Impact case study (REF3b)

Institution: University of Southampton
Unit of Assessment: 11 Computer Science and Informatics
Title of case study: 11-03 Web Science: Designing a Pro-Human World Wide Web

1. Summary of the impact

Research over two decades at the University of Southampton into the structure and development of the World Wide Web has led to the establishment of a new scientific field, which has earned recognition – and direct funding – from governments and industry around the world. Web Science is the study of the Web as a sociotechnical system. Southampton's work has influenced the Web strategies of the world's biggest companies, including Microsoft, IBM and Google, informed international Web standards and government information policies, led to a network of international laboratories working with industry to advance the Web's development through the provision of highly skilled people taking up specialist roles that draw on their research training.

2. Underpinning research

The World Wide Web is the world's largest and most complex engineered environment. Understanding its development and growth is vital for preserving its capacity for furthering global knowledge and communication, and maintaining a 'pro-human' Web that benefits society.

Research into the architecture of the Web at the University of Southampton stretches back to the early 1990s when Wendy Hall, Professor of Computer Science (at Southampton since 1984) Paul Lewis, Professor of Computer Science (1985-present), Hugh Davis, Professor of Learning Technologies (1987-present), David De Roure, Professor of Computer Science (1986 - 2010) and Leslie Carr, Professor of Web Science (1986-present), developed some of the earliest open hypermedia systems. This gave rise to Microcosm, a link service that separated the link structure from the data in the hypermedia system, making it more easily modifiable and customisable.

The group was joined by Nigel Shadbolt, Professor of Artificial Intelligence (2000-present) and Luc Moreau, Professor of Computer Science (1995-present) and applied these early insights into the Web as an infrastructure for distributed information management to the development of the Semantic Web vision: the conversion of the existing Web into content that can be better interpreted by machines. Shadbolt, Hall and Tim Berners-Lee, the Web's inventor, published an influential reformulation of the Semantic Web into a set of basic principles for a Web of Linked Data **[3.1]** in 2006. A year later in 2007, Berners-Lee was appointed to a part-time Chair at Southampton.

In a paper published in *Science* in 2006, Shadbolt, Hall, Berners-Lee et al articulated their vision for a new research paradigm for the study of the Web **[3.2]**. They called for the establishment of Web Science as the interdisciplinary study of the evolution and impact of the Web, to ensure it supports the basic social values of trustworthiness, privacy and respect for social boundaries. The Web is a highly complex sociotechnical system and understanding its dynamics requires simultaneous study of technology and social engagement, underpinned by the gathering and curation of massive amounts of data **[3.3, 3.4]**.

Working towards this vision, research at Southampton involved the design of methods and tools for the provision of semantics and metadata to support trust and deployment of resources across domains. The £7.5 million Advanced Knowledge Technologies **[Grant 1]** funded by EPSRC (2000 -2007) showed how heterogeneous information could be harvested and integrated **[3.5]**. The £1.94 million follow-on project EnAKTinG **[Grant 2]** developed Linked Data technologies capable of supporting information management on a global scale.

As part of this new interdisciplinary and holistic approach, PASOA **[Grant 3]** and the EU Provenance project **[Grant 4]** led by Moreau, set out to define what provenance – information about entities, activities and people involved in producing a piece of data, which can be used to assess its quality, reliability or trustworthiness – means for the Web and to investigate trust, based on provenance information. It resulted in the first comprehensive open specification of a data model for provenance and related protocols, which led to the Open Provenance Model (OPM) **[3.6]**.

Southampton's Web Science research supported significant contributions to World Wide Web

Consortium (W3C) standards for rich linking (XLINK) the Web Ontology Language (OWL) the SPARQL Query Language and the PROV language (derived from OPM) for representing and reasoning about provenance. In 2009, Southampton's School of Electronics and Computer Science was pivotal in organising the first International Web Science Conference. The Association for Computing Machinery formally adopted Web Science as a research discipline in 2011.

3. References to the research (best three are starred)

3.1   * Shadbolt, N., Hall, W., & Berners-Lee, T. (2006). The semantic web revisited. *IEEE Intelligent Systems*, 21, (3), 96-101.

3.2   * Berners-Lee, T. B., Hall, W., Hendler, J., Shadbolt, N., & Weitzner, D. J. (2006). Creating a science of the web. *Science*, 313(5788), 769–771.

3.3   Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T., & Weitzner, D. (2008). Web science: an interdisciplinary approach to understanding the web. *Communications of the ACM*, 51, (7), 60-69.

3.4   Shadbolt, N., Hall, W., Hendler, J. and Dutton, W.H. (2013) *Philosophical Transactions of the Royal Society A*, "Web Science: A New Frontier". 1987(371).

3.5   Shadbolt, N., Gibbins, N., Glaser, H., Harris, S. and schraefel, m. c. (2004). CS AKTive Space or how we stopped worrying and learned to love the Semantic Web. *IEEE Intelligent Systems,* 19 (3), pp. 41-47.

3.6   * Luc Moreau, Ben Clifford, Juliana Freire, Joe Futrelle, Yolanda Gil, Paul Groth, Natalia Kwasnikowska, Simon Miles, Paolo Missier, Jim Myers, Beth Plale, Yogesh Simmhan, Eric Stephan, and Jan Van den Bussche (2011). The open provenance model core specification (v1.1). *Future Generation Computer Systems*, 27(6):743-756.

Selected grants

[Grant 1] PI Shadbolt, EPSRC Funded Advanced Knowledge Technologies (AKT) IRC £7.5m (2000-07 GR/N15764/01) AKT – rated outstanding scoring 35 out 36 at final review

[Grant 2] PI Shadbolt, EPSRC Funded Large Grant EnAKTinG the unbounded Web of Data £1.94m (2009-2012 EP/G008493/1)

[Grant 3] PI: Moreau, Rana, CI: Walker, EPSRC funded, PASOA: Provenance Aware Service Oriented Architecture 02/2004-10/2007, http://www.pasoa.org/

[Grant 4] PI: Moreau, EU IST Project, Provenance,Enabling and Supporting Provenance in Grids for Complex Problems. Total value €3.1m, 2005-2007, http://www.gridprovenance.org/

4. Details of the impact

Southampton's work has had major impact in a short period, attracting investment from governments, multinational businesses and research institutions. Eric Schmidt, Google CEO, has said: "Web Science represents a pretty big next step in the evolution of information. This … is likely to have a lot of influence of the next generation of researchers, scientists and … entrepreneurs who will build new companies from this." **[5.1]**

Impact on international standards and practice

Southampton created (and now hosts) the Web Science Trust (WST) in 2009, a charity for which Hall and Shadbolt are directors. It supports the global development of Web Science by engaging with industry through a network of world-class laboratories known as WSTNet. It is or has been supported and sponsored by organisations that include BT, IBM, InfoSys, Edelman, Switch, Wipro, NESTA, the Media Standards Trust, the Web Foundation and Highlands and Islands Enterprise.

Building on Southampton's research into Web Science, WSTNet now comprises 15 research labs in the UK, continental Europe, USA, Brazil and Asia **[5.2]** which together employ hundreds of scientists, and have attracted funding to develop research programmes on the impact of the Web in business and society and on innovation on the Web. The WST also coordinates the Web Observatory, a project set up in 2011 to develop a global data resource for the advancement of economic and social prosperity. The Observatory is focused on education, promotion and support

and has engaged technology partners (including Infoys, Edelman, Switch, BT, Twitter, Microsoft, HP), key repositories such as the Internet Archive, and the British Library, and engagement groups for data such as the Open Data Institute. Southampton has coordinated a number of Web Observatory workshops to promote, develop and disseminate the concept around the world.

The WST is influencing teaching practice through its curriculum development programme (with prominent contributions from Carr), which has informed the content of Web Science degrees in the UK, US, China, South Korea and Europe **[5.3]**. This is equipping cohorts of graduates with the sociotechnical skillset required for understanding and developing the Web. In 2009, the University of Southampton secured £6 million in EPSRC funding for a Doctoral Training Centre for Web Science, which has partnerships with about 40 private and public organisations, including IBM, BT, Dow Jones, BBC, The British Library, NHS, SOCA and Ordnance Survey **[5.4]** who are given exclusive early access to latest research developments and the opportunity to recruit outstanding students. The DTC has produced, or is in the process of producing, over 70 highly skilled postgraduates that have taken up specialist roles that draw on their research with some of the most prestigious companies in the world. Examples of organisations hiring our DTC graduates include Price Waterhouse Coopers, UK Civil Service Overseas Territories Directorate, Ordnance Survey, White Space Analysis (market analysis consultants) and Switch Concepts (2nd fastest growing tech company in UK, Sunday Times ranking 2013).

Southampton's research into the basic foundations of the Web has led to its central involvement in the work of the World Wide Web Consortium (W3C), an international community co-led by Berners-Lee to develop standards that will allow the Web to reach its full potential. Contributions outlined in section 2 have put many Southampton innovations at the core of the linked data Web. The W3C PROV Working Group, co-chaired by Moreau, has produced standards whose core constructs are derived from OPM, and has over 50 participants from academia, government and industry, demonstrating widespread and international agreement as to the importance of this work **[5.5]**. Participants include the Oracle Corporation, IBM, the Mayo Clinic, TopQuadrant, Revelytix Inc, the National Archives, the Library of Congress and NASA, several of which have adopted the standard specifications in their products.

Impact on the economy

Web Science at Southampton, particularly the areas of the Semantic Web, Linked Data and Provenance, has engaged sectors ranging from medicine to defence, leading to substantial engagement with IBM, ARA, General Dynamics, QinetiQ, Fraunhofer Gesellschaft among others. It has also created commercial applications and spin-out businesses. Garlik, a company that helps consumers protect themselves from the risk of identity theft and fraud, was spun out of the Semantic Web research at Southampton; the company employed 18 full time employees and had a turnover of £2.3 million. Garlik was awarded Technology Pioneer status by the Davos World Economic Forum and won the UK national BT Flagship IT Award in 2008. By December 2011 it had over half a million users and was acquired by Experian, the global information services company **[5.6]**. Oracle has commercialised a product that extends PROV with a 100-strong team of developers, led by a member of the W3C Working Group. Oracle described it as "a significant achievement for Oracle to implement this standard and demonstrates Oracle's leadership position in the technology industry" **[5.7]**. Technology firms Clark & Parsia **[5.11]** and 3 Round Stones **[5.12]** have also implemented PROV in their linked data platforms.

Impact on public policy and debate

The UK government has adopted linked data standards; the commitment to use linked data is embedded in the public data principles recommended by Southampton **[5.8]**. Linked data standards underpin the UK's lead in Open Data - the subject of a separate impact case study from Southampton. Use of linked data produced applications such as the National Archives' legislation.gov.uk and DCLG's opendatacommunities.org, which serve to improve the efficiency and transparency of governance. The London Gazette is published using linked data, while PROV is mandated for all UK Official Gazettes (the government journals of record) to log the creation and processing of all new artefacts; this follows the Stationery Office, which launched a data enrichment service that generates OPM-compliant provenance.

In 2010, Shadbolt and Hall were invited to co-organise *Web Science: a new frontier*, one of ten

public discussion meetings held to celebrate the 350th anniversary of the Royal Society **[5.9].** Hall and Shadbolt edited a special issue of *Philosophical Transactions A*, with the proceedings, including contributions from top researchers in the field of Web research and analytics. Shadbolt acted as Series Consultant to the BBC's 2010 BAFTA Award-winning Virtual Revolution documentary series about the Web **[5.10]**.

Impact on the Environment

The 2013 National Climate Assessment report to US Congress makes complete provenance using PROV specifications available to the public. This includes reference to a huge author team and over 550 direct technical inputs (papers, datasets, graphs, etc.) which each have their own provenance tracing back to even more data, research, models, analyses, sensors, satellites, etc. Public access to that provenance information will be made through API and human browsable interfaces to PROV information. NASA is involved in compiling this report, and has a representative **[5.7]** on the W3C Provenance working group. NASA has also incorporated OPM into its provenance encodings for its environmental satellite programme.

5. Sources to corroborate the impact

5.1   Quote from Eric Schmidt: http://www.nytimes.com/2006/11/01/technology/01iht-compute.3360734.html?_r=0

5.2   Link to WSTnet research labs: http://webscience.org/WSTNet.html

5.3   Links to curricula wikis and programme pages: http://wiki.websciencetrust.org/w/Curriculum

5.4   DTC industrial board: http://dtc.webscience.ecs.soton.ac.uk/industry-partners/ and EPSRC DTC grant: PI Hall EPSRC Funding Doctorial Training Centre in Web Science £5.99M (2009-2018 EP/G036926/1)

5.5   Link to PROV W3C Standards: http://www.w3.org/TR/prov-overview/

5.6   CEO, Garlik
http://semanticweb.com/experian-acquires-garlik-ltd_b25580

5.7   Vice President of Development at Oracle. See also link to industrial engagement with PROV:
https://blogs.oracle.com/FinancialsMkting/entry/oracle_advanced_controls_uptakes_w3c

5.8   Published Letter to Francis Maude, MCO at start of Coalition Governments Transparency Board http://data.gov.uk/sites/default/files/Transparency%20Board%20-%20letter%20from%20Nigel%20Shadbolt%20to%20MCO%2014.06.10.pdp__0.pdf   and the resultant UK public data principles http://data.gov.uk/blog/public-data-statement-of-principles

5.9   Royal Society Discussion Meeting http://royalsociety.org/events/2010/web-science/

5.10   The Virtual Revolution http://www.bbc.co.uk/virtualrevolution/interviews.shtml and series consultant credit http://www.imdb.com/name/nm3801047/

5.11   Senior Research Scientist at Clark & Parsia, clarkparsia.com

5.12   Lead Software Engineer at 3 Round Stones, 3roundstones.com