Web Science

Michalis Vafopoulos[*] and Michalis Kenteris[*]

[*] Cultural Informatics Department, Saphous and Arionos st., 81100 Mitilini, Greece. Email: vaf@aegean.gr

Abstract

Notwithstanding the Web is the most popular piece of software in human society has been unstudied in many aspects. What is the social, political and economical change after Web invention? Could be a useful tool for developing countries? Do we need a new privacy framework? Do we need regulation or liberalization? These and many other questions motivate emergence of a framework for a new trans-disciplinary field; Web science. The major scientific question of Web science could be what technological and other changes need to be incorporated in evolution of Web technologies to best serve human beings. What is Web Science? Is it a new discipline or a new name for an old discipline? Is it a genuine academic discipline at all? What is a Web Science methodology? In this context, privacy is briefly demonstrated as a Web science research case. Finally, the major research challenges and a strategic sketch for the Web Science Research Initiative are presented.

Keywords: Web, Semantic Web, Web science, interdisciplinarity, trans-disciplinarity, privacy.

1. Web technologies

a. Web 1.0: Web of documents

The World Wide Web (known as "WWW" or "Web") is now the primary source of information and communication for many people [Cole, Suman, Schramm, Lunn, Aquino, (2003); Fox (2002)]. The Web began as a networked information project at CERN, where Tim Berners-Lee, now Director of the World Wide Web Consortium (W3C), developed a vision of the project [Berners-Lee (1989)]. According to [Berners-Lee (2004)] "The dream behind the Web is of a common information space in which we communicate by sharing information. Its universality is essential: the fact that a hypertext link can point to anything, be it personal, local or global, be it draft or highly polished. There was a second part of the dream, too, dependent on the Web being so generally used that it became a realistic mirror (or in fact the primary embodiment) of the ways in which we work and play and socialize. That was that once the state of our interactions was on line, we could then use computers to help us analyse it, make sense of what we are doing, where we individually fit in, and how we can better work together.” The Web is a system of interlinked, hypertext documents accessed via the Internet. With a web browser, user views web pages that may contain text, images, videos, and other multimedia and navigates between them using hyperlinks. The Web is an information space in which the items of interest, referred to as resources, are identified by global identifiers called Uniform Resource Identifiers (URI) using the Hypertext Transfer Protocol (HTTP). The Web has a body of software, and a set of protocols and conventions. Web technologies are successful because are based in architecture (e.g. HTTP, URI, HTML) which is simple, networked, based on open standards, extensible, tolerant, universal (regardless hardware platform, software platform, application software, network access, public, group, or personal scope, language and culture operating system and ability), free or cheap, fun and powerful.

b. Web 2.0: the Social Web

The term “Web 2.0” initiated in a conference brainstorming session between O'Reilly and Dougherty, concluded that the Web had outlived the collapse of the Internet bubble of 2001 and became more important than ever, with exciting new applications and websites which allowed broader user participation. Web 2.0 is a term depicting the trend in the use of World Wide Web technology (especially, after 2001) that facilitates naïve user creativity, information sharing, and, most importantly, massive collaboration among users (inter-creativity). These websites and applications have led to the development of Web based services, such as social networking sites, wikis, blogs, and folksonomies (i.e. facebook, myspace, youtube etc). Despite the fact that the term misleadingly suggests a new version of the Web (as it was described in the previous section), it does not refer to an update to any technical specifications, but to changes in the ways software developers and end users use the Web. Newly emerged techniques such as AJAX do not supersede underlying technologies like HTTP and URI, but add an additional layer of abstraction on top of them. Despite the fact that the ideas of Web 2.0 had already been featured in Tim Berners-Lee’s original vision, it could be useful to define 1990’s Web as the Web 1.0 decade and 2000’s as the Web 2.0 decade, since the establishment of Web as the major social and business artifact became clear. User friendliness, mass participation, collective intelligence, network effects from user contributions and market dominance are the innovative characteristics of Web 2.0.

c. Web 3.0: the Semantic Web

The Semantic Web (SW) is an evolving extension of the World Wide Web in which content can be expressed not only in natural language, but also in a format that can be read and used by software agents, thus permitting them to find, share and integrate information more easily. It derives from Berners-Lee's vision of the Web as a universal medium for data, information, and knowledge exchange. According to [Berners-Lee (2004)] "The great need for information about information, to help us categorize, sort, pay for, own information is driving the design of languages for the Web designed for processing by machines, rather than people. The Web of human-readable document is being merged with a web of machine-understandable data. The potential of the mixture of humans and machines working together and communicating through the Web could be immense." In few words, the vision of the SW is as an extension of Web principles from documents to data. Web hides computers and analogously, SW hides documents. At its basis, the SW is composed of an enveloping philosophy, a set of architectural principles and a variety of enabling technologies. Some elements of the SW are expressed as prospective future possibilities that have yet to be put into practical effect (e.g. Rule Interchange Format, RIF) [Ginsberg, Hirtle (2006)]. Other elements of the SW are symbolized in formal specifications. Some of these include Resource Description Framework (RDF) [Manola, Miller (2004)], which unifies a variety of data interchange formats (e.g. RDF/XML, N3, Turtle, N-Triples) using XML as the interchange syntax, and notations such as RDF Schema (RDFS) and the Web Ontology Language (OWL) [D. [McGuinness, van Harmelen (2004)], all of which are purposed to provide a formal description of concepts, terms, and relationships within a specific knowledge domain. Until today, the major effort in SW research and development has been focused on expert system-like applications with complicated semantics, excluding popular social networking applications [Hendler, Golbeck, (2008)].

There are two sound exceptions in this rule: twine and Friend of Friend ontology (FOAF). Twine (twine.com) is a new service that helps users organize, share and discover information about their interests, with networks of like-minded people. Twine is powered by semantic understanding, automatically organizes information, learns about users' interests and makes connections and recommendations. The FOAF project (foaf-project.org) is creating (mostly automatically) a Web of machine-readable pages describing people, the links between them and the things they create and do. Today, there are millions of FOAF profiles which provide connections among various social networks. An interesting extension to FOAF vocabulary is the Semantically-Interlinked Online Communities initiative (SIOC, sioc-project.org). SIOC intends to enable the integration of online community information by providing a SW ontology for representing rich data from the Social Web in RDF. It has recently achieved significant adoption through its usage in a variety of commercial and open-source software applications and published as a W3C Member Submission, which was submitted by 16 organisations.

d. Web 4.0: Metacomputing or WebOS

In all the Web eras, namely Web 1.0, Web 2.0 and Web 3.0, the fundamental hyper-linking property is actually the same. What differentiates last two eras is that are currently exploiting link spaces to different benefit. Web 2.0 earns value from the social dimension of linkage properties and Web 3.0 from URI-based semantic linkages [Hendler, Golbeck, (2008)] (Table 1).

eras	description	basic value source
Web 1.0: 1990's ***documents***	"Surfing" Web: The browser is the platform	hyper-linking of documents
Web 2.0: 2000's ***people***	Social Web: The Web is the platform	social dimension of linkage properties
Web 3.0:2010's ***data***	Semantic Web: The Graph is the platform	URI-based semantic linkages
Web 4.0:2020's ***abilities***	Metacomputing: The network is the platform	+ processing power hyper-linking

Table 1: *The basic value source in all four Web eras.*

Future Internet (FI) technologies and metacomputing paradigms promise that the Web could be used as a distributed universal computer that can take advantage of the universal information space and various heterogeneous devices and sensors. In this context, there are various definitions for metacomputing, but [Foster, Kesselman (1996)] offer an all-embracing definition by determining metacomputing as "a networked virtual supercomputer, constructed dynamically from geographically distributed resources linked by high-speed networks." The major applications of the metacomputing paradigm are considered to be (a) desktop supercomputing (application of high-end graphics capabilities with remote supercomputer and data set), (b) smart instruments (connect users to instruments such as microscope), (c) collaborative environments (multiple virtual environments for interaction of users in different places) and (d) distributed supercomputing (multiple computers to tackle hard problems) [Foster, Kesselman (1996)]. First to foresee the emergence of metacomputing was John McCarthy, at MIT Centennial in 1961. He declared that "If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility... The computer utility could become the basis of a new and important industry". Since then, many scientific terms emerged in order to describe new trends and directions in utility computing [Ross, Westerman (2004)] or on demand computing [Fellenstein (2005)] or computing on demand [Hunter (2003)]. The decisive boost for metacomputing, during the last decade, came from Grid computing [Foster, Kesselman (1999)], Internet [Comer (1997)] and Web technologies which a common communication platform accelerating the physical and virtual connection between machines, digital content and people. In this context, g-work paradigm [Vafopoulos (2005); Vafopoulos, Aggelis, Platis (2005); Vafopoulos, Gravvanis, Platis (2005)] have been introduced g-work as a Web 4.0-based holistic analytical framework advancing Web benefits for local development.

In parallel, Web Operating Systems (or simply WebOS) portray network services for Internet scale distributed computing. The term WebOS was coined from a project at the University of California, Berkeley in 1996 as part of the Network of Workstations project. It was completed in 1998 with the NOW finale. WebOS provide basic operating systems services needed to build applications that are geographically distributed, highly available, incrementally scalable, and dynamically reconfiguring. These applications include mechanisms for resource discovery, a global namespace, remote process execution, resource management, authentication, and security. The most popular WebOS platforms (or Webtops) that enhance some of the above characteristics are DesktopTwo, G.ho.st, YouOS, Browser OS, eyeOS, Stoneware webOS, Jooce and Webxos.

Nevertheless, the concept of WebOS or Web 4.0 could be used in broader sense to include a possible paradigm shift to metacomputing as the primary way of solving everyday problems in science, business and personal level providing new abilities (Table 1).

2. *Creating a science of the Web*

a. *Introduction*

The rapid development of Information and Communication Technologies (ICT), especially Web technologies, are influencing in an increasing rhythm the ways we live and communicate. The Web has been transformational (it's the largest human information artifact). With 1.5 billion users, the Web is now fundamental to our daily activities and of high economic, social and cultural significance. Yet we know relatively little about why the Web is so successful or how it works. We need to understand it better, to anticipate future developments, and to identify opportunities and threats. In order to engineer Web's future and to promote its social benefit we need a new trans-disciplinary research framework that we call Web Science.
What is Web Science? Is it a new discipline or a new name for an old discipline? Is it a genuine academic discipline at all? What is a Web Science methodology? What is the core knowledge set that Web Science practitioners share?

The proposed trans-disciplinarity approach for Web science should go one step further from interdisciplinarity, as it will be based upon a common analytical framework and metaphors. In trans-disciplinary research, the point is not just application of given methodologies, but also implication, a result of creating entirely new possibilities for what disciplines can do. The "trans" in trans-disciplinarity is about recognizing the holistic approach of this process of investigation which transforms mainstream definitions of research.

One of the envelope questions of Web Science could be **what technological and other changes need to be made in order for the Web to work better?** Web Science is inherently trans-disciplinary and experience from other non-disciplinary fields, such as econometrics, suggests success requires a clear functional framework. As a decentralized information system, the Web itself could act as a major catalyst for effective trans-disciplinary research. For the first time, researchers have a well-conceivable and practical common tool to use to help them build the necessary contact points. Existing approaches are insufficient and a new science is needed. Current engineering-based approaches are limited-scope and failed to provide the social insights and policy implications expected. Although substantial research methods and data are available on Web developments and trends, we need to conceive in a much panoramic way what the data is telling us. Web Science, or the 'Science of the Web', offers a means of studying the Web as a decentralised information system with its own intrinsic structure, rules and laws.

Web Science draws on a diverse range of disciplines, from computer science and engineering, the physical and mathematical sciences, to social sciences and policy-making approaches. It should be a true science, with hypotheses that can be clearly articulated, tested and verified. The foundations for Web Science will emerge through interaction between the many different contributing disciplines. The next step is to orchestrate a systematic scientific methodology based on the synergies of technological and social approach. This is what motivates the development of a framework for a new trans-disciplinary field; Web science (Figure 1).

New disciplines emerge where befitting methods and tools could not be found within existing fields. This is essentially the case today with Web Science. To study the Web we need to evolve, redesign and create techniques, methodologies and theories based on fields such as mathematics, economics, sociology and computer science. At the same time, we need to distinguish Web Science from existing disciplines (i.e. network science); the focus is specifically the Web as a techno-social phenomenon on its own and not just the study of all "Web-like things".

b. Research agenda

Web Science is about emergence. The central challenge of this new science is to understand human interaction and communication in the context of the Web. How and why do simple principles and behaviours create complex structures and phenomena, and how can we anticipate these in advance? In particular, Web Science research could provide innovation on networked information and knowledge – i.e. how people relate to and use information in the group context and how consensus (or "social wisdom") emerges from individual behaviours.

The Web we use today is an achievement of computer science. Web Science, building on inputs from many different communities and disciplines, could be expected to result in a different kind of Web – one that is more human-centric and effective in solving everyday problems in micro and macro scale. ICT would look very different as a result. For instance, nobody invented the blogosphere; it just emanated as a result of the mass take-up of a particular Web application. But what would happen if we could design such spaces or applications?

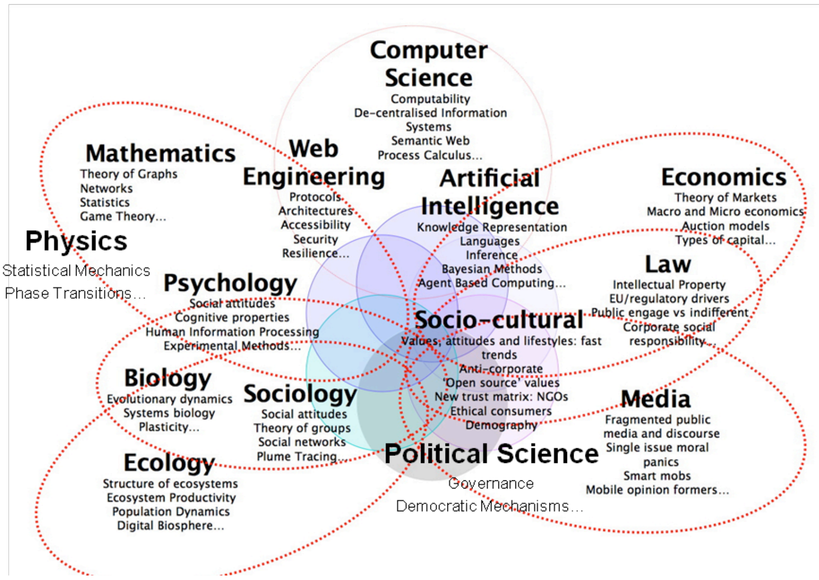

Figure 1*: The trans-disciplinary nature of Web science (webscience.org)*

The research questions are numerous and challenging. We need to understand how the Web changes society and how society changes the Web. What are the features that make the Web work? What is it about the Web's design that has led to success? What are the threats to the Web? What do we want from it? The research challenges closely related to ICT may refer to (a) privacy, security and trust and confidence, (b) human-computer interaction, (c) more effective collaboration, (d) better Web applications and interfaces and (e) creating social value.

Research and development in *Web and Semantic Web engineering* is at the center of Web science. In the course of developments in *Mathematical Logic* many different types of reasoning were originated, but, so far, only deductive linear reasoning and statistical models have been implemented in an automated way. In this context, what alternative methods can the Web facilitate? Which are the logics that are appropriate for the Web, or the Semantic Web? *Probability* is important in conceiving the future

Web. The Web is not static, deterministic and asynchronous, but posses multifold stochastic elements. *Economics* would contribute in areas such as game theory, value creation, information asymmetries and dynamic policy frameworks. Web and new computing paradigms (i.e. Grid) raise the question of how lots of relatively autonomous individuals can work together to produce mutually beneficial results (either results beneficial to each individual, or to society as a whole). Are there upper limits to the utility of the freedom that decentralisation has produced? As the number of users increases, will the chances that the choices that one makes impinge on the range of choices available to others increase, or is that an illegitimate extrapolation from the real world with fixed spatial parameters? In a decentralised and growing Web, where there are no "owners" as such, can we be sure that decisions that make sense for an individual do not damage the interests of users as a whole? [Berners-Lee, (2006)] The Web is a socially embedded technology. The major *social dimension* of the Web could be reflected in the following question. Is the social status quo of the physical world (partially) reflected in decentralized systems and virtual worlds? What are the transformation mechanisms of "physical" personalities and societies in "virtual" ones? Specifically, what do people and communities want from the Web, and what online behavior is required for the Web to work? Furthermore, the subjective beliefs formed by the readers of a certain webpage could be described under a single word: trust. A sociological question about trust could be: Do we trust the machines and automated processes that are put under way when we work or play on the Web? The basic issue in *governing* the Web is how things should be regulated to ensure the steady and fruitful development of the Web? In addition, a group of serious political issues can be summarized in the question if the Web is a liberal construct.

c. Privacy as a Web science research case

The Web is today one of the key components of critical infrastructures and is transforming our lives in personal and social level. It generates opportunities for new forms of socialization and personal development. But at the same time with the Web's vast growth has been a growth in computer-related crimes. The range of criminal activity that the Web supports is enormous and multifold, including consumer, business and government threats.

Privacy violation is a basic part of this criminal activity. Privacy concerns include crimes target personally identifiable information for theft and misuse and more general non-crime-related issues about the amount of data we create in our electronic personalities and how that data can be collected, aggregated, analyzed, disseminated, and used. In this context, we demonstrate a potential research workflow of the need for new privacy framework in Web science. The theoretical model used has been introduced by [Hendler, Shadbolt, Hall, Berners-Lee, Weitzner (2008)] and it constitutes cyclical flows of knowledge creation and social interaction in five basic steps.

1. In the first step, human values impose several issues/problems for scientific investigation. In the present case, the issue is considered to be the need for a new privacy framework in digital age.
2. Creativity in the academic and/or business sphere generates innovative ideas in order to provide solutions in a specific problem. The information accountability [Weitzner, Abelson, Berners-Lee, Feigenbaum, Hendler, Sussman, (2008)] proposed solution framework, could be part of the vision not to create a "secure world." On the contrary, comprehending that the physical world will never be completely safe, we need a cyber world reasonably safe under certain operational contexts.
3. The third step is twofold, including the technical and social design path. In the technical part of privacy, it is proposed to design systems that are oriented toward information accountability, transparency and appropriate use, rather than information security and access restriction. In these systems, users should be able to make (a minimized number of) trust decisions and they should be ensured with effective information and an intuitive interface that encourages the right choices. A Web engineering choice could be dependency reasoning. In this context, trust in cyberspace should be a process, interconnected to the rest of our life, not just a set of policy rules. These systems, in order to be feasible, functional and sustainable need social appreciation and minimal consensus which can be achieved through legal rules and changes in user's mentality.
4. The fourth step refers to the micro implications of the proposed implementations and is supported through the analysis between techno-social interplay. In our case, an effective, acceptable and sustainable privacy framework of technologies and institutions which includes accountable usage and usage tracking could drive productivity in personal and social level.
5. The final step examines the consequences of micro foundations in the macro scale. Degrees of complexity in aggregating micro foundations in macro policy implications and suggestions, is considered to be a key point. Productivity can foster economic development in local and global scale.

The cyclical workflow ends first round by analyzing feedback from the macro implementation and proposals to redefine solutions are expected.

d. Web Science Research Initiative

The Web Science Research Initiative [www.webscience.org] was launched in November 2006 as a long-term partnership between the University of Southampton and the Massachusetts Institute of Technology. It brings together academics, scientists, sociologists, entrepreneurs and decision-makers from around the world to create the first multidisciplinary research body to examine the World Wide Web and offer practical solutions to help guide its future use and design. WSRI is lead by co-directors Tim Berners-Lee (MIT/Southampton), Wendy Hall (Southampton), Nigel Shadbolt (Southampton) and Daniel Weitzner (MIT).

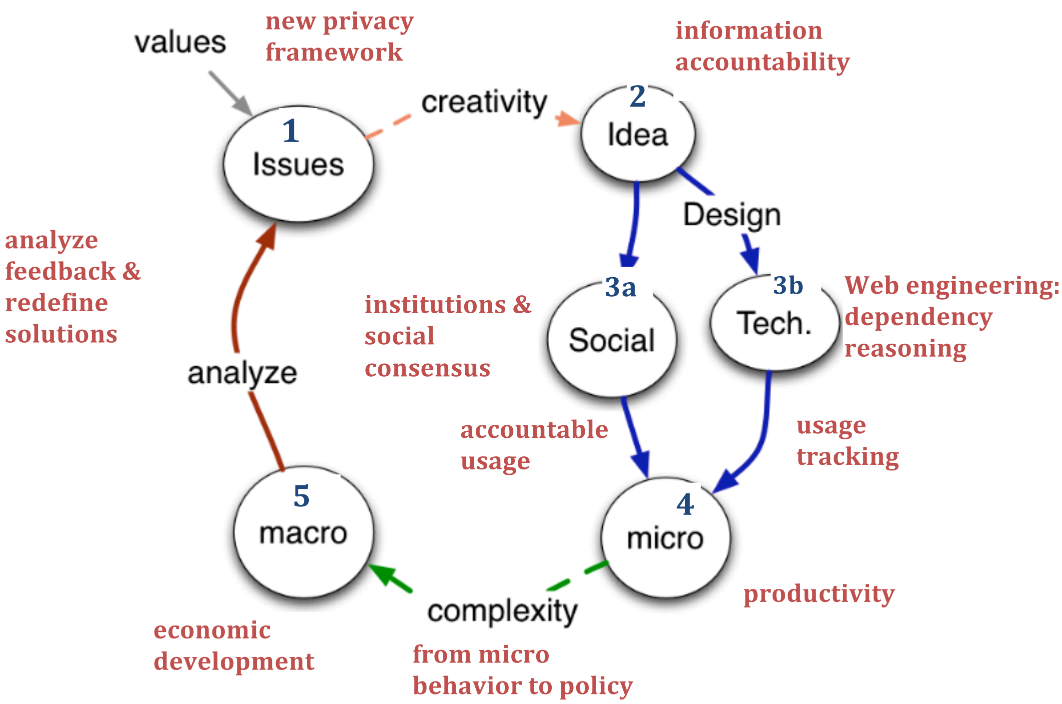

Figure 2: *The need for a new privacy framework: cyclical flows of knowledge creation and social interaction in the context of Web science*

i. Strategy

The Initiative has three main goals: research, thought leadership and education (Figure 3). It undertakes research into the scientific, technical and social factors that

drive the growth of the Web, so as to better understand past and present phenomena (such as the rapid growth of the Blogosphere), and anticipate new ones (for example the data Web). As a global forum of experts, it also aims to raise awareness, lead thinking, disseminate information and provide corporations, governments and regulators with the capability to anticipate future developments. In addition, it brings together educationalists to develop degree curricula and build capacity in this new field. These three fundamental pillars aim to create a trans-disciplinary approach to understand how the Web will evolve, anticipate issues that will arise and proactively address these issues to ensure that the Web remains as a positive enabler for our society.
The next steps of Web science could be (a) find evangelists and successful trans-disciplinary paradigms, (b) building the Web science community, (c) creating a common agenda and (d) develop functional metaphors.

A. Research	**B. Thought leadership**	**C. Education**
1. Evolution of the web i. Structure ii. Protocols iii. Social effects iv. Fragmentation v. Rates of change *2. Applications* i. Education ii. Science iii. Health iv. eGovernment	*3. Global forum of experts* i. Academia, Government, Industry ii. Raise awareness, lead thinking, disseminate *4. Workshops* i. Issue generation ii. Consultancy iii. Conferences iv. Input Sessions	*5. Web Science curriculum development at undergraduate and graduate levels* *6. Creation of Web Science capabilities and talent* *7. Role of web in innovation* *8. Web as driver for development*

integrated, trans-disciplinary perspective
capacity for new field of Web Science
solutions to anticipated issues
coordination on a global scale

Figure 3: *The strategic pillars and goals of the Web Science Research Initiative (presented by W. Hall and N. Shadbolt at Web science workshop 2008 in Athens)*

3. Conclusion

Web is the most popular piece of software in human society. What is the social, political and economical change after Web invention? Could be a useful tool for developing countries? How can we ensure global accessibility to computing resources? Do we need regulation or liberalization? Or both? These and many other questions motivate emergence of a framework for a new trans-disciplinary field; Web science. The major scientific question of Web science could be what technological and other changes need to be incorporated in evolution of Web technologies to best serve human beings. It was argued that the foundations for Web Science would emerge through interaction between the many different contributing disciplines. No single community will have all the answers; each will approach the issue from its own perspective and have something interesting to say. The key is in bringing communities together so as to create synergies. This would require those concerned to develop a common language and common foundations, in particular in a way that bridges the 'hard' and 'soft' sciences.

Up to now the Web has largely been about creating value for commercial operators. But there are many ways in which it could be used to create social value as well. Web science will lead to a new kind of Web. This new discipline aims both to understand the Web and to focus its development so as to better meet human needs. The benefits will be felt in all areas of science, business and government, and by individual users in their daily lives.

6. References

Cole, J. I., Suman, M., Schramm, P., Lunn, R., & Aquino, J. S. (2003) The ucla internet report surveying the digital future year three. UCLA Center for Communication Policy. http://www.ccp.ucla.edu/pdf/ucla-internet-report-year-three.pdf.

Fox, S., (2002). Search engines The Pew Internet & American Life Project. http://www.pewinternet.org/reports/toc.asp.

Berners-Lee, T. (2004). The World Wide Web: A very short personal history. http://www.w3.org/People/Berners-Lee/ShortHistory.html

Berners-Lee, T. (1989). Information management: a proposal. CERN, March 1989. http://www.w3.org/History/1989/proposal.html

A. Ginsberg and D. Hirtle, eds., (2006) RIF Use Cases and Requirements, 2006. http://www.w3.org/TR/rif-ucr.

F. Manola and E. Miller, eds., RDF Primer, 2004. http://www.w3.org/TR/rdf-primer.

D. L. McGuinness and F. van Harmelen, eds., (2004) OWL Web Ontology Language Overview, 2004. http://www.w3.org/TR/owl-features

Ian Foster and Carl Kesselman (1996). Globus: A Metacomputing Infrastructure Toolkit. Int. J. Supercomputer Applications, 1996.

Hendler, J. and Golbeck, J. (2008).Metcalfe's Law, Web 2.0, and the semantic Web Journal of Web Semantics 6, 1

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), pp. 60-69.

Berners-Lee, T. (2006). *A framework for web science.* Foundations and trends in web science, vol. 1, issue 1. Boston: Now.

J. W. Ross and G. Westerman (2004). "Preparing for utility computing: The role of IT architecture and relationship management", *IBM Systems Journal*, 43, 1, pp. 5-19, 2004.

C. Fellenstein (2005). On demand computing: technologies and strategies, On demand series, Upper Saddle River, NJ, Pearson Education, Inc., 2005.

P. Hunter (2003). "High-Performance Computing On-Demand", *The Scientist*, 17(23):38, 2003.

D. Comer (1997). The internet book: everything you need to know about computer networking and how the internet works, Upper Saddle River, N.J., Prentice-Hall, 1997.

Foster and C. Kesselman (1999). The Grid: blueprint for a new computing infrastructure, San Francisco, Morgan Kaufmann Publishers, 1999.

M.N Vafopoulos (2005). "A roadmap to the GRID e-workspace", Proceedings of the Advances in Web Intelligence: Third International Atlantic Web Intelligence Conference, AWIC 2005, Lodz, Poland, June 6-9, Lecture Notes in Computer Science 3528 Springer, ISBN 3-540-26219-9, 504 – 509, 2005.

M.N. Vafopoulos, V. Aggelis and A.N Platis (2005). "HyperClustering: from digital divide to the GRID e-workspace", Proceedings of the VI Data Mining: Data Mining, Text Mining and their Business Applications, WIT Transactions on Information and Communication Technologies, vol. 34, ISSN 1743-3517, 311-321, 2005.

M.N. Vafopoulos, G.A. Gravvanis and A.N. Platis (2005), "The personal Grid e-workspace", in: M.P. Bekakos, G.A. Gravvanis and H.R. Arabnia, eds., Grid Technologies: Emerging from Distributed Architectures to Virtual Organizations, WIT Press, 2005.

Weitzner, D., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., and Sussman, G. (2008) Information accountability. Commun. ACM 51, 6 (June 2008).