The Evolution of the Web and Implications for e-Research

Wendy Hall, David De Roure and Nigel Shadbolt
School of Electronics and Computer Science
University of Southampton

Abstract

The hypertext visionaries foresaw the potential of richly inter-linked global information systems for advancing human knowledge. The Web provided the infrastructure to enable those ideas to become reality, and it quickly became a platform for collaborative research and data sharing. As the Web has evolved, new ways of using it for e-Research have emerged, such as the social networking facilities enabled by Web 2.0 technologies. The next generation of the Web – the so-called Semantic Web – is now on the horizon, which will again enable new types of collaborative research to emerge. If we are to understand and anticipate these new modes of collaboration we need a discipline that studies the Web as whole. Web Science is this discipline.

Keywords

Hypertext, Web 2.0, Semantic Web, e-Research, Web Science

Introduction

Long before the Web existed hypertext visionaries foresaw a richly inter-linked global information network. The Web provided the infrastructure to enable those ideas to become reality. Nevertheless the Web remains a difficult environment in which to create meaningful links. Websites are notoriously difficult to design and maintain and we rely on the search engines to navigate our way around hyperspace. As the Web has evolved to support increasing collaboration we have seen the growth of Web 2.0 technologies such as social networking. Today the Web is used by millions around the world to link to communities of people with whom they have common interests. But richly connected information environments are still difficult to set-up and manage.

Researchers across all disciplines are taking advantage of new technologies to do new research. Much of this user-centred activity is drawing on the Web as a distributed application platform, with “mash-ups” for integration, easy access to computational resources "in the cloud", and social networking to share the results and practice of digital science. The Semantic Web will enable new developments in this respect and will continue the trend of technologies empowering the individual. We are seeing an evolution from the current Web of documents towards a Web of linked data and the broad benefits this brings. Once again, those using the Web for scientific endeavour are the pioneers of the Web’s evolution.

However, there is a growing realization among many researchers that if we want to model the Web and understand this future trajectory, if we want to understand the architectural principles that have provided for its growth, and if we want to be sure that it supports the basic social values of trustworthiness, privacy, and respect for social boundaries, then we must chart out a research agenda that targets the Web as a primary focus of attention. The emergence of this exciting new discipline, which we call Web Science, is discussed at the end of this paper. We argue that by studying the evolution of the Web and developing new methodologies for understanding this, we will better understand its forward trajectory: e-Researchers using the Web provide a case study in its evolution, and Web Science itself will better enable researchers to take advantage of the platform it offers for new types of e-Research[1].

A brief history of hypertext and the Web

The terms hypertext and hypermedia are often used quite interchangeably. Hypertext in the strict sense only applies to text-based systems; hypermedia is simply the extension of hypertext to include multimedia data. The invention of both terms is credited to Ted Nelson in 1965. His vision of a universal hypermedia system, Xanadu, is most fully explored in his book "Literary Machines" [Nelson, 1981]. Nelson defines hypertext as *non-sequential writing* and views hypertext as a literary medium, but the ideas the term encapsulates are wider than that and include cross-referencing and the association of ideas. Nelson acknowledges that his ideas came from the writings of Vannevar Bush and the pioneering work of Douglas Engelbart.

Bush, who was scientific advisor to President Roosevelt during the Second World War, proposed a theoretical design for a system that we would now call a hypertext system [Bush, 1945]. He foresaw the explosion of scientific information and predicted the need for a machine to help scientists follow developments in their discipline. Bush called his system Memex (memory extender) which he described as "a sort of mechanised private file and library". Bush talked of *trails*, which users build as they move through the information so that their paths of discovery can be saved and recalled later or passed on to other researchers.

Douglas Engelbart, one of the early pioneers in the computing industry, is credited with the invention of word processing, screen windows and the mouse and thus inspiring the developments in graphical user interfaces that have taken place over the last twenty years. In 1962 Engelbart, working at Stanford University, started work on his Augment project [Engelbart, 1963]. He anticipated a world of instant text access on screens, interconnections that can be made and shared, a new style of shared work amongst colleagues, and the use of computers to augment the human intellect.

Apple's release of HyperCard free on every Macintosh computer in 1987 did more to popularise hypertext in the late 1980's than any other event. HyperCard introduced the concepts of hypertext to the computer-using community at large. It ceased to be solely a topic for research and became a widely accepted technique for application development, particularly in education. By the early 1990's, many new products on the market claimed some sort of hypertext or hypermedia functionality.

Meanwhile, the hypertext research community was continuing to explore the development of hypertext systems that handled information on a large scale and in distributed environments. Tim

[1] The UK's Joint Information Systems Committee (JISC) defines e-Research as the development of, and the support for, information and computing technologies to facilitate all phases of research processes. The term e-e-Research originates from the term e-Science but expands its remit to all research domains not just the sciences. It's concerned with technologies that support all the processes involved in research including (but not limited to) creating and sustaining research collaborations and discovering, analysing, processing, publishing, storing and sharing research data and information.

Berners-Lee, the inventor of the Web, first started working on its development at CERN, the high-energy physics laboratory in Switzerland, in 1989, although he had been building hypertext systems long before this [Berners-Lee, 1999]. The aim of the project was to provide a distributed hypertext environment to enable physicists to easily share and distribute information. Main features of the design included ease of use, accessibility from anywhere and the provision of open protocols.

The open protocols on which its client-server model is based - Hypertext Transfer Protocol (HTTP) and Hypertext Mark-up Language (HTML) - were the cornerstones of its success. The original Web viewer at CERN worked over line-oriented telnet connections, meaning it could be used essentially from any computer in the world. Early viewers implemented at CERN were also editors, which enabled easy creation of HTML documents by users. The introduction of the graphical Mosaic browser from NCSA (the National Centre for Supercomputing Applications at the University of Illinois at Urbana-Champagne) was critical to the Web's success.

The growth of the Web has since been phenomenal. It now impacts every aspect of the way we live and work and has the potential to significantly change our culture and society. As the Web grew over a certain size it became increasingly difficult to find information simply by following hypertext links or keeping a list of useful websites. A new technology – search engines – was needed. Early search systems were based on the frequency of the occurrence of search terms on Web pages. The innovative algorithm on which Google was built [Brin & Page, 1998] determined the most relevant pages by estimating the importance of Web pages containing the user's search terms. Now of course it is difficult to imagine using the Web without search technology.

In line with Bush's vision, the Web has changed how we carry out research. We now assume in most subject fields that we will be able to find any recent publication on the Web – either directly from the publisher or via the author's website or appropriate open access repository. This has dramatically changed the research culture and is driving developments to create integrated research repositories which can be analysed to present a comprehensive picture of the latest state of the art in the various research fields including detailed citation analyses.

Generally we use a combination of the Web and our preferred search engine as our first port of call to find and share information in today's research world. This is enabling multidisciplinary, interdisciplinary and collaborative research work on a global scale and at a tempo not possible for earlier generations of researchers. We are increasingly using Web 2.0 environments to do this. Web 2.0 is defined in Wikipedia, which of course is a free on-line encyclopaedia developed using Web 2.0 technology, as follows

> ***Web 2.0*** is a term describing the trend in the use of World Wide Web technology and web design that aims to enhance creativity, information sharing, collaboration and functionality of the web.

Web 2.0 technologies support the generation and sharing of user generated content (UGC). Meanwhile, the deluge of information and data from users and enterprises, individuals and groups, humans and machines has been a driving force behind another important set of developments collectively referred to as the Semantic Web.

The Semantic Web

The original *Scientific American* article on the Semantic Web appeared in 2001 [Berners-Lee et al, 2001]. It described the evolution of a Web that consisted largely of documents for humans to read to one that included data and information for computers to manipulate. The Semantic Web is a Web of actionable information—information derived from data through a semantic theory for interpreting the symbols. The semantic theory provides an account of "meaning" in which the logical connection of terms establishes interoperability between systems. This was not a new vision. Tim Berners-Lee articulated it at the very first World Wide Web Conference in 1994.

A Web of data and information would look very different from the Web we experience today. It would routinely let us recruit the right data for a particular use context—for example, opening a calendar and seeing business meetings, travel arrangements, photographs, and financial transactions appropriately placed on a time line. The *Scientific American* article assumed that this would be straightforward, but it's still difficult to achieve in today's Web. The article included many scenarios in which intelligent agents and bots undertook tasks on behalf of their human or corporate owners. Of course, shopbots and auction bots abound on the Web, but these are essentially handcrafted for particular tasks; they have little ability to interact with heterogeneous data and information types. Because we haven't yet delivered large-scale, agent-based mediation, some commentators argue that the Semantic Web has failed to deliver [McCool, 2005].

In a more recent paper [Shadbolt et al, 2006] we argue that agents can only flourish when standards are well established and that the Web standards for expressing shared meaning have progressed steadily over the past five years. This is crucial as researchers are beginning to build a linked Web of data and information.

The basic building blocks of the Semantic Web are the Resource Description Framework (RDF), Universal Resource Identifiers (URIs), triple stores and ontologies. The original Web took hypertext and made it work on a global scale; the vision for RDF was to provide a minimalist knowledge representation for the Web. It provides a simple but powerful triple-based representation language for the URIs which enable the identification of resources because they have a global scope and are interpreted consistently across contexts. Associating a URI with a resource means that anyone can link to it, refer to it or retrieve a representation of it. URI's allows machines to process data directly enabling the shift from a Web of documents to a Web of data.

In February 2004 RDF Schema (RDFS) became a W3C recommendation. It took the basic RDF specification and extended it to support the expression of structured vocabularies and simple ontologies. As RDF and RDFS have gained ground, the need for repositories that can store RDF content has grown. These so-called triple stores vary in their capabilities and the key to their successful use is the recent availability of SPARQL (Simple Protocol and RDF Query Language, http://www.w3.org/TR/rdf-sparql-query) which provides reliable and standardized data access into the RDF they hold[2].

The final building blocks are the ontologies that provide the common conceptualizations to enable data integration and for which there is now an agreed representation standard OWL (Web Ontology

[2] Techniques and methods have also been developed to enable SPARQL queries to traditional Relational Databases.

Language, http://www.w3.org/TR/owl-features). Today the increasing use of ontologies in the e-Science community is presaging the ultimate success for the Semantic Web—just as the use of HTTP within the CERN particle physics community led to the success of the original Web. One important incubator for the Semantic Web has been the life sciences, where research needs demand the integration of diverse and heterogeneous data sets that originate from distinct communities of researchers. Now many other disciplines are adopting this approach. For example environmental science is looking to integrate data from hydrology, climatology, ecology and oceanography.

The need to understand systems across ranges of scale and distribution is evident across a broad range of scientific disciplines and presents a pressing requirement for data and information integration. It is not exclusive to science of course. The requirement to integrate diverse information resources can be seen in engineering and education, in the private and public sectors. Methodologies for developing Semantic Web applications are now well understood [Alani et al, 2008]. But what is interesting about the evolution of the Web is that as more and more communities start developing applications using the new evolving technologies, hitherto unforeseen consequences emerge that have a profound impact on those very same communities. We can see this reflected in the way the e-Science community has evolved – a case study we discuss in detail in the next section.

The Web and e-Science

The UK e-Science programme [Hey & Trefethen, 2002] led the world in creating a coordinated, multidisciplinary research and development programme utilising an emerging set of technologies that were set to enable large scale collaboration and resource-sharing. Distinctively data-centric, the programme focused on handling the data deluge that was enabled by new parallel and high throughput experimental practices, from sensor networks in the environment and earth observation to DNA microarrays and combinatorial chemistry. In particular, the technologies of Grid computing have proven to be an important part of the e-Science infrastructure, enabling distributed data and computational resources to be combined in 'virtual organisations' in order to process the large data volumes, models and simulations.

Early in 2001, a number of researchers working at the intersection of the Semantic Web, Grid and software agent research and development communities were increasingly conscious of the gap between the aspirations of e-Science and the then-current practice in Grid computing. This was captured in the "Semantic Grid" report, presenting a research agenda for the e-Science infrastructure which drew not just on Grid computing but also Web, Web Services, software agents and Knowledge Technologies [De Roure et al, 2005]. The Semantic Grid initiative has seen considerable activity through e-Science projects and workshops in the intervening years, demonstrating the value of a Semantic Web approach working with e-Science data and also in enabling an increasing scale of automation in the infrastructure [Goble et al, 2006].

A particularly powerful aspect of Semantic Web in e-Science was explored by the CombeChem e-Science pilot project, which built a "Semantic DataGrid" by using Semantic Web principles to describe data, its context and its provenance [Taylor et al, 2005]. The underlying maxim of CombeChem was that data on its own is meaningless – rather it is necessary to record an interlinked provenance trail from laboratory bench to scholarly output, so that the data can be interpreted, reused and trusted. CombeChem demonstrated the power of using shared identifiers to interlink data. Just as significantly, CombeChem took a holistic view of the scholarly knowledge cycle and demonstrated how data in repositories can be interlinked with scholarly output [Duke et al,2005], establishing an ethos of publishing data for reuse rather than warehousing it within a project. In many ways, CombeChem demonstrated the power of Semantic Web in bringing scientific data to the Web, through applying hypertext thinking in the context of science. Parallel to this, the myGrid pilot project brought Semantic Web to bear on Web Services, another key evolution in the Web and an important intersection with Grid computing, through building tools to facilitate their use in the field of bioinformatics [Wolstencroft et al, 2007].

The successes of e-Science and Grid computing have tended to focus on large projects with coordinated infrastructure, such as [Droegemeier, 2008]. The Wikipedia definition of e-Science notes: "Due to the complexity of the software and the backend infrastructural requirements, e-Science projects usually involve large teams managed and developed by research laboratories, large universities or governments". However, the individual scientist is also experiencing a transformation in their practice, as research is increasingly conducted using digital techniques within very many disciplines. Given the high complexity of learning to use the Grid or the Semantic Web, some scientists are turning to an array of easy-to-use web-based tools which give immediate benefit to their work. Significantly many of these are collaborative tools, an area recognised by the e-Science programme but not comprehensively explored. Meanwhile developers are turning to "cloud services" such as the storage and compute services offered by Amazon, accessed through simple Web programming interfaces.

This emerging *new* e-Science practice can be characterised as follows:

Everyday researchers doing everyday research. As the entry costs to working digitally have come down, and the benefits are increasingly evident, we are seeing the bulk of researchers working with digital artefacts and digital tools to facilitate their work. This means that there is a greater volume and variety of research data online, but very significantly it also means there are a great many online users – this is the so-called 'long tail'. Furthermore, researchers are exploiting the increasing power of the hardware on their desks (for example, multicore processors), and interacting through everyday devices from laptop to PDA to mobile phone.

A data-centric perspective. As well as the volume of the 'data deluge', the data is increasingly rich, complex and real-time, and often generated locally. Furthermore there is tremendous new value in data, through new digital artefacts and through metadata e.g. context, provenance and scientific workflows. This is not to suggest that computation is unimportant, but rather that many scientists are benefitting from interaction designed around data.

Collaborative and participatory. The process of science has always been collaborative and participatory – it involves publication, peer review, critique and reuse. Now we see the social process of science being revisited in the digital age, using collaborative tools such as blogs and Wikis. Science is about content creation, so these UGC tools come into play. For example, chemistry researchers are using blogs to create electronic lab books with reusable data and provenance records, so that results can be interpreted and trusted, where not only the researchers but also the laboratory instruments are recording data on the blog [Neylon]. OpenWetWare (openwetware.org) is a Wiki with significant uptake in the scientific community and demonstrates the willingness of scientists to make their scientific protocols visible.

Benefitting from the scale of digital science. As the process becomes increasingly digital and increasingly large-scale, we achieve network effects not just through participation and contribution of content. Thus we benefit from collaborative filtering, enabling automatic recommendations based on previous activity and outcomes as well as reviews and tags. This is a new and powerful effect – a new instrument of scale.

Increasingly open. The power of sharing all forms of scientific content is being realised by new mechanisms for publishing, discovery and reuse. Preprint servers, institutional repositories and open access journals, are all mechanisms for making content available for sharing, and Science Commons provides licensing approaches that facilitate this. The Open Archives Initiative provides standards for metadata exchange that promote discovery of content, for example through aggregators. Significantly, the emerging Object Reuse and Exchange standard (http://www.openarchives.org/ore/) provides a mechanism for describing collections of digital artefacts, enabling us not just to work with individual files but rather with compound objects, for example, all the elements that may comprise an experiment.

Better not Perfect. The technologies that researchers are choosing to use are not perfect but scientists choose them because there is an immediate benefit, and often the promise of a longer term benefit too. They are also easy or familiar to use. Sometimes the user requirements evolved through use. It is not therefore possible to deliver perfect tools by following a traditional software engineering method based on requirements capture at the outset – a more agile process is sought.

Empowering researchers. Many of the success stories of e-Science come from researchers who have learned to use ICT and/or have domain ICT experts who are creating the solutions. However, researchers need a sense of ownership of the tools. Moreover, anything that takes autonomy away from researchers is likely to be resisted. Successful e-Science projects have demonstrated the power of giving scientists the tools to assemble a new software "apparatus", rather than building a solution and obliging them to use it.

Using pervasive computing. e-Science is about the intersection of the digital and physical worlds. On the one hand we have sensor networks delivering more data more often from more places, on the other we are interacting with the digital world not just through portals in web browsers but through handheld devices and new forms of display.

These eight characteristics described above correspond to the Web 2.0 design patterns [O'Reilly, 2005] and this should be no surprise. In some sense, the Web 2.0 patterns are about the contemporary relationship between computers and users, so we would expect to see the same with e-Science. However, we need to ask if scientists do use Web 2.0 technologies differently.

The myExperiment project (www.myexperiment.org) has built a social networking environment for scientists in order to test these principles and explore this specific question [DeRoure et al, 2007]. To meet the special needs of scientists it pays due attention to attribution, licensing, ownership and sharing policies. It supports the datatypes of scientists, and in particular it supports collections of information into compound *Research Objects*. The initial digital artefacts supported by myExperiment are scientific workflows, each of which captures a piece of scientific activity, like a protocol [Gil et al, 2007]. Scientific workflows provide an important case study because they are used by a broad, decoupled community of individual scientists who stand to gain by being able to discover workflows, reuse and repurpose them, and by publishing the workflows they create.

myExperiment is particularly interesting because it acknowledges the social aspect of science – rather than making scientific content available as a library or repository, it provides a social infrastructure to encourage sharing. It supports informal exchange and annotation, while fitting in with the more formal scholarly process of the learned journal. Indeed, it may be interesting to compare this with the 17th century 'Invisible College' – the scientists collaborating through informal exchange (by letter) and annotation (of books), which was a precursor of the Royal Society [Zuccala, 2006].

e-Science applications using Web 2.0 technologies can be developed by Web developers, in contrast to the more highly trained specialists needed to work with Grid and the Semantic Web. It is uncontroversial to suggest that the Web can therefore be seen as a kind of 'usability' layer between the Grid infrastructure and the user applications for both users and developers. However, one might go further: Web 2.0 provides a means of coupling together resources in a flexible manner to meet the scientist's needs, so this might be seen as an alternative to Grid computing, i.e. can we use the Web as a distributed application platform for e-Science?

In some cases the answer is clearly yes, but Web 2.0 demands robust underlying services and the techniques of Grid computing are one way of providing these. The shift we anticipate is that increasingly the Web will be used for assembly of functionality over a variety of robust infrastructure resources, and these will include computing clouds, supercomputers and grids as appropriate – so the role of Grid as integrator of distributed resources will give way to achieving that with the Web. This is an exciting prospect, enabling ease of scientific exploration and achieving new – not just faster – scientific outcomes.

Towards a Science of the Web

The pioneering use of Web in e-Science illustrates the evolution of the Web in its context of use, bringing together Semantic Web, Web 2.0 and Web Services. Semantic Web has been successfully adopted in specific areas, such as bioinformatics where the circumstances were ready for an immediate gain from these technologies; in turn the bioinformatics experience provides use cases for the evolution of Web tools and standards. Web 2.0 is used for collaboration, sharing, mashups and the Web is increasingly used as a distributed application platform; e-Science is leading to new Web 2.0 tools such as myExperiment. Web Services have been harnessed by scientific workflow systems, producing new digital artefacts for sharing.

As the Web continues to evolve it will offer ever greater opportunities for e-Research. What are the implications of this for science? Can we learn how to anticipate the effects of such developments? The interaction of these evolving technologies seems difficult to predict; for example, one might now conjecture that some of the promise of the Semantic Web can better be realised now that there is Web 2.0 content in place, such as tags, folksonomies and personal profiles. This was the subject of a 2008 workshop which explored the evolution of the Web and provides examples of how this may be studied [De Roure, 2008]. How do we ensure that future generations of researchers are trained to understand these phenomena and harness their power to produce more sophisticated research methodologies? To answer these sorts of question we believe nothing less than a new discipline is required – Web Science.

Web Science is the emerging interdisciplinary field that views the World Wide Web as an important entity to be studied in its own right [Berners-Lee et al, 2006a, Berners-Lee et al, 2006b]. Physical science is commonly regarded as an analytic discipline that aims to find laws that generate or explain observed phenomena; computer science is predominantly (though not exclusively) synthetic, in that formalisms and algorithms are created in order to support particular desired behaviour. Web science deliberately seeks to merge these two paradigms. The Web needs to be studied and understood as a phenomenon, but it also needs to be engineered for future growth and capabilities. At the micro scale, the Web is an infrastructure of artificial languages and protocols; it is a piece of engineering. However, it is the interaction of human beings creating, linking and consuming information that generates the Web's behaviour as emergent properties at the macro scale.

The Web's macro properties are often surprising and require analytic methods to understand them. Some properties are desirable, and therefore to be engineered in, others are undesirable, and if possible should be engineered out. We also need to keep in mind that the Web's use is part of a wider system of human interaction – the Web has had profound effects on society, with each emerging wave creating both new challenges and new opportunities in making information of different kinds available to wider sectors of the population than ever before.

How do we can design systems to have the eventual effect we envision? Currently, the best we can do is to design and build the micro elements hoping for the best – but how do we know if we've built in the right elements/functionality to ensure large-scale, macroscopic take up? How do we predict what the side effects and emergent properties of the large-scale will be? Further, as the success or failure of a Web technology may involve aspects of social interactions between users understanding the Web requires more than a simple analysis of technological issues, but also an understanding of social dynamics. Given the breadth of the Web, and its inherently multi-user, social nature, its science is necessarily interdisciplinary, involving at least mathematics, computer science, artificial intelligence, sociology, psychology, biology and economics.

It is very early days for Web Science. The Web Science Research Initiative (www.webscience.org) was launched in November 2006 and its methodologies for analysing the Web, its development and evolution are only just emerging [Hendler, 2008]. The Web is different from other previously studied systems in that it is changing at a rate which is of the same order as, or greater than, our ability to observe it. The effect of this on e-Research is potentially profound and it could be argued that all scientists in the future would do well to study elements of Web Science as part of their basic education so that they can understand and contribute to the future development of this remarkable infrastructure and construct.

Conclusions

The Web has changed the way we do research. However, e-Science and e-Research are changing the Web. These activities are bringing a huge volume of data into Web content and making it reusable, and they are establishing tools and methods for collaboration which enhance the social process of science. The techniques we establish for discovery, reuse, review, trust and curation of research data are set to be more broadly applicable to other Web data content.

But the story won't stop here. Just as Web 1.0 resulted from how we used the read-only Web (e-commerce, search engines etc.), and Web 2.0 has resulted from the applications that have been built based on the interactive Web (blogs, wikis, social networks etc.), so Web 3.0 will be the result of applications that are built based on the Semantic Web and a Web of linked data. As yet, we have no way of really knowing what these applications will be until the Semantic Web moves from islands of data with relatively small populations of users to much larger ecologies. We believe that to understand and anticipate what these possibilities are we need a new discipline that studies the Web as a first class object of study.

References

Alani, H.; Hall, W., O'Hara, K., Shadbolt, N., Szomszor, M., Chandler, P. "Building a Pragmatic Semantic Web", IEEE Intelligent Systems, Volume 23, Issue 3, May-June 2008. Pages 61-68. DOI 10.1109/MIS.2008.42

Berners-Lee, T. "Weaving the Web". Texere Publishing. 1999.

T. Berners-Lee, J. Hendler and O. Lassila "The Semantic Web" Scientific American, May 2001, pp 29-37.

T. Berners-Lee, W. Hall, J. Hendler, N. Shadbolt and D. Weitzner "Creating a Science of the Web" Science, 311, 2006a.

T. Berners-Lee, W. Hall, J. Hendler, K. O'Hara, N. Shadbolt and D. Weitzner "A Framework for Web Science" Foundations and Trends in Web Science, 1(1), 2006b.

Sergey Brinn and Lawrence Page "The Anatomy of a Large-Scale Hypertextual Web Search Engine" Computer Networks and ISDN Systems, Vol 30, Issues 1-7, April 1998 pp 107-117, Proceedings of the Seventh International World Wide Web Conference

Bush, V. (1945) "As We May Think". Atlantic Monthly, pp 101 – 108, July 1945 .

De Roure, D., Jennings, N.R. and Shadbolt, N.R. *The Semantic Grid: Past, Present, and Future*, Proceedings of the IEEE, Volume 93, Issue 3, March 2005, Pages 669-681.

David De Roure, Carole Goble and Robert Stevens. Designing the myExperiment Virtual Research Environment for the Social Sharing of Workflows. e-Science 2007 - Third IEEE International Conference on e-Science and Grid Computing, 2007. Bangalore, India, 10-13 December 2007. Pages 603-610.

De Roure, D. and Hall, W. (eds). Proceedings of the First International Workshop on Understanding Web Evolution (WebEvolve2008): A prerequisite for Web Science. Web Science Research Initiative, Southampton, UK. 2008.

Droegemeier, "Transforming the sensing and numerical prediction of high impact local weather through dynamic adaptation". In this volume.

Monica Duke, Michael Day, Rachel Heery, Leslie A. Carr, and Simon J. Coles. Enhancing access to research data: the challenge of crystallography. In JCDL '05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries, pages 46–55, New York, NY, USA, 2005. ACM.

Engelbart, D. (1963) “A conceptual framework for the augmentation of man’s intellect”. In *Vistas of Information Handling*, Vol. 1. London: Spartan Books.

Gil, Y., Deelman, E., Ellisman, M., Fahringer, T., Fox, G., Gannon, D., Goble, C., Livny, M., Moreau, L., and Myers, J. 2007. Examining the Challenges of Scientific Workflows. Computer 40, 12 (Dec. 2007), 24-32. DOI http://dx.doi.org/10.1109/MC.2007.421.

Carole A. Goble, Oscar Corcho, Pinar Alper and David De Roure, *e-Science and the Semantic Web: A Symbiotic Relationship*, Discovery Science, 9th International Conference (DS2006), Barcelona, Spain, October 7-10, 2006. Lecture Notes in Computer Science 4265, Pages 1-12. 9th International Conference, DS 2006, Springer. ISBN 3-540-46491-3

Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T., and Weitzner, D. 2008. Web science: an interdisciplinary approach to understanding the web. Commun. ACM 51, 7 (Jul. 2008), 60-69. DOI http://doi.acm.org/10.1145/1364782.1364798

Tony Hey and Anne E. Trefethen. The UK e-Science core programme and the Grid. Future Gener. Comput. Syst., 18(8):1017–1031, 2002.

Rob McCool, "Rethinking the Semantic Web, Part 1," IEEE Internet Computing ,vol. 9, no. 6, pp. 88, 86-87, November/December, 2005. DOI http://doi.ieeecomputersociety.org/10.1109/MIC.2005.133

Nelson, T. (1981) “Literary Machines” Published by the author

Neylon, C. Blog, Science in the Open. http://blog.openwetware.org/scienceintheopen/

T. O’Reilly. What is Web 2.0 – design patterns and business models for the next generation of software. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

Shadbolt, N.R., Berners-Lee, T. and Hall, W. “The Semantic Web Revisited”. IEEE Intelligent Systems, Volume 21 , Issue 3 (May 2006). Pages 96 – 101. DOI http://dx.doi.org/10.1109/MIS.2006.62

Taylor, K., Gledhill, R., Essex, J. W., Frey, J. G., Harris, S. W., and De Roure, D. 2005. A Semantic Datagrid for Combinatorial Chemistry. In Proceedings of the 6th IEEE/ACM international Workshop on Grid Computing (November 13 - 14, 2005). IEEE Computer Society, Washington, DC, 148-155. DOI http://dx.doi.org/10.1109/GRID.2005.1542736.

Wolstencroft, K., Alper, P., Hull, D., Wroe, C., Lord, P. W., Stevens, R. D., and Goble, C. A. 2007. The myGrid ontology: bioinformatics service discovery. Int. J. Bioinformatics Res. Appl. 3, 3 (Sep. 2007), 303-325. DOI http://dx.doi.org/10.1504/IJBRA.2007.015005

Zuccala, A. 2006. Modeling the invisible college. *J. Am. Soc. Inf. Sci. Technol.* 57, 2 (Jan. 2006), 152-168. DOI http://dx.doi.org/10.1002/asi.v57:2