The Development of Web Science

A Global Enterprise

Professor Dame Wendy Hall
28th May 2012
The Web – most successful information architecture in history
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
Web 3.0?
The Semantic Web
A Web of Linked Data
The Semantic Web

Semantic Web LayerCake (Berners-Lee, 99; Swartz-Hendler, 2001)
Why Web Science?
“Web Science represents a pretty big next step in the evolution of information. This kind of research is likely to have a lot of influence on the next generation of researchers, scientists and, most importantly, the next generation of entrepreneurs who will build new companies from this.”

Dr Eric Schmidt, CEO, Google Inc.
www.webscience.org

Exists to promote and help coordinate

– Research

• Research Agenda
• WSTNet Labs

– Education

• Curriculum Development
• Summer Schools
  – DERI 6-13 July 2011
  – Netherlands July 2012

– Thought Leadership

• Workshops
• Dissemination

info@webscience.org
The Web Science Network of Laboratories (WSTNet) combines some of the world’s leading academic researchers in Web Science, with new academic programmes that will enhance the already growing influence of Web Science. The member Labs will provide valuable support for the ongoing development of Web Science. There are now 13 WSTNet labs.

Southampton, UK
MIT, USA
North-Western, USA
Tsinghua Graduate School, Shenzhen, China
DERI, Galway, Ireland
KAIST, Korea
Hannover, Germany

Oxford Internet Institute, UK
RPI, USA
Anaheim School of Communication, USC, USA
VU, Amsterdam, The Netherlands
Koblenz, Germany
Rio, Brazil
The Web is a socio-technical system

Tim Berners-Lee didn’t create the Web
technological determinism
(the way engineers and computer scientists often view things)

- views the web as the product or outcome of scientific advances

- encourages us to see technology (e.g. the web) as ‘a thing’ (a process of reification) with effects on society.

- Focuses on reengineering the web or improving the technology

© Professor Cathy Pope
social shaping (or what we often refer to as co-constitution)

• Sees the relationship between science and technology as far more messy, contingent and complex.

• Suggests that technologies do not have ‘an essence’ (they are not ‘a thing’, they are not fixed) and our relationship with technologies is reciprocal (we shape the web and the web shapes us) i.e. mutually constitutive
The Web – a socio-technical system

Web Science is the theory and practice of social machines
What is a social machine?

“Real life must be full of all kinds of social constraints – the very processes from which society arises. Computers help if we can use them to create abstract social machines on the Web, processes in which people do the creativity and the machine does the administration. The stage is set for an evolutionary growth of new social engines. The ability to create new forms of social process would be given to the world at large and development would be rapid”

*Tim Berners-Lee and Mark Fischetti, Weaving the Web, 1999*
Examples of social machines

• The Web, Google, FaceBook, Twitter, Wikipedia ………

• Captcha (Luis von Ahn)

• Trip Advisor

• Galaxy Zoo

• Ushahidi – open source project which allows users to crowd source crisis information to be sent via mobile

• The OpenStreet Map of Haiti created after the earthquake

• The list goes on … design your own
Social Machines in Context

- More machines
- More people

- Big Data
- Big Compute

- Conventional Computation

- Social Machines
- Social Networking
Web Science across continents

• Astronomers obtain a very high resolution picture of the sky from small telescopes a long distance apart.

• Many web labs, contributing across the globe, help build an accurate picture of human activity at planetary scale.
  – *transcending parochial social, political, economic, legal interpretations*
Web Science
the grand challenge

• Web Science Observatory
• Researchers around the world gathering and sharing data and evidence
• Sharing tools, methods and techniques
• Web Science Collaboratories
• Longitudinal studies
The Web Observatory
a global research platform
facilitated by the Web Science Trust

DRAFT 0.2

Joanna Lewis
Craig Gallen
09-02-12
Opportunity: ‘Data is the New Oil’

‘Data is the new raw material of the 21st century - a resource that gets more plentiful every day. Generated and disseminated by users of the Internet and World Wide Web, a data deluge is changing our world.

Data drives transactions and decisions of every kind. Transport, retail, health, education, leisure, along with every aspect of our lives, depend on an evolving data ecosystem.

The science and engineering of data is fundamental to the modern world.

(source: Open Data: Powering the Information Age, Technology Visionaries lecture series)

We cannot find the ‘needle in the haystack’
— Exponential growth in public data sets
— You cannot trust what you find

The cost of data collection is high
— Like the ‘genome’ project, nobody can afford to do it alone
— Danger that one or two companies hold all the data.

So you’ve got all this data – now what?
— What are the correct, simplest and cheapest ways to analyse the data
— The value is in the analysis – not in the data
— Combining data sets reveals new information & new added value

Lies, Damned lies and statistics
— Need authoritative recommendations
— Need to be able to validate or challenge the conclusions
— Need trusted, transparent independent analysis
— Publishing papers with data validates the research
Opportunity: a critical role for a neutral player?

Locate Data: Finding Data Sets
Where is all the data?
“There must be someone who has this information”
“How does my data compare with others”
“Who would like to help me collect this data”
“If only I could enrich my data with this other information I would have an answer”

Gather Data: Collecting and Storing Data Sets
How do I collect this data?
“Are there any tools someone else has written”
“I would have to build a huge system to collect this information”
“This data is private how do I securely share it”

Process Data: Analysing Data Sets
How do I Analyse this data
“What statistical techniques should I use”
“Is there anyone else who does this type of analysis”
“Is it legal for me to this”
“What tools can analyse data encoded this way”
“Could you just check my results”

Disseminate Data: Reporting Findings
So What does all this mean?
“can someone simplify this for me”
“what do you recommend we do”
“I don’t believe your conclusions can I see the data and how you processed it”