Could the Web be a Temporary Glitch?

Leslie Carr, Cathy Pope, Susan Halford
University of Southampton
Southampton
United Kingdom
lac@ecs.soton.ac.uk

ABSTRACT

In this paper we begin with the assertion that Web Science is the study of the technologies and policies that support the co-construction of a linked online environment by a networked society, and we end by questioning whether the Web that we currently enjoy is a permanent and fundamental phenomenon, or merely a fashionable popular enthusiasm for a novel kind of information sharing.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information Services

General Terms

Management, Documentation, Economics, Security, Human Factors, Legal Aspects.

Keywords

Web science, open access, information sharing

1. INTRODUCTION

The Web is a co-construct of society and technology mutually transforming each other(Halford, Pope, & Carr, 2010). The technology (an Internet-based protocol for information exchange) facilitates individuals sharing information, which in turn causes society to experiment with mass public and open forms of knowledge transfer, which in turn forces the technologists to improve their protocols, lawyers to refine their legislation and industry to rewrite its business models.

As society has developed and grown through history, information has become more valuable and increasingly more available. Whether for the ends of agriculture, trade, warfare, technology, manufacturing or government, more people have been involved in gathering and creating information and more sophisticated ways have been developed to extract new value from increasing amounts of information. The development of society is inextricably related to the technology of information provision, consumption and dissemination (*e.g.* writing, reading, printing, education), and so it is unsurprising that a new information technology could alter society's operation.

Web Science Conf. 2010, April 26-27, 2010, Raleigh, NC, USA.

Society is not a monolithic entity; there exist a variety of social institutions with different roles and objectives and hence different relationships with knowledge, and the technology that is used to acquire and disseminate that knowledge.

Academy	Create and transmit knowledge
Commerce	Trade goods
Press	Report news
Media	Broadcast content
Military	Defend society
Government	Control society & share resources

These different social institutions have different objectives, concerns and modus operandi, and hence impose different requirements on their IT infrastructures that lead to incompatible expectations of a Web of information. Arguments over the relative importance of openness, security, transparency and privacy are hardly surprising therefore.

The Web was developed at CERN (the European Centre for Nuclear Research) to support large-scale scientific experiments, undertaken by large teams of experimenters requiring access to a diverse range of documentation using a heterogeneous mix of computing infrastructure (Berners-Lee, 2000). This new environment migrated beyond the lab into the wider research community, and subsequently the public sphere.

Of the societal institutions listed above then, it is only the academy's values and practices that informed the initial creation of the Web. Granted, the Web is built on Internet protocols and norms, which are grounded in the *military* requirement for decentralized organizational structures that can withstand damage inflicted on the hardware infrastructure of the underlying network (Abbate, 1999). However, the military influence of the networking layers did not inform the design of the Web's information superstructure.

What are the principal values of academia? Although it may seem a horribly complex question in current political environment, it is uncontroversial to say that the academy operates on fundamentally different rules to the commercial

Sponsor	System	Scope	Date	Important Properties
Press	Reuters	Professional, centralised	1850	Fast access to news & stock information (originally carrier pigeon and subsequently telegraph)
Private Institution	Mundaneum	Public, centralised	1910	Extended a library with indexing technology (the library card) and remote query via telephone
Military	Memex	Scholarly, individual, centralised	1945	Helping scientists and technologists to cross discipline boundaries.
Media*	Xanadu	Public, decentralised	1960s	Organising personal ideas and the universal literature; focused on DRM and author reimbursement
Media	Teletext	Public, national, centralised	1976	Broadcast information services, linked, not participative
Government	Minitel	Public, national, centralised	1982	Interactive commercial services and information
Academy (CS & HEP)	FTP / Archie / Anarchie	Public, decentralised	1985	Downloading software and PostScript documents to hard drives for printing on LaserWriters.
Commerce	Hypercard, HyperTIES	Private, centralised	1988	Personal applications, sometimes tied to multimedia resources on CDROMs / video disks
Academy (HEP)	WWW	Public, global, decentralised	1990	Document exchange. Universal naming, linking, interoperability. Open and participative but no writing or indexing.
Academy (CS)	Microcosm	Private, centralised	1990	Sophisticated linking and openness for personal information stores.
Academy (CS)	HyperG	Public, centralised	1990	Alternative to Web (and subsequently an extension to Web) with support for writing, indexing and consistency management.
Commerce	AOL, CompuServ	Public, centralised	1990	Dialup access to (closed) email, forums, chat rooms and information resources.

Figure 1: Historical Information Sharing Technologies

world. Supported by government funding and philanthropic donations, academic institutions exist to further knowledge and to benefit society by helping to apply that knowledge. This entails that their research findings are made public without payment and that researchers collaborate together in common endeavor with a wider responsibility to their disciplines.

The values of academia that have been imprinted on its operation during its development are now affecting the whole of society. The Web's open information system that shares content with no inbuilt mechanism for payments or rights management impacts the way that information, content and media are traded, developed and sustained.

However, it is not only academia and its values that have developed and shaped information environments. Other institutions had influenced previous attempts at developing widespread information systems.

2. OTHER SPONSORS, OTHER WEBS

Historically, there have been a number of attempts at developing widespread information systems over the previous century (summarized in Figure 1). Some of these have worked within the constraints of contemporary technology, others have tried to push those boundaries with less success and are relegated to the status of 'never implemented' (*e.g.* Vannevar Bush's Memex) or never com,pleted (Xanadu). Even so, we consider them here because they were serious proposals for achievable information systems that were highly influential.

Developments in communications preceded storage and processing technology, with the electric telegraph, characterized by Standage as 'the Victorian Internet' (Standage, 1998), achieving instantaneous trans-global communication. But the capability that this internet provided – instantaneous communication between two places hundreds or thousands of miles apart – was so incomprehensible to the contemporary audience that it took years for the possibilities to be realized in a world in which the speed of communication was limited to the speed of a galloping horse and messages between business partners might take weeks or months. Reuters was one of the first companies to base their business on this new (private, point-to-point) communications ability, transmitting stock market quotations between London and Paris.

Microfilm technology was seen as the information storage solution for the proposed documentation systems the World Brain (Wells, 1937) and the Memex (Bush, 1945). Neither system addressed any communications capability – both assumed that copies (of the hardware and information resources) would be manufactured and shipped to every individual or organization needing to access the world's scientific knowledge. The 'World Brain' was inspired by the failure of academia to undertake 'the thought and knowledge organization of the world' (Wells, 1937). The design of the Memex (notably introducing a form of hypertext link) was a response to the shortcomings of the war effort; the need of the military to engage a cross-disciplinary perspective on problem solving that saw scientists 'burying their old professional competition in the demand of a common cause' (Bush, 1945).

By contrast, the aim of Paul Otlet's earlier Mundaneum (van den Heuvel, 2008) was not to store or transmit knowledge *per se*, but to provide an efficient, publicly accessible index to a comprehensive library of texts based on 15 million index cards. Queries for this 'search engine' were originally received by post, but later expected to be received by phone and wireless. His work was inspired by the end of the Great War, and the hope of building a new society, based on centralised access to shared knowledge and understanding.

Building on the capabilities of the new electronic computers of the 1960s, Xanadu (Nelson, 1987) was one of the early pioneers of adapting a 'calculating environment' to the needs of documentation and the capture and manipulation of ideas. Credited with coining the word 'hypertext' and articulating its capabilities and a mechanism for achieving those capabilities, Nelson considered himself initially a film-maker and consequently his motivation was not for "shared global knowledge" but for the ability to organize and use creative material in such a way that professional authors always received recompense for any quotation or reuse of their material in the emergent global interconnected literature.

In the 1970s a system for incorporating pages of text within public television broadcasts (Teletext) was adopted by a number of national broadcasting agencies as a centralised mechanism for information dissemination (Graziplene, 2000). Although paid-for advertising was accepted by some broadcasters, it was the later French Minitel system that provided genuine interaction and market-led services in the 1980s by adopting the telephone network, the modem and the microcomputer to produce a widely used national information service (Cats-Baril & Jelassi, 1994).

Microcomputers and personal computers delivered personal (*i.e.* isolated) information services throughout the 1980s and 1990s – using floppy disks, CDROMs and even Videodisks as the dissemination mechanism for databases. Apple's commercial Hypercard popularized a graphical front-end to hypertext resources (Goodman, 1987), and research systems like HyperTIES (Shneiderman, 1989) and Microcosm (Hall, Davis, & Hutchings, 1996) exploring closer integration with researchers' local desktop working environments.

Throughout the 1980s, the Internet became adopted more and more widely, although due to its government-funded status in the US and UK it was deployed for non-commercial purposes only throughout academia and the research industry. In that limited environment it began to have an enormous impact, with email changing the dynamics of scholarly communication, and FTP archives providing a convenient new mechanism for the dissemination of reports, documents and even academic papers in many technical disciplines such as computer science, economics and high energy physics.

Commercial dial-up services such as CompuServe (Bowen & Peyton, 1990) grew up during the 1980s, independently of the Internet, and targeting the general public rather than the research community to provide a subscription-based 'walled garden' of centralized information resources.

The Web (Berners-Lee, 2000), developed at CERN around 1991 as a solution to the problems of exchanging information in highly collaborative international research projects, took hold in the wider academic internet and from there established itself in other areas of public activity (media, commerce and government).

However, the Web didn't succeed independently of these other efforts. The Web needed a 'back end' to handle authentication of users and the upload of created content; this was mainly provided by the Internet FTP service. Existing commercial dialup services such as CompuServe and AOL became Web portals, so feeding the growth of the Web, but slowly perishing in the process as the dominant operational model of the Web became "free advert-supported content". (Compuserve had already integrated its email with the internet in 1989 and in 1995 started to provide access to the WWW from its bespoke 'information manager' client.)

The fact that the Web was developed in the context of a highly collaborative academic research environment set the initial assumptions and design parameters in place that led to the dominance of free content: sharing and openness were the keys. By comparison, Xanadu, an earlier system with similar ambitions to provide a global interconnected web of literature, was designed by a film-maker and writer whose principal concern was to provide fundamental support for Digital Rights Management and micropayments to track and reward the use of materials(Nelson, 1987).

3. CO-CONSTRUCTION IN ACADEMIA

Pre-web, scientific and scholarly publication had traditionally involved the donation of a research manuscript and transfer of its copyright to commercial publishers, who then make economic profits by selling subscriptions to their publications. By charging for access to research, publishers guarantee a sustainable business model for scholarly communications, at the cost of limiting the readership and impact of research – the so-called Faustian bargain (Harnad, 1995).

However, BOAI, the Budapest Open Access initiative(Chan, et al., 2002) declared that 'an old tradition and a new technology have converged to make possible an unprecedented public good'. The old tradition is the academic practice of publishing the results of academic research without expecting payment, the new technology is the Web[1] and the new public good is completely free and unrestricted world-wide electronic distribution of the research literature for "scientists, scholars, teachers, students, and other curious minds" (normally referred to as *Open Access*).

In conjunction with the technical standards of the Open Archiving Initiative(Lagoze & Van de Sompel, 1999), which enabled repositories[2] to exchange metadata about their contents with third party services, a new class of web application emerged: the Open Access Institutional Repository whose role was to provide access to research articles for individuals and institutions across the world that lacked a subscription to the particular journals in which those articles were published.

Based on open source repository packages such as EPrints(Carr, 1999), the repository also represents a new kind of scholarly activity – the curation and management of an institution's intellectual output on behalf of its own scholars for the benefit of the whole scholarly community.

[1] The declaration itself refers to the internet, but it is clear from the context that it is the Web that is intended

[2] previously dumb FTP archives, but by then becoming sophisticated Web applications

We see the following chain of events

- The research ethos of academia determines the design of the Web
- The Web allows researchers (and anyone) to create web sites to disseminate information at apparently zero cost
    - and share their publications online through Web sites and repositories
- Commercial publishing companies realize the threat of the Web to their business models
    - and use copyright to restrict the author's ability to share their own articles

Apparently ignoring publishers, or removing them from the scholarly publication chain, the BOAI caused (and still causes) debate in the academic community about the economics of scholarly communication and the application of Intellectual Property and Copyright to its activities.

Other knowledge co,mmunities have adapted to (and even been formed by) the Web. A prime example of the circular influence of technological and social development is Wikipedia. A form of open, collaborative website (the Wiki) has been adapted to the role of encyclopedia, but using 'open sharing' as the guiding principle rather than 'closed expertise'. New knowledge is edited and managed on the web through processes that are discussed and managed through the Web. Wikipedia only exists because of the Wikipedia community; the Wikipedia community only exists because of the Web. Both are a linked resource with an emerging set of values and standards embodied in an evolving set of processes.

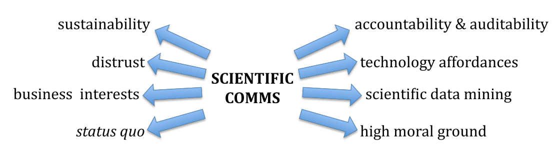

Figure 2: Pressures on Academic Community for and against increasing openness

By contrast, academia can seem much more conservative – new knowledge is created privately (by individual researchers and groups) and published through subscription journals. Centuries-old processes and values pre-exist in academic organisations (quality and independence) and are re-interpreted for the prevailing technology. Although a decade of the Web's influence has spawned new agendas for the academic community – Open Access (2002), Open Data (2004), Open Educational Resources (2005), Creative Commons and Science Commons (2006) – these agendas are still not adopted by the whole of the scholarly and scientific community.

It is not just inertia against changing practice, or unfamiliarity with technology that mitigates against increasing openness of scholarly communications. There is a genuine issue of service sustainability to be addressed: web publishing may be less costly than the manufacturing and distribution of paper products, but there is still some cost associated with it that must be born by someone. Publishing is more than simple dissemination (including aspects of quality control, editorial responsibility, marketing and preservation of the academic record), but even dissemination has its costs (*e.g.* server and software maintenance, information management, security patching).

Beyond the issues of cost recovery (or even business profitability), there are problems raised by the community's relationship to the research information it generates, and the community's social relationships. When it comes to sharing research data through the Web, researchers express reservations (RIN, 2009) about trusting each other (open data allows others to steal an advantage) and about trusting each others' data (which may not have been compiled with appropriately high standards).

Contrariwise, the issue of trust is at the forefront of the recent ClimateGate[3] scandal in the UK. By keeping climate research data private, and refusing to respond to Freedom of Information requests made by critical climate scientists, in order to protect their conclusions from antagonistic scrutiny, scientists at the University of East Anglia were criticized by the House of Commons Science and Technology Committee, who demanded 'greater transparency':

> *We consider that climate scientists should take steps to make available all the data that support their work, including raw data... and full methodological workings, including the computer codes*[4]

In a world in which governments are investing huge amounts of public money in policies that are supported by scientific research, the pressure is on for scientific information to be open and auditable, rather than private and hidden away.

But equally, in a world where the financial trading of privately owned information underpinned by copyright (through subscription products such as journals or magazine) is the only model for ensuring continuity of information production, there is genuine nervousness that individuals adopting a radically new modus operandi for information transfer will damage the overall information flow in society.

The aim of this paper is not to fully rehearse the arguments for and against Open Access (of which the author is a supporter), but to note that the Web's affordance for free and open information exchange is not uncritically embraced by a community, or even that there are no disadvantages to its adoption.

[3] http://en.wikipedia.org/wiki/Climatic_Research_Unit_email_controversy

[4] http://news.bbc.co.uk/1/hi/sci/tech/8595483.stm

4. CONCLUDING REMARKS

The Web is not an independent thing, it is a socio-technical phenomenon, brought together and held together by people, historical contexts, antecedents, (sub-)cultural norms and expectations. In this paper we have illustrated this complex interplay of actors by examining the Web's impact on the academic community. Both dynamic and contingent, the Web has the potential to move (grow and develop) but it does so subject to chance and conditional factors (temporality). It is of its moment, and responds to current concerns (such as openness, free speech, censorship, pedophilia, commerce, intellectual property, DRM and cybercrime).

Web Science has to understand this Web (the Web that we have produced today) sufficiently well that we can engage in a systematic, scientifically informed futurology concerning the range of potential Webs that can be constructed with broadly the same technology but different social drivers. How can the Web (our Web) adapt as other social institutions impose their requirements on it? How can we maximize societal benefits if (or when) the expectations of security, policing or energy availability take future priority over free exchange of information and the unrestricted transfer of knowledge?

In particular, are the public and open aspects of the Web a fundamental change in our society's information processes? Are open source, open access, open science & creative commons efficient alternatives to fee-based knowledge transfer that should be maintained and extended in the face of commercial or political resistance? Openness is a property of the Web architecture and a contributory factor in the success of its adoption, but it is not an inevitable property of the user experience in the coming decade.

5. ACKNOWLEDGMENTS

This work was funded in part by the JISC Institutional Data Management Blueprint project (IDMB) and the EPSRC Digital Repositories E-Science Network.

6. Works Cited

Abbate, J. (1999). *Inventing the Internet.* MIT Press.

Andrews, K., Kappa, F., & H., M. (1995). Hyper-G and harmony: towards the next generation of networked information technology. *CHI '95.*

Berners-Lee, T. (2000). *Weaving the Web.* New York, NY, USA: Harper Business.

Bowen, C., & Peyton, D. (1990). *Compuserve information manager : the complete sourcebook.* Bantam Books.

Bush, V. (1945, July). As We May Think. *Atlantic Monthly* .

Carr, L. (1999). Retrieved from EPrints Software: http://www.eprints.org/

Cats-Baril, W. L., & Jelassi, T. (1994). The French Videotex System Minitel: A Successful Implementation of a National Information Technology Infrastructure. *MIS Quarterly , 18* (1).

Chan, L., Cuplinskas, D., Eisen, M., Friend, F., Genoa, Y., Guédon, J.-C. H., et al. (2002, Feb 14). *Budapest Open Access Initiative.* Retrieved from Open Society Institute: http://www.soros.org/openaccess/read.shtml

Endrijonas, J. (1996). *America Online discovered.* Random House.

Goodman, D. (1987). *The Complete Hypercard Handbook.* New York, NY, USA: Random House Inc.

Graziplene, L. R. (2000). *Teletext: its promise and demise.* Lehigh University Press.

Halford, S., Pope, C., & Carr, L. (2010). A Manifesto for Web Science. *WebSci10: Extending the Frontiers of Society On-Line.*

Hall, W., Davis, H. C., & Hutchings, G. A. (1996). *Rethinking Hyopermedia: the Microcosm Approach.* Kluwer Academic Publishers.

Harnad, S. (1995). Electronic Scholarly Publication: Quo Vadis? *Serials Review , 21* (1), 70-72.

Lagoze, C., & Van de Sompel, H. (1999). Retrieved from Open Archives Initiative: http://www.openarchives.org/

Nelson, T. (1987). *Literary Machines* (87.1 ed.).

RIN. (2009). *Patterns of information use and exchange: case studies of researchers in the life sciences.* Research Information Network.

RIN. (2009). *Patterns of information use and exchange: case studies of researchers in the life sciences.* Retrieved from Research Information Network: http://www.rin.ac.uk/case-studies

Shneiderman, B. (1989). Reflections on Authoring, Editing, and Managing Hypertext. In E. Barrett, *The society of text: hypertext, hypermedia, and the social construction of information* (pp. 115-131). Cambridge, MA, USA: MIT Press.

Standage, T. (1998). *The Victorian Internet.* New York, New York, USA: Walker and COmpany.

van den Heuvel, C. (2008). Architectures of Global Knowledge: The Mundaneum and the World Wide Web. *Destination Library , 15*, 48-53.

Wells, H. (1937). World Brain: The Idea of a Permanent World Encyclopaedia. In *Encyclopédie Française.*