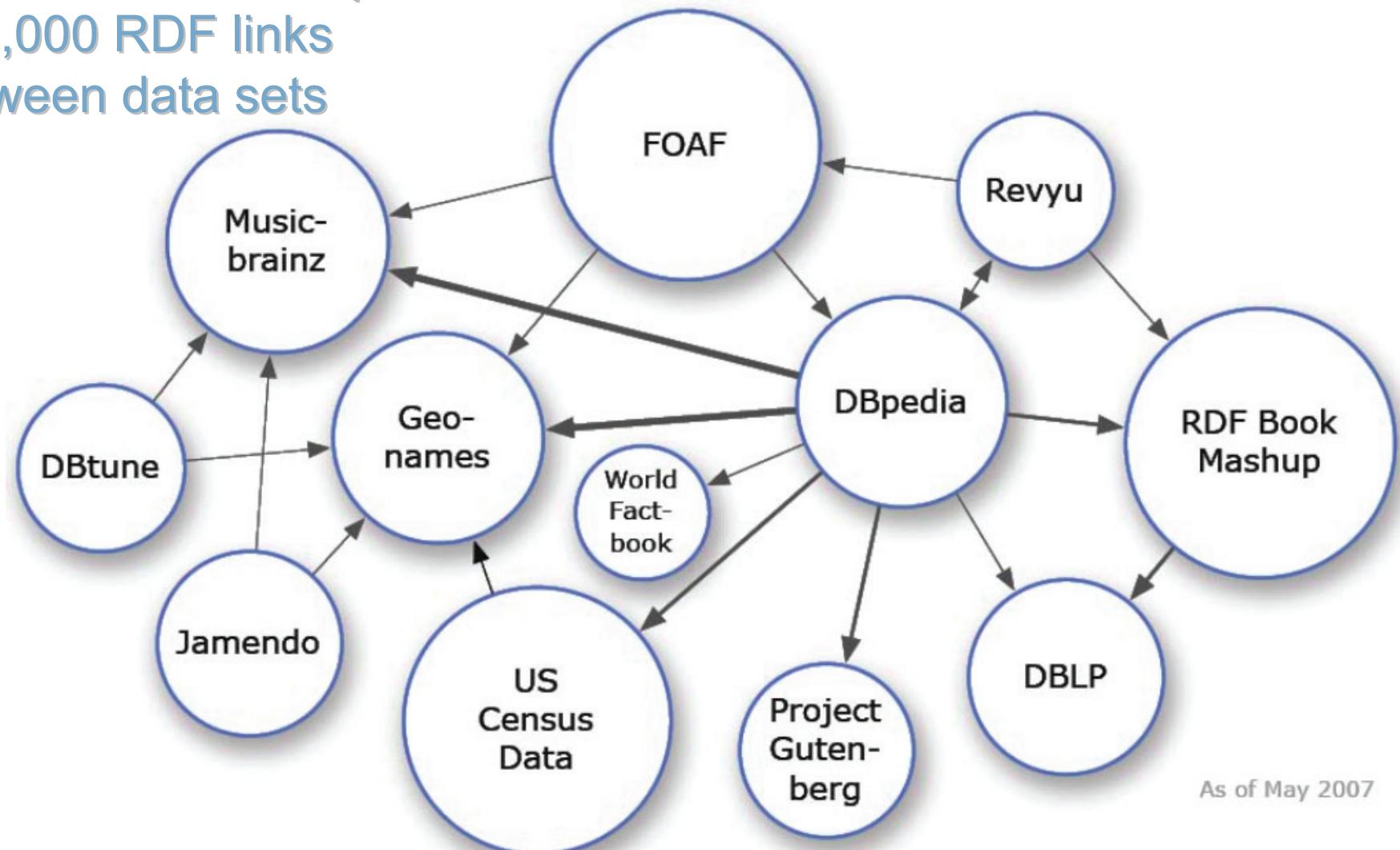
research / thought leadership / insight

Linked Data on the Web: May 2007

500 Million RDF Triples

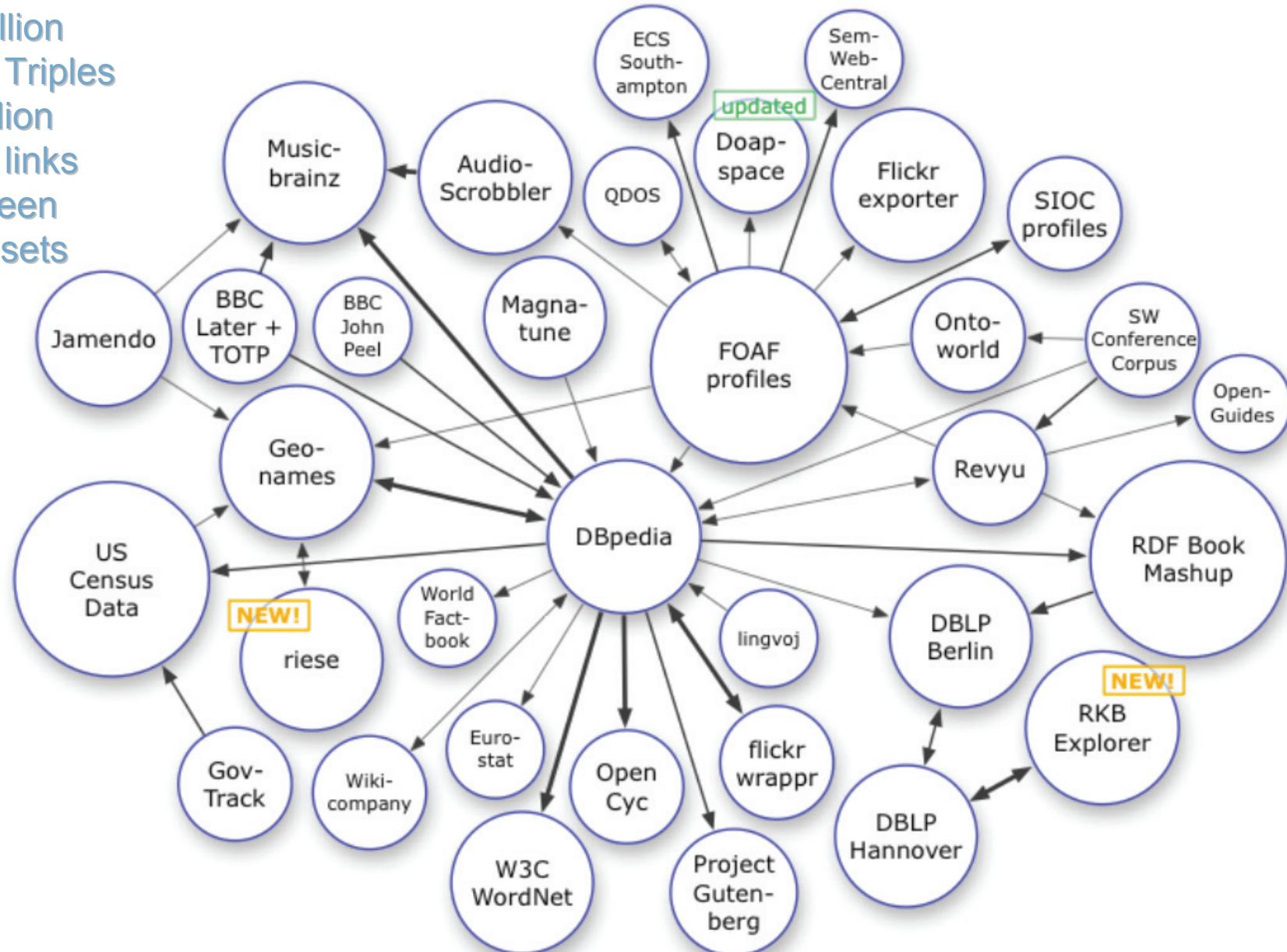
120,000 RDF links

between data sets

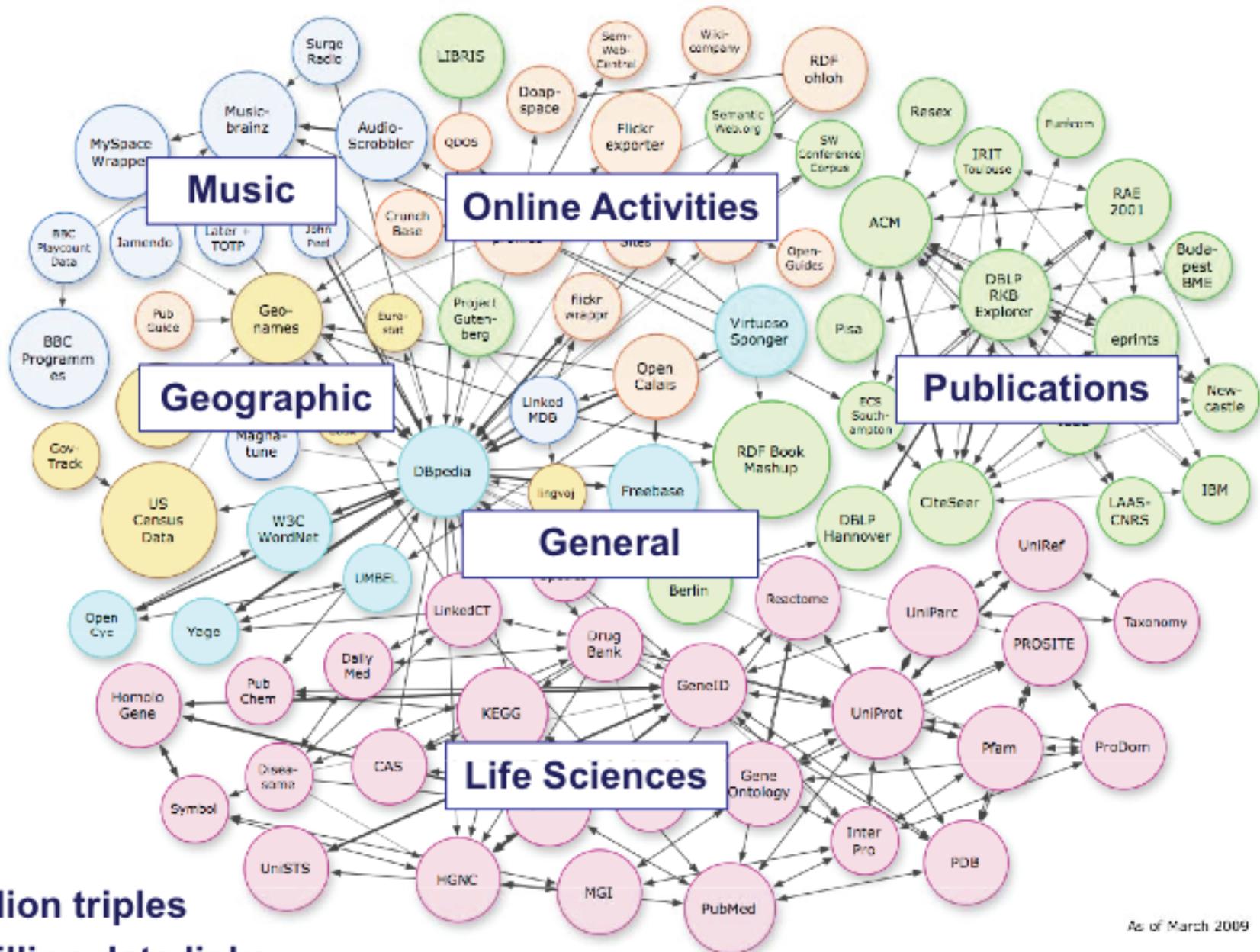


Linked Data on the Web: April 2008

23 billion
RDF Triples
3 million
RDF links
between
data sets



LOD Datasets on the Web: March 2009



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.



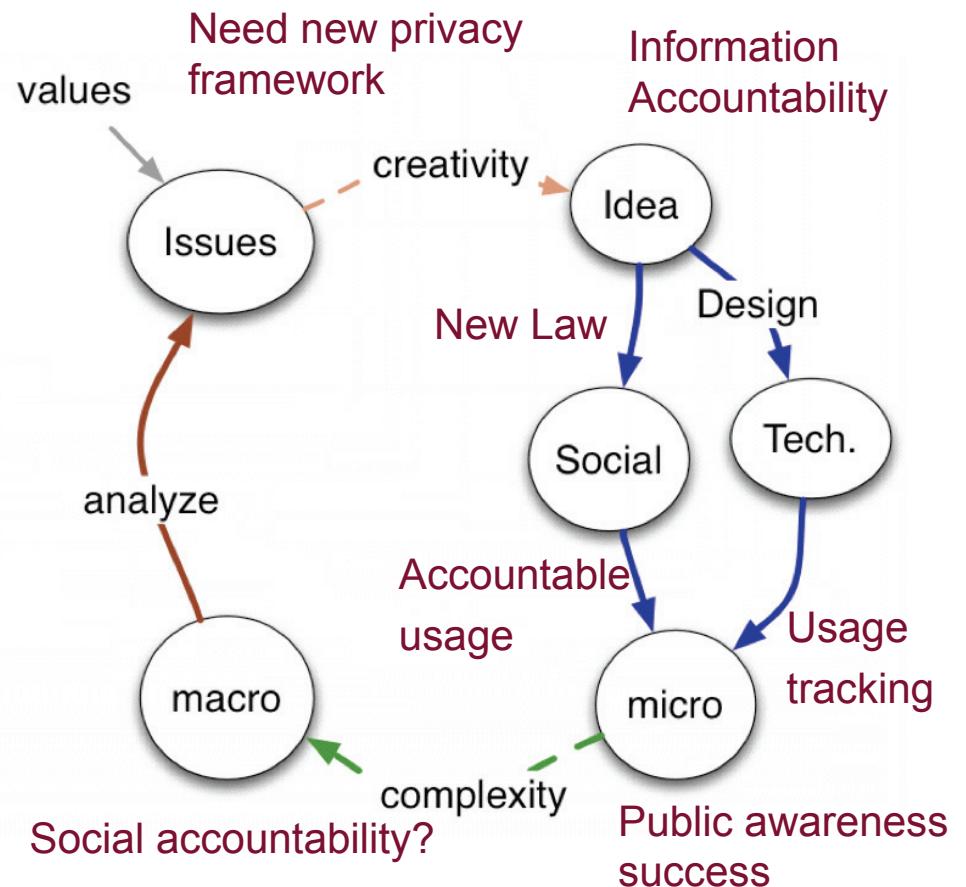
Wendy Hall is a professor of computer science in the School of Electronics and Computer Science at Southampton University. Contact her at wh@ecs.soton.ac.uk.

The Semantic Web Revisited

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

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 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 in computing significant science was launched as a discipline in November 1947 when the two of us and colleagues at the Massachusetts Institute of Technology and the University of Southampton, England announced the creation of a Web Science Research Initiative involving researchers from 16 of the world's universities that have since expanded on that theme.

This new discipline will model the structure, articulate the architectural principles that have fueled its phenomenal growth, cover how online human interactions affect and can change social conventions. It will codify the principles that can ensure that work continues to grow productively as complex issues such as privacy protection, intellectual-property rights. To achieve this, Web science will draw on mathematics, computer science, psychology, ecology, ergonomics, law, political science, economics, art and design.

Of course, we cannot predict what this scientific endeavor might reveal. Yet Web science has already generated crucial insights presented here. Ultimately, the pursuit will answer fundamental questions: What evolutionary patterns have driven the Web? Could they burn out? How do tipping points arise, and can that be altered?

Insights Already

Although Web science as a discipline is still in its infancy, earlier research has revealed the potential of such work. As the 1990s progressed, researchers began looking for information by looking for key words among the mounting number of pages returning more and more irrelevant results. The founders of Google, Larry Page and Sergey Brin, realized they needed to prioritize 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 other pages linking to it. The difficulty was that this definition is recursive: the importance of a page is determined by the importance of the pages linking to it.

research / thought leadership / insight

COMMUNICATIONS OF THE ACM

ACM.ORG

07/08 VOL.51 NO.7

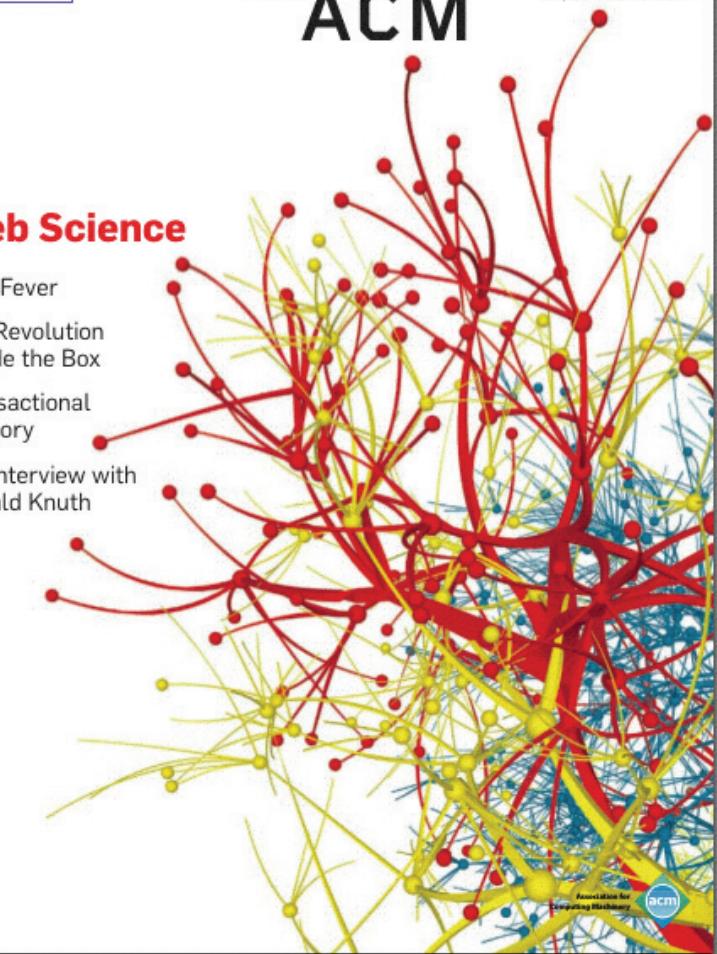
Web Science

XML Fever

The Revolution Inside the Box

Transactional Memory

An Interview with Donald Knuth



WSRI

web science research initiative



WSRI

web science research initiative

Activities

Outreach and Thought Leadership

- Refined Research Agenda with Sci Council (MIT Nov 2008)
- Influence on funding agencies – UK EPSRC (Digital Economy), European Commission, NSF?, China?
- Workshops – WebEvolve2008 @WWW2008
- More workshops in the pipeline for 2009 – scholarship on the Web, critical infrastructure, transparency in journalism, e-government II (Washington, October) and lots more to come
- Web Science 2009, Athens, 18-20 March 2009
- Web Science 2010, Rayleigh-Durham

research / thought leadership /
education



Education

- Doctoral Summer School, joint with the Oxford Internet Institute, July 2008
- Curriculum workshop, September 2008
- Curriculum wiki launched
- Curriculum workshop in Athens @ WebSci'09, more planned
- Doctoral Training Centre @ Southampton

research / thought leadership /
education



WSRI Affiliation Activities

- WSRI Affiliated Labs (WAL's)
- Wider network of Web Science research groups
- Curriculum Development
- WSRI Ambassadors and Evangelists

research / thought leadership /
education



WSRI Affiliated Web Science Labs

- Developing a network of Web Science Labs around the world
- Pursuing a coordinated programme of work
 - Research – annual meeting of research directors
 - Doctoral Summer Schools (21-28 July 2009 @RPI, 2010 @Tsinghua)
 - Curriculum Development
 - Technology Transfer



research / thought leadership /
education

Web Science

why this matters

- the Web matters
- an essential part of humanity
- understanding the Web is a major challenge as big as any other global cause
- www.webscience.org



research / thought leadership /
education