Technology & Standards Watch

What is Web 2.0?
Ideas, technologies and implications for education

by
Paul Anderson

This report was peer reviewed by:

Mark Hepworth
Senior Lecturer, Department of Information Science
Loughborough University

Brian Kelly
Policy and Advice Team Leader and Web Focus
UKOLN

Randy Metcalfe
Manager, OSSWatch
University of Oxford

Lawrie Phipps
Programme Manager, Users and Innovation
JISC

Web 2 Executive summary

Within 15 years the Web has grown from a group work tool for scientists at CERN into a global information space with more than a billion users. Currently, it is both returning to its roots as a read/write tool and also entering a new, more social and participatory phase. These trends have led to a feeling that the Web is entering a 'second phase'—a new, 'improved' Web version 2.0. But how justified is this perception?

This TechWatch report was commissioned to investigate the substance behind the hyperbole surrounding 'Web 2.0' and to report on the implications this may have for the UK Higher and Further Education sector, with a special focus on collection and preservation activities within libraries. The report argues that by separating out the discussion of Web technologies (ongoing Web development overseen by the W3C), from the more recent applications and services (social software), and attempts to understand the manifestations and adoption of these services (the 'big ideas'), decision makers will find it easier to understand and act on the strategic implications of 'Web 2.0'. Indeed, analysing the composition and interplay of these strands provides a useful framework for understanding its significance.

The report establishes that Web 2.0 is more than a set of 'cool' and new technologies and services, important though some of these are. It has, at its heart, a set of at least six powerful ideas that are changing the way some people interact. Secondly, it is also important to acknowledge that these ideas are not necessarily the preserve of 'Web 2.0', but are, in fact, direct or indirect reflections of the power of the network: the strange effects and topologies at the micro and macro level that a billion Internet users produce. This might well be why Sir Tim Berners-Lee, the creator of the World Wide Web, maintains that Web 2.0 is really just an extension of the original ideals of the Web that does not warrant a special moniker. However, business concerns are increasingly shaping the way in which we are being led to think and potentially act on the Web and this has implications for the control of public and private data. Indeed, Tim O'Reilly's original attempt to articulate the key ideas behind Web 2.0 was focused on a desire to be able to benchmark and therefore identify a set of new, innovative companies that were potentially ripe for investment. The UK HE sector should debate whether this is a long-term issue and maybe delineating Web from Web 2.0 will help us to do that.

As with other aspects of university life the library has not escaped considerable discussion about the potential change afforded by the introduction of Web 2.0 and social media. One of the key objectives of the report is to examine some of the work in this area and to tease out some of the key elements of ongoing discussions. For example, the report argues that there needs to be a distinction between concerns around quality of service and 'user-centred change' and the services and applications that are being driven by Web 2.0 ideas. This is particularly important for library collection and preservation activities and some of the key questions for libraries are: is the content produced by Web 2.0 services sufficiently or fundamentally different to that of previous Web content and, in particular, do its characteristics make it harder to collect and preserve? Are there areas where further work is needed by researchers and library specialists? The report examines these questions in the light of the six big ideas as well as the key Web services and applications, in order to review the potential impact of Web 2.0 on library services and preservation activities.

CONTENTS

Introduction

At the end of 2006, Time magazine's Person of the Year was 'You'. On the cover of the magazine, underneath the title of the award, was a picture of a PC with a mirror in place of the screen, reflecting not only the face of the reader, but also the general feeling that 2006 was the year of the Web - a new, improved, 'second version', 'user generated' Web. But how accurate is our perception of so-called 'Web 2.0'? Is there real substance behind the hyperbole? Is it a publishing revolution or is it a social revolution? Is it actually a revolution at all? And what will it mean for education, a sector that is already feeling the effects of the demands of Internet-related change?

In this TechWatch report I argue for the distinction between Web technologies (ongoing Web development overseen by the W3C), the more recent applications and services that are emerging as a result of this ongoing technological development (social software), and attempts to understand the manifestations and adoption of these newer applications and services. I start with a brief discussion of the historical context, with Sir Tim Berners-Lee and his vision for a single, global, collaborative information space and contrast this story of the technology with the ideas of Tim O'Reilly, who has attempted to understand the ways in which knowledge about the technologies, and the adoption of the technologies, can be used to make predictions about technology markets.

Media coverage of Web 2.0 concentrates on the common applications/services such as blogs, video sharing, social networking and podcasting—a more socially connected Web in which people can contribute as much as they can consume. In chapter two I provide a brief introduction to some of these services, many of them built on the technologies and open standards that have been around since the earliest days of the Web, and show how they have been refined, and in some cases concatenated, to provide a technological foundation for delivering services to the user through the browser window (based on the key idea of the Web, rather than the desktop, as the technology platform). But is this Web 2.0? Indeed, it can be argued that these applications and services are really just early manifestations of ongoing Web technology development. If we look at Web 2.0 as it was originally articulated we can see that it is, in fact, an umbrella term that attempts to express explicitly the framework of ideas that underpin attempts to understand the manifestations of these newer Web services within the context of the technologies that have produced them.

In section three I articulate six 'big' ideas, based on concepts originally outlined by Tim O'Reilly, which can help us to explain and understand why Web 2.0 has had such a huge impact. In short, these are ideas about building something more than a global information space; something with much more of a social angle to it. Collaboration, contribution and community are the order of the day and there is a sense in which some think that a new 'social fabric' is being constructed before our eyes. These ideas though, need technology in order to be realised into the functioning Web-based services and applications that we are using.

Education and educational institutions will have their own special issues with regard to Web 2.0 services and technologies and in section five I look at some of these issues. By special request, particular attention has been given to libraries and preservation and the issues that present themselves for those tasked with preserving some of the material produced by these services and applications. Finally, I look to the future. What are the technologies that will affect the next phase of the Web's development: what one might call, rather reluctantly, Web 3.0?

1. 'Web 2.0' or 'Web 1.0'?: a tale of two Tims

Web 2.0 is a slippery character to pin down. Is it a revolution in the way we use the Web? Is it another technology 'bubble'? It rather depends on who you ask. A Web technologist will give quite a different answer to a marketing student or an economics professor.

The short answer, for many people, is to make a reference to a group of technologies which have become deeply associated with the term: blogs, wikis, podcasts, RSS feeds etc., which facilitate a more socially connected Web where everyone is able to add to and edit the information space. The longer answer is rather more complicated and pulls in economics, technology and new ideas about the connected society. To some, though, it is simply a time to invest in technology again—a time of renewed exuberance after the dot-com bust.

For the inventor of the Web, Sir Tim Berners-Lee, there is a tremendous sense of *déjà vu* about all this. When asked in an interview for a podcast, published on IBM's website, whether Web 2.0 was different to what might be called Web 1.0 because the former is all about connecting people, he replied:

"Totally not. Web 1.0 was all about connecting people. It was an interactive space, and I think Web 2.0 is of course a piece of jargon, nobody even knows what it means. If Web 2.0 for you is blogs and wikis, then that is people to people. But that was what the Web was supposed to be all along. And in fact, you know, this 'Web 2.0', it means using the standards which have been produced by all these people working on Web 1.0."[1]

Laningham (ed.), developerWorks Interviews, 22nd August, 2006.

To understand Sir Tim's attitude one needs look back at the history of the development of the Web, which is explored in his book *Weaving the Web* (1999). His original vision was very much of a collaborative workspace where everything was linked to everything in a 'single, global information space' (p. 5), and, crucially for this discussion, the assumption was that 'everyone would be able to edit in this space' (IBM podcast, 12:20 minutes). The first development was Enquire, a rudimentary project management tool, developed while Berners-Lee was working at CERN, which allowed pages of notes to be linked together and edited. A series of further technological and software developments led to the creation of the World Wide Web and a browser or Web client that could view *and edit* pages of marked-up information (HTML). However, during a series of ports to other machines from the original development computer, the ability to edit through the Web client was not included in order to speed up the process of adoption within CERN (Berners-Lee, 1999). This attitude to the 'edit' function continued through subsequent Web browser developments such as ViolaWWW and Mosaic (which became the Netscape browser). Crucially, this left people thinking of the Web as a medium in which a relatively small number of people published and most browsed, but it is probably more accurate to picture it as a fork in the road of the technology's development, one which has meant that the original pathway has only recently been rejoined.

The term 'Web 2.0' was officially coined in 2004 by Dale Dougherty, a vice-president of O'Reilly Media Inc. (the company famous for its technology-related conferences and high quality books) during a team discussion on a potential future conference about the Web (O'Reilly, 2005a). The team wanted to capture the feeling that despite the dot-com boom and subsequent bust, the Web was 'more important than ever, with exciting new applications and sites popping up with surprising regularity' (O'Reilly, 2005a, p. 1). It was also noted, at the same meeting, that companies that had survived the dot-com firestorms of the late 90s now appeared to be stronger and have a number of things in common. Thus it is important to note that the term was not coined in an attempt to capture the essence of an identified group of technologies, but an attempt to capture something far more amorphous.

[1] A transcript of the podcast is available at: http://www-128.ibm.com/developerworks/podcast/dwi/cm-int082206.txt [last accessed 17/01/07].

The second Tim in the story, Tim O'Reilly himself, the founder of the company, then followed up this discussion with a now famous paper, *What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software,* outlining in detail what the company thought they meant by the term. It is important to note that this paper was an attempt to make explicit certain features that could be used to identify a particular set of innovative companies, including business characteristics, such as the fact that they have control over unique, hard-to-recreate data sources (something that could become increasingly significant for H&FE), or that they have lightweight business models. The paper did, however, identify certain features that have come to be associated with 'social software' technologies, such as participation, user as contributor, harnessing the power of the crowd, rich user experiences etc., but it should be noted that these do not constitute a *de facto* Web (r)evolution. As Tim Berners-Lee has pointed out, the ability to implement this technology is all based on so-called 'Web 1.0' standards, as we shall see in section four, and that, in fact, it's just taken longer for it to be implemented than was initially anticipated. From this perspective, 'Web 2.0' should not therefore be held up in opposition to 'Web 1.0', but should be seen as a consequence of a more fully implemented Web.

This distinction is key to understanding where the boundaries are between 'the Web', as a set of technologies, and 'Web 2.0'—the attempt to conceptualise the significance of a set of outcomes that are enabled by those Web technologies. Understanding this distinction helps us to think more clearly about the issues that are thrown up by both the technologies and the results of the technologies, and this helps us to better understand why something might be classed as 'Web 2.0' or not. In order to be able to discuss and address the Web 2.0 issues that face higher education we need to have these conceptual tools in order to identify why something might be significant and whether or not we should act on it.

For example, Tim O'Reilly, in his original article, identifies what he considers to be features of successful 'Web 1.0' companies and the 'most interesting' of the new applications. He does this in order to develop a set of concepts by which to benchmark whether or not a company is Web 1.0 or Web 2.0. This is important to him because he is concerned that 'the Web 2.0 meme has become so widespread that companies are now pasting it on as a marketing buzzword, with no real understanding of just what it means' (O'Reilly, 2005a, p.1). In order to express some of the concepts which were behind the original O'Reilly discussions of Web 2.0 he lists and describes seven principles: The Web as platform, Harnessing collective intelligence, Data is the next 'Intel inside', End of the software release cycle, Lightweight programming models, Software above the level of single device, and Rich user experiences. In this report I have adapted some of O'Reilly's seven principles, partly to avoid ambiguity (for example, I use 'harnessing the 'power of the crowd'', rather than 'collective intelligence' as I believe this more accurately describes the articulation of the concept in its original form), and partly to provide the conceptual tools that people involved in HE practice and decision making have expressed a need for.

2. Key Web 2.0 services/applications

There are a number of Web-based services and applications that demonstrate the foundations of the Web 2.0 concept, and they are already being used to a certain extent in education. These are not really technologies as such, but services (or user processes) built using the building blocks of the technologies and open standards that underpin the Internet and the Web. These include blogs, wikis, multimedia sharing services, content syndication, podcasting and content tagging services. Many of these applications of Web technology are relatively mature, having been in use for a number of years, although new features and capabilities are being added on a regular basis. It is worth noting that many of these newer technologies are concatenations, i.e. they make use of existing services. In the first part of this section we introduce and review these well-known and commonly used services with a view to providing a common grounding for later discussion.

NB * indicates an open source or other, similar, community or public-spirited project.

2.1 Blogs

The term web-log, or *blog*, was coined by Jorn Barger in 1997 and refers to a simple webpage consisting of brief paragraphs of opinion, information, personal diary entries, or links, called *posts*, arranged chronologically with the most recent first, in the style of an online journal (Doctorow *et al.*, 2002). Most blogs also allow visitors to add a *comment* below a blog entry.

This posting and commenting process contributes to the nature of blogging (as an exchange of views) in what Yale University law professor, Yochai Benkler, calls a 'weighted conversation' between a primary author and a group of secondary comment contributors, who communicate to an unlimited number of readers. It also contributes to blogging's sense of immediacy, since 'blogs enable individuals to write to their Web pages in journalism time – that is hourly, daily, weekly – whereas the Web page culture that preceded it tended to be slower moving: less an equivalent of reportage than of the essay' (Benkler, 2006, p. 217).

Well-known or education-based blogs:
http://radar.oreilly.com/
http://www.techcrunch.com/
http://www.instapundit.com/
http://blogs.warwick.ac.uk/ *
http://jiscdigitisation.typepad.com/jisc_digitisation_program/ *

Software:
http://wordpress.org/ *
http://www.sixapart.com/typepad/
http://www.blogger.com/start
http://radio.userland.com/
http://www.bblog.com/

Blog search services:
http://technorati.com/
http://www.gnosh.org/
http://blogsearch.google.com/
http://www.weblogs.com/about.html

Each post is usually 'tagged' with a keyword or two, allowing the subject of the post to be categorised within the system so that when the post becomes old it can be filed into a standard, theme-based menu system[2]. Clicking on a post's description, or tag (which is displayed below the post), will take you to a list of other posts by the same author on the blogging software's system that use the same tag.

Linking is also an important aspect of blogging as it deepens the conversational nature of the blogosphere (see below) and its sense of immediacy. It also helps to facilitate retrieval and referencing of information on different blogs but some of these are not without inherent problems:

- The *permalink* is a permanent URI which is generated by the blogging system and is applied to a particular post. If the item is moved within the database, e.g. for archiving, the permalink stays the same. Crucially, if the post is renamed, or if the content is changed in any way, the

[2] Blog content is regularly filed so that only the latest content is available from the homepage. This means that returning to a blog's homepage after several weeks or months to find a particular piece of content is potentially a hit and miss affair. The development of the permalink was an attempt to counter this, but has its own inherent problems.

permalink will still remain unchanged: i.e. there is no version control, and using a permalink does not guarantee the content of a post.

- *Trackback* (or *pingback*) allows a blogger (A) to notify another blogger (B) that they have referenced or commented on one of blogger B's posts. When blog B receives notification from blog A that a trackback has been created, blog B's system automatically creates a record of the permalink of the referring post. Trackback only works when it is enabled on both the referring and the referred blogs. Some bloggers deliberately disable trackback as it can be a route in for spammers.

- The *blogroll* is a list of links to other blogs that a particular blogger likes or finds useful. It is similar to a blog 'bookmark' or 'favourites' list.

Blog software also facilitates *syndication,* in which information about the blog entries, for example, the headline, is made available to other software via RSS and, increasingly, Atom. This content is then aggregated into feeds, and a variety of blog aggregators and specialist blog reading tools can make use of these feeds (see Table 1 for some key examples).

The large number of people engaged in blogging has given rise to its own term – *blogosphere* – to express the sense of a whole 'world' of bloggers operating in their own environment. As technology has become more sophisticated, bloggers have begun to incorporate multimedia into their blogs and there are now photo-blogs, video blogs (vlogs), and, increasingly, bloggers can upload material directly from their mobile phones (mob-blogging). For more on the reasons why people blog, the style and manner of their blogging and the subject areas that are covered, see Nardi *et al*., 2004.

2.2 Wikis

A *wiki*[3] is a webpage or set of webpages that can be easily edited by anyone who is allowed access (Ebersbach *et al*., 2006). Wikipedia's popular success has meant that the concept of the wiki, as a collaborative tool that facilitates the production of a group work, is widely understood. Wiki pages have an edit button displayed on the screen and the user can click on this to access an easy-to-use online editing tool to change or even delete the contents of the page in question. Simple, hypertext-style linking between pages is used to create a navigable set of pages.

Unlike blogs, wikis generally have a *history* function, which allows previous versions to be examined, and a *rollback* function, which restores previous versions. Proponents of the power of wikis cite the ease of use (even playfulness) of the tools, their extreme flexibility and open access as some of the many reasons why they are useful for group working (Ebersbach *et al*., 2006; Lamb, 2004).

Examples of wikis:
http://wiki.oss-watch.ac.uk/ *
http://wiki.cetis.ac.uk/CETIS_Wiki *
http://en.wikipedia.org/wiki/Main_Page *
http://www.ch.ic.ac.uk/wiki/index.php/Main_Page
http://www.wikihow.com

Software:
http://meta.wikimedia.org/wiki/MediaWiki *
http://www.socialtext.com/products/overview
http://www.twiki.org/
http://uniwakka.sourceforge.net/HomePage

Online notes on using wikis in education:
http://www.wikiineducation.com/display/ikiw/Home *

There are undeniably problems for systems that allow such a level of openness, and Wikipedia itself has suffered from problems of malicious editing and vandalism (Stvilia *et al*., 2005). However, there are also those who argue that acts of vandalism and mistakes are rectified quite quickly by the self-

[3] Ebersbach *et al*. traces this from the Hawaiian word, wikiwiki, meaning 'quick' or 'hurry' from Ward Cunningham's concept of the wikiwikiWeb, in 1995.

moderation processes at work. Alternatively, restricting access to registered users only, is often used for professional, work group wikis (Cych, 2006).

2.3 Tagging and social bookmarking

A tag is a keyword that is added to a digital object (e.g. a website, picture or video clip) to describe it, but not as part of a formal classification system. One of the first large-scale applications of tagging was seen with the introduction of Joshua Schacter's del.icio.us website, which launched the 'social bookmarking' phenomenon.

Social bookmarking systems share a number of common features (Millen *et al*., 2005): They allow users to create lists of 'bookmarks' or 'favourites', to store these centrally on a remote service (rather than within the client browser) and to share them with other users of the system (the 'social' aspect). These bookmarks can also be tagged with keywords, and an important difference from the 'folder'-based categorisation used in traditional, browser-based bookmark lists is that a bookmark can belong in more than one category. Using tags, a photo of a tree could be categorised with both 'tree' and 'larch', for example.

Examples of tagging services:
http://www.connotea.org/
http://www.citeulike.org/*
http://www.librarything.com/
http://del.icio.us/
http://www.sitebar.org
http://www.furl.net/index.jsp
http://www.stumbleupon.com/
http://www.blinklist.com/
http://www.digg.com/
http://www.rawsugar.com
http://del.icio.us/elearningfocus/web2.0 *

The concept of tagging has been widened far beyond website bookmarking, and services like Flickr (photos), YouTube (video) and Odeo (podcasts) allow a variety of digital artefacts to be socially tagged. For example, the BBC's Shared Tags[4] project is an experimental service that allows members of the public to tag BBC News online items. A particularly important example within the context of higher education is Richard Cameron's CiteULike[5], a free service to help academics to store, organise and share the academic papers they are reading. When you see a paper on the Web that interests you, you click a button and add it to your personal library. CiteULike automatically extracts the citation details, so you don't have to type them in. This tool was used during the research for this report.

The idea of tagging has been expanded to include what are called *tag clouds*: groups of tags *(tag sets)* from a number of different users of a tagging service, which collates information about the frequency with which particular tags are used. This frequency information is often displayed graphically as a 'cloud' in which tags with higher frequency of use are displayed in larger text.

Large organisations are beginning to explore the potential of these new tools and their concepts for knowledge management across the enterprise. For example, IBM is investigating social bookmarking through their intranet-based DogEar tool (Millen *et al*., 2005). In education, JISC's e-Learning Focus service has set up a del.icio.us account at: http://del.icio.us/elearningfocus [last accessed 07/02/07].

Folksonomy versus collabulary

One outcome from the practice of tagging has been the rise of the 'folksonomy'. Unfortunately, the term has not been used consistently and there is confusion about its application. More will be said about this in the section on network effects, but for now it is sufficient to note that there is a distinction between a folksonomy (a collection of tags created by an individual for their own personal use) and a collabulary (a collective vocabulary).

[4] http://backstage.bbc.co.uk/prototypes/archives/2005/05/bbc_shared_tags.html [last accessed 16/01/07].
[5] http://www.citeulike.org/ [last accessed 16/01/07].

2.4 Multimedia sharing

One of the biggest growth areas has been amongst services that facilitate the storage and sharing of multimedia content. Well known examples include YouTube (video) Flickr (photographs) and Odeo (podcasts). These popular services take the idea of the 'writeable' Web (where users are not just consumers but contribute actively to the production of Web content) and enable it on a massive scale. Literally millions of people now participate in the sharing and exchange of these forms of media by producing their own podcasts, videos and photos. This development has only been made possible through the widespread adoption of high quality, but relatively low cost digital media technology such as hand-held video cameras.

Well known photo sharing services:
http://www.flickr.com/
http://www.ourpictures.com/
http://www.snapfish.com/
http://www.fotki.com/

Well known video sharing services:
http://www.youtube.com/
http://www.getdemocracy.com/broadcast/ *
http://eyespot.com/
http://ourmedia.org/ *
http://vsocial.com
http://www.videojug.com/

Well known podcasting sites:
http://www.apple.com/itunes/store/podcasts.html
http://btpodshow.com/
http://www.audblog.com/
http://odeo.com/
http://www.ourmedia.org/ *
http://connect.educause.edu/ *
http://juicereceiver.sourceforge.net/index.php
http://www.impala.ac.uk/ *
http://www.law.dept.shef.ac.uk/podcasts/ *

2.5 Audio blogging and podcasting

Podcasts are audio recordings, usually in MP3 format, of talks, interviews and lectures, which can be played either on a desktop computer or on a wide range of handheld MP3 devices. Originally called audio blogs they have their roots in efforts to add audio streams to early blogs (Felix and Stolarz, 2006). Once standards had settled down and Apple introduced the commercially successful iPod MP3 player and its associated iTunes software, the process started to become known as podcasting[6]. This term is not without some controversy since it implies that only the Apple iPod will play these files, whereas, in actual fact, any MP3 player or PC with the requisite software can be used. A more recent development is the introduction of video podcasts (sometimes shortened to vidcast or vodcast): the online delivery of video-on-demand clips that can be played on a PC, or again on a suitable handheld player (the more recent versions of the Apple iPod for example, provide for video playing).

A podcast is made by creating an MP3 format audio file (using a voice recorder or similar device), uploading the file to a host server, and then making the world aware of its existence through the use of RSS (see next section). This process (known as *enclosure*) adds a URL link to the audio file, as well as directions to the audio file's location on the host server, into the RSS file (Patterson, 2006).

Podcast listeners subscribe to the RSS feeds and receive information about new podcasts as they become available. Distribution is therefore relatively simple. The harder part, as those who listen to a lot of podcasts know, is to produce a good quality audio file. Podcasting is becoming increasingly used in education (Brittain *et al*., 2006; Ractham and Zhang, 2006) and recently there have been moves to establish a UK HE podcasting community[7].

2.6 RSS and syndication

RSS is a family of formats which allow users to find out about updates to the content of RSS-enabled websites, blogs or podcasts without actually having to go and visit the site. Instead, information from the website (typically, a new story's title and synopsis, along with the originating website's name) is

[6] Coined by Ben Hammersley in a Guardian article on 12th February 2004: http://technology.guardian.co.uk/online/story/0,3605,1145689,00.html [last accessed 14/02/07].
[7] See: http://www.podcasting.blog-city.com/tags/?/ukhepodnet [last accessed 10/02/07].

collected within a feed (which uses the RSS format) and 'piped' to the user in a process known as syndication.

In order to be able to use a feed a prospective user must install a software tool known as an *aggregator* or *feed reader*, onto their computer desktop. Once this has been done, the user must decide which RSS feeds they want to receive and then *subscribe* to them. The client software will then periodically check for updates to the RSS feed and keep the user informed of any changes.

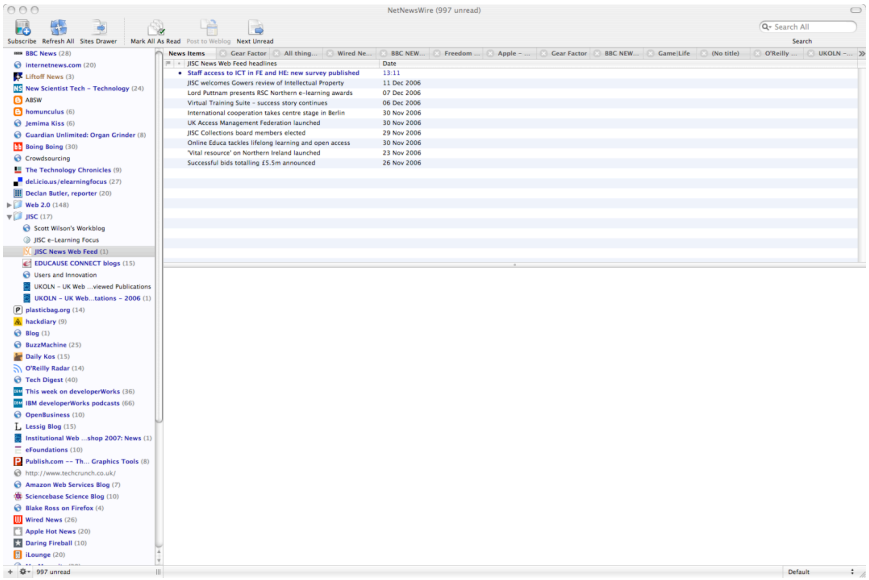

Illustration 1: Example of an RSS feed aggregation tool (NetNewsWire).

Technically, RSS is an XML-based data format for websites to exchange files that contain publishing information and summaries of the site's contents. Indeed, in its earliest incarnation, RSS was understood to stand for Rich Site Summary (Doctorow, 2002). For a variety of historical reasons there are a number of RSS formats (RSS 0.91, RSS 0.92, RSS 1.0, RSS 2.0) and there are some issues of incompatibility[8]. It is worth noting that RSS 2.0 is not simply a later version of RSS 1.0, but is a different format. As it has become more widely used for blog content syndication, in later versions RSS became known as Really Simple Syndication[9]. A lot of blogging tools now create and publish these RSS feeds automatically and webpages and blogs frequently display small RSS icons and links to allow a quick process of registering to get a feed from the site (see above, right).

In 2003 a new syndication system was proposed and developed under the name *Atom* in order to clear up some of the inconsistencies between RSS versions and the problems with the way they interoperate. This consists of two standards: the Atom Syndication Format, an XML language used for Web feeds, and the Atom Publishing Protocol (APP), a HTTP-based protocol for creating and updating Web resources. There is considerable discussion between proponents of RSS and Atom as to which is the best way forward for syndication. The two most important differences between the two are, firstly, that the development of Atom is taking place through a formal and open standards process within the IETF[10], and, secondly, that with Atom the actual content of the feed item's encoding (known as the payload container) is more clearly defined. Atom can also support the enclosure of

[8] See: http://blogs.law.harvard.edu/tech/rssVersionHistory for a history of the versions [last accessed 14/02/07].
[9] See RSS Advisory Board service: http://www.rssboard.org/ [last accessed 14/02/07].
[10] The Internet Engineering Task Force.

more than one podcast file at a time (see podcasting section) and so multiple file formats of the same podcast can be syndicated at the same time[11].

2.7 Newer Web 2.0 services and applications

As we have seen, there are a number of technology services that are often posited as representing the Web 2.0 concept in some way. In recent months, however, there has been an explosion of new ideas, applications and start-up companies working on ways to extend existing services. Some of these are likely to become more important than others, and some are certainly more likely to be more relevant to education than others. There is such a deluge of new services that it is often difficult to keep track of what's 'out there' or to make sense of what each provides. I suggest there are two ways of helping with this process. Firstly, to make sense of what the service is trying to do in the context of the overall Web 2.0 'big ideas' presented in section three. Secondly, as new services become available they can be categorised roughly in terms of what they attempt to do, e.g. aggregate user data, construct a social network etc.

In Table 1 I make a first attempt at such a categorisation process based on a small range of some of the newer services. Such a table is only the beginning of the process and can only be snapshot as this is a fluid market with new tools and start-up companies being announced on almost a daily basis (see, for example, TechCrunch's regular updates[12] on start-ups and new ideas; or eConsultant's Web 2.0 directory which recently listed over 1,200 services in fifty categories ranging from blogging to Wifi)[13].

[11] More technical detail of the Atom standard can be found at: http://www.ietf.org/html.charters/atompub-charter.html and http://www-128.ibm.com/developerworks/xml/library/x-atom10.html [last accessed 14/02/07].
[12] TechCrunch is a blog dedicated to profiling and reviewing new Internet products and companies: www.techcrunch.com
[13] http://www.econsultant.com/web2/

Table 1: Newer Web 2.0 services

Categorisation (based on what they attempt to do)	*Explanation and indicative links to the big ideas of Web 2.0 (see section 3 for more detail)*	*Examples of service*
Social Networking	Professional and social networking sites that facilitate meeting people, finding like minds, sharing content—uses ideas from harnessing the power of the crowd, network effect and individual production/user generated content.	Professional networking: http://www.siphs.com/aboutus.jsp https://www.linkedin.com/ http://www.zoominfo.com/ Social networking: www.myspace.com www.facebook.com http://fo.rtuito.us/ http://www.spock.com/ (test beta only) http://www.flock.com/ http://www.bebo.com/
Aggregation services	Gather information from diverse sources across the Web and publish in one place. Includes news and RSS feed aggregators and tools that create a single webpage with all your feeds and email in one place—uses ideas from individual production/user generated content.	http://www.techmeme.com/ http://www.google.co.uk/nwshp?hl=en http://www.blogbridge.com/ http://www.suprglu.com/ http://www.netvibes.com/
	Collect and aggregate user data, user 'attention' (what you look at) and intentions—uses ideas from the architecture of participation, data on epic scale and power of the crowd.	http://www.attentiontrust.org/ http://www.digg.com/
Data 'mash-ups'	Web services that pull together data from different sources to create a new service (i.e. aggregation and recombination). Uses, for example, ideas from data on epic scale and openness of data.	http://www.housingmaps.com/ http://darwin.zoology.gla.ac.uk/~rpage/ispecies/ http://www.rrove.com/set/item/59/top-11-us-universities http://www.blears.net/weather/ (world weather from BBC RSS feed)
Tracking and filtering content	Services that keep track of, filter, analyse and allow search of the growing amounts of Web 2.0 content from blogs, multimedia sharing services etc. Uses ideas from e.g. data on epic scale.	http://technorati.com/about/ http://www.digg.com/ http://www.blogpulse.com http://cloudalicio.us/about/
Collaborating	Collaborative reference works (like Wikipedia) that are built using wiki-like software tools. Uses ideas from harnessing the power of the crowd.	http://www.squidoo.com/ http://wikia.com/wiki/Wikia
	Collaborative, Web-based project and work group productivity tools. Uses architecture of participation.	http://vyew.com/always-on/collaboration/ http://www.systemone.at/en/technology/overview# http://www.37signals.com/
Replicate office-style software in the browser	Web-based desktop application/document tools. Replicate desktop applications. Based on technological developments.	http://www.google.com/google-d-s/tour1.html http://www.stikkit.com/ http://www.backpackit.com/tour
Source ideas or work from the crowd	Seek ideas, solutions to problems or get tasks completed by out-sourcing to users of the Web. Uses the idea of power of the crowd.	http://www.mturk.com/mturk/welcome http://www.innocentive.com/

3. The big ideas behind Web 2.0

As outlined in section one, there is considerable speculation as to what Web 2.0 might be, and it is inevitable that some of this would become confused as various people vie for attention in the ongoing conversation. What I have tried to do in this section is to uncover what I believe are the core ideas and to show, where possible, points at which various strands of related thought start to be developed. I also try to raise some questions about how closely these strands are related to some kind of evidence base. By looking at the history of, for example, network theory, it is possible to see how assumptions made about the rate at which networks grow could have contributed to the last technology boom and bust. This is important, not only for avoiding a similar situation in the future, but also, for getting a more realistic understanding of the role that Web 2.0 might play within education.

In this section I put forward six 'big' ideas, based on concepts originally outlined by Tim O'Reilly, that can help us to explain and understand why Web 2.0 has had such a huge impact. In short these are ideas about building something more than a global information space; something with much more of a social angle to it. Collaboration, contribution and community are the order of the day and there is a sense in which some think that a new 'social fabric' is being constructed before our eyes. However, it is also important to acknowledge that these ideas are not necessarily the preserve of 'Web 2.0', but are, in fact, direct or indirect reflections of the power of the network: the strange effects and topologies at the micro and macro level that a billion Internet users produce.

	Key Idea
1	Individual production and User Generated Content
2	Harness the power of the crowd
3	Data on an epic scale
4	Architecture of Participation
5	Network Effects
6	Openness

3.1 Individual production and User Generated Content

*'I have always imagined the information space as something to which everyone has immediate and intuitive access, and not just to browse, but to **create**.'*

Tim Berners-Lee, 1999, p. 169

'We don't hate the media, we become the media'

Jello Biafra (Eric Boucher), 2001[14]

In the 1980s the punk rock adage of "I can do that" led to thousands of young people forming local bands and writing their own fanzines. Today's generation are pressing 'record' on their video cameras and hitting their mouse keys. With a few clicks of the mouse a user can upload a video or photo from their digital camera and into their own media space, tag it with suitable keywords and make the content available to their friends or the world in general. In parallel, individuals are setting up and writing blogs and working together to create information through the use of wikis. What these tools have done is to lower the barrier to entry, following in the same footsteps as the 1980s self-publishing revolution sparked by the

[14] From the spoken recording *Become the media* (Alternative Tentacles, 2001) available online at: http://www.alternativetentacles.com/product.php?product=380 [last accessed 12/01/07].

introduction of the office laser printer and desktop publishing software pioneered by Apple (Hertzfeld, 2005). There has been an out-pouring of production on the Web.

Much of recent media attention concerning the rise of the Web 2.0 phenomenon has focused on what's been given the rather ugly moniker of *user generated content* (UGC). Alternatives to this phrase include content self-publishing, personal publishing (Downes, 2004) and 'self expression'.

Media interest in this is derived, in part, because the media itself is undergoing a period of profound change as the true implications of the Web and in particular the new capability of the viewers, or as the journalist Dan Gillmor (2004) describes them, the former audience, to contribute materials for programmes, newspapers and websites. The widespread adoption of cheap, fairly high quality digital cameras, videos, mobile and smartphones, have all contributed to a rise in what's sometimes called 'citizen journalism' or 'witness contributions', in which newspapers and TV programmes make use of viewer's clips of news events. Many media organisations are undertaking major reviews of how they generate content and investing in facilities to allow the public to have more of a role in newsgathering. For example, The Sun newspaper now provides a single mobile phone number for members of the public to submit copy and photos, and in South Korea the OhmyNews service has an army of 40,000 citizen journalists edited by 50 professionals (Anderson, 2006). Meanwhile, the BBC is working on a Creative Archive which will allow users to view and make use of old, archived TV material, possibly 'mashing-up' their own versions of TV content. Many commentators think we are entering a new era in which news is more of a 'conversation' and this kind of change in people's perception of who has the authority to 'say' and 'know' is surely set to be a challenge within education.

So why do people engage in peer production like this? Chris Anderson (2006) says: 'the motives to create are not the same in the head as they are in the tail' (see section 3.5.4). People are driven by monetary motives at the head, but the coin of the realm at the lower end of the tail *is reputation*' (p. 73). We are living in more of an exposure culture, where 'getting noticed is everything' (Tim Wu, Professor of Law, in Anderson, 2006, p. 74).

To some commentators the increasing propensity for individuals to engage in the creation and manipulation of information and digital artefacts is a major positive benefit. There are, of course those who worry about where this might take us. The Chief Scientist at Xerox, John Seely Brown worries about the loss of the structure and authority of an edited newspaper as an institution in which a process of selection and reflection takes place (Brown and Duguid, 2000). The RSS feed is organised temporally, but what is the more important news? A designed newspaper has a headline, an 'above the fold' story, and the editors have selected the news based on lots of factors. There are also those who are sceptical over the true scale of actual participation in all this. Over 10 million of the 13 million blogs in Blogger, a major blog provider, are inactive according to Charles Mann (2006) who thinks that: 'The huge mass of dead blogs is one reason to maintain a healthy scepticism about the vast growth of the blogosphere' (p. 12).

3.2 Harnessing the power of the crowd

The term 'harnessing collective intelligence' as used by Tim O'Reilly has several problems associated with it: firstly, what kind of 'intelligence' are we referring to? If we equate 'information' to 'intelligence' then many of his examples stand up to scrutiny. However, if your understanding of 'intelligence' more naturally focuses on the idea of having or showing some kind of intellectual ability, then the phrase becomes more problematic. O'Reilly acknowledges this inherently by bringing in the concept of 'the wisdom of crowds' (WoC), but this, in turn, brings its own set of problems (see below).

Related to this is the problem of what we mean by 'collective intelligence'. Again, the WoC ideas are drafted in by O'Reilly to try to help with this, but there is a critical gap between the explication of 'wisdom of crowds' in its original form, as expressed by James Surowiecki, and its application to Web 2.0 issues, that should give us cause to pause for thought.

3.2.1 The Wisdom of Crowds

The Wisdom of Crowds is the title of a book written by James Surowiecki, a columnist for the New Yorker. In it, he outlines three different types of problem (which he calls cognition, co-ordination and co-operation), and demonstrates how they can be solved more effectively by groups operating according to specific conditions, than even the most intelligent individual member of that group. It is important to note that although Surowiecki provides caveats on the limitations to his ideas, the book's subtitle ('why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations') tends to gloss over some of the subtleties of his arguments. The book has been very influential on Web 2.0-style thinking, and several writers have adapted Surowiecki's ideas to fit their observations on Web and Internet-based activities.

An example of one of the ways in which WoC has been adapted for Web 2.0 is provided by Tim O'Reilly in his original paper (2005a). He uses the example of Cloudmark, a collaborative spam filtering system, which aggregates 'the individual decisions of email users about what is and is not spam, outperforming systems that rely on analysis of the messages themselves' (p. 2). What this kind of system demonstrates is what Surowiecki would describe as a type of cognitive decision making process, or what fans of the TV show *Who wants to be a millionaire* would call 'ask the audience'. It is the idea that, by acting independently, but collectively, the 'crowd' is more likely to come up with 'the right answer', in certain situations, than any one individual. The Cloudmark system implements an architecture of participation to harness this type of distributed human intelligence.

This is a fairly unproblematic application of Surowiecki's ideas to the Internet, but some of the wider claims are potentially more difficult to reconcile. Whilst a detailed examination of the issue is beyond the scope of this report, it is important to note that some examples that supposedly demonstrate the connective forces of WoC to Web 2.0 are really closer to collaborative production or crowdsourcing (see below) than collective 'wisdom'. As Suroweicki does not use the Web to demonstrate his concepts (although he has gone on record as saying that 'the Web is 'structurally congenial' to the wisdom of crowds'[15]) it is difficult to objectively establish how far it should be used for understanding Web 2.0 and therefore used as an accurate tool for benchmarking how 'Web 2.0' a company might be. However, regardless of this, the way in which WoC is generally understood reinforces a powerful *zeitgeist* and may therefore discourage a deep level of critical thinking. In fact, one of the interesting things about the power of this idea is the implication it may have for the traditional ways in which universities are perceived to accumulate status as 'knowers' and how knowledge can legitimately be seen to be 'acquired'.

3.2.2 Crowdsourcing: the rise of the amateur

The term *crowdsourcing* was coined by *Wired* journalist Jeff Howe to conceptualise a process of Web-based out-sourcing for the procurement of media content, small tasks, even solutions to scientific problems from the crowd gathered on the Internet. At its simplest level, crowdsourcing builds on the popularity of multimedia sharing websites such as Flickr and YouTube to create a second generation of websites where UGC is made available for re-use. ShutterStock, iStockphoto and Fotolia are examples of Web-based, stock photo or video

[15] http://www.msnbc.msn.com/id/12015774/site/newsweek/page/2/ [last accessed 14/02/07].

agencies that act as intermediaries between amateur content producers and anyone wanting to use their material. These amateur producers are often content with little or no fee for their work, taking pride, instead, from the inherent seal of approval that comes with being 'chosen'.

This type of crowdsourcing has been chipping away at the edges of the creative professions for a while now. Photographers in particular have started to feel the pinch as websites make it increasingly difficult for professionals to find a market for their work. Whilst the quality of the images may vary considerably (it is often only good enough for low-end brochures and websites) purchasers are often not able to see the poor quality or just don't care.

At the other end of the spectrum Howe demonstrates how, over the last five years or so, companies such as InnoCentive and YourEncore have been using their websites to match independent scientists and amateur or retired researchers with their clients' R&D development challenges. The individual who comes up with the solution to a particular unsolved R&D problem receives a 'prize' that runs to tens of thousands of dollars.

More recently, Canadian start-up company Cambrian House has taken the crowdsourcing model and experimented with open source software-type development models to create a model that is more closely aligned to the WoC ideal. In the Cambrian House model, members of the crowd suggest ideas that are then voted on (again, by 'the crowd') in order to decide which ones should go forward for development. This model not only sources ideas and innovations from the crowd, but also uses them to select the idea that will be the most successful, accepting that, collectively, the decision of the crowd will be stronger than any one individual's decision.

3.2.3 Folksonomies: individuals acting individually yet producing a collective result.

The term *folksonomy* is generally acknowledged to have been coined by Thomas Vander Wal, whose ideas on what a folksonomy is stem, in part, from his experience of building taxonomy systems in commercial environments and finding that successful retrieval was often poor because users could not 'guess' the 'right' keyword to use. He has, however, expressed concern in the recent past about the way the term has been mis-applied and his definition, taken from a recent blog posting, attempted to clarify some of the issues:

> 'Folksonomy is the result of personal free tagging of information and objects (anything with a URL) for one's own retrival [*sic*]. The tagging is done in a social environment (shared and open to others). *The act of tagging is done by the person consuming the information.*' [my italics].
>
> VanderWal, 2005, blog entry.

Although folksonomy tagging is done in a social environment (shared and open) Vander Wal emphasises that it is not collaborative and it is not a form of categorisation. He makes the point that tagging done by one person on behalf of another ('in the Internet space' is implied here) is not folksonomy[16] and that the value of a folksonomy is derived from people using their own vocabulary in order to add explicit meaning to the information or object they are consuming (either as a user or producer): 'The people are not so much categorizing as providing a means to connect items and to provide their meaning in their own understanding.' (Vander Wal, 2005). By aggregating the results of folksonomy production it is possible to see how additional value can be created.

[16] he describes this as 'social tagging'

Vander Wal states that the value of a folksonomy is derived from three key data elements: the person tagging, the object being tagged (as an entity), and the tag being attached to that object. From these three data elements you only need two in order to find the third. He provides an example from del.icio.us which demonstrates that if you know the object's URL (i.e. a webpage) and have a tag for that webpage, you can find other individuals that use the same tag on that particular object (sometimes known as 'pivot browsing'). This can then potentially lead to finding another person who has similar interests or shares a similar vocabulary, and this is one of Vander Wal's key points concerning what he considers to be the value of folksonomy over taxonomy: that groups of people with a similar vocabulary can function as a kind of 'human filter' for each other.

Another key feature of folksonomy is that tags are generated again and again, so that it is possible to make sense of emerging trends of interest. It is the large number of people contributing that leads to opportunities to discern contextual information when the tags are aggregated (Owen *et al*., 2006), a wisdom of crowds-type scenario. One author describes such unconstrained tagging, in the overall context of the development of hypertext, as 'feral hypertext': 'These links are not paths cleared by the professional trail-blazers Vannevar Bush dreamed of, they are more like sheep paths in the mountains, paths that have formed over time as many animals and people just happened to use them' (Walker, 2005, p. 3).

3.3 Data on an epic scale

'Information gently but relentlessly drizzles down on us in an invisible, impalpable electric rain'

von Baeyer, 2003, p.3

In the Information Age we generate and make use of ever-increasing amounts of data. Some commentators fear that this *datafication* is causing us to drown. Many Web 2.0 companies feel that they offer a way out of this, and in the emerging Web 2.0 universe, data, and lots of it, is profoundly important. Von Baeyer's invisible rain is captured by Web 2.0 companies and turned into mighty rivers of information. Rivers that can be fished.

In his original piece on the emergence of Web 2.0, Tim O'Reilly (2005a) discusses the role that data and its management has played with companies like Google, arguing that for those services, 'the value of the software is proportional to the scale and dynamism of the data it helps to manage' (p. 3). These are companies that have database management and networking as core competencies and who have developed the ability to collect and manage this data on an epic scale.

A recent article in Wired magazine emphasised the staggering scale of the data processing and collection efforts of Google when it reported on the company's plans to build a huge new server farm in Oregon, USA, near cheap hydro-electric power supplies once used to smelt aluminium (Gilder, 2006). Google now has a total database measured in hundreds of peta-bytes[17] which is swelled each day by terabytes of new information. This is the network effect working at full tilt.

Much of this is collected indirectly from users and aggregated as a side effect of the ordinary use of major Internet services and applications such as Google, Amazon and Ebay. In a sense these services are 'learning' every time they are used. As one example, Amazon will record your book buying choices, combine this with millions of other choices and then mine and sift this data to help provide targeted recommendations. Anderson (2006) calls these companies

[17] 10 to power 15 (a million, billion)

long tail *aggregators* who 'tap consumer wisdom collectively by watching what millions of them do' (p. 57).

This data is also made available to developers, who can recombine it in new ways. Lashing together applications that take rivulets of information from a variety of Web 2.0 sources has its own term—a *mash-up*. As an early, oft-quoted example, Paul Rademacher's HousingMaps.com combined Google Maps (an online mapping service) with the USA-based CraigsList of flats available for rent. These kinds of mash-ups are facilitated by what are known as 'open APIs'–Application Programming Interfaces (see section 4.5).

Much as these services have made life easier on the Web (who can imagine life without Google now?) there is a darker side. Who owns this data? Increasingly, data is seen as something – a resource – that can be repurposed, reformatted and reused. But what are the privacy implications? Google's mission is 'to organise the world's information' and in part this means *yours*. There is a tension here. Some argue that a key component of Web 2.0 is the process of freeing data, in a process of exposure and reformatting, through techniques like open APIs and mash-ups (Miller, 2005, p. 1). Others are not so sure. Tim O'Reilly makes a telling point: 'the race is on to own certain classes of core data: location, identity, calendaring of public events, product identifiers and namespaces' (2005a, p. 3). Brown and Duguid (2000) argue that the mass dis-intermediation of the Web is actually leading to centralization.

3.4 Architecture of Participation

This is a subtle concept, expressing something more than, and indeed building on, the ideas of collaboration and user production/generated content. The key to understanding it is to give equal weight to both words[18]: this is about architecture as much as participation, and at the most basic level, this means that the way a service is actually designed can improve and facilitate mass user participation (i.e. low barriers to use).

At a more sophisticated level, the architecture of participation occurs when, through normal use of an application or service, the service itself gets better. To the user, this appears to be a side effect of using the service, but in fact, the system has been designed to take the user interactions and utilise them to improve itself (e.g. Google search).

It is described in Tim O'Reilly's original paper (2005a) in an attempt to explain the importance of the decentralised way in which Bit Torrent works i.e. that it is the network of downloaders that provides both the bandwidth and data to other users so that the more people participate, the more resources are available to other users on the network. O'Reilly concludes: 'BitTorrent thus demonstrates a key Web 2.0 principle: *the service automatically gets better the more people use it*. There's an implicit 'architecture of participation', a built-in ethic of cooperation, in which the service acts primarily as an intelligent broker, connecting the edges to each other and harnessing the power of the users themselves.' (p. 2).

3.4.1 Participation and openness.

This concept pre-dates discussions about Web 2.0, having its roots in open source software development communities. Such communities organise themselves so that there are lowered barriers to participation and a real market for new ideas and suggestions that are adopted by popular acclamation (O'Reilly, 2003). The same argument applies to Web-based services. The most successful seem to be, the argument goes, those that encourage mass participation and provide an architecture (easy-of-use, handy tools etc.) that lowers the barriers to

[18] Indeed, Chris Anderson, in The Long Tail, seems to get a little confused, equating the architecture of participation to a simple blurring of the lines between producers and consumers.

participation. As a Web 2.0 concept, this idea of opening up goes beyond the open source software idea of opening up code to developers, to opening up content production to all users and exposing data for re-use and combination in so-called 'mash-ups'.

3.5 Network effects, power laws and the Long Tail

'Think deeply about the way the internet works, and build systems and applications that use it more richly, freed from the constraints of PC-era thinking, and you're well on your way.'

Tim O'Reilly, O'Reilly Radar, 10th Dec 2006.

The Web is a network of interlinked nodes (HTML documents linked by hypertext) and is itself built upon the technologies and protocols of the Internet (TCP/IP, routers, servers etc.) which form a telecommunications network. There are over a billion people online and as these technologies mature and we become aware of their size and scale, the implications of working with these kinds of networks are beginning to be explored in detail. Understanding the topology of the Web and the Internet, its shape and interconnectedness, becomes important.

There are two key concepts which have a bearing on a discussion of the implications of Web 2.0. The first is to do with the size of the Internet or Web as a network, or, more precisely, the economic and social implications of adding new users to a service based on the Internet. This is known as the Network Effect. The second concept is the power law and its implications for the Web, and this leads us into a discussion of the Long Tail phenomenon. At the heart of Tim O'Reilly's comment about the importance of the Internet as a network is the belief that understanding these effects and the sheer scale of the network involved, and working 'with the grain', will help to define who the Web 2.0 winners and losers will be.

3.5.1 The Network Effect

The Network Effect is a general economic term used to describe the increase in value to the existing users of a service in which there is some form of interaction with others, as more and more people start to use it (Klemperer, 2006; Liebowitz and Margolis, 1994). It is most commonly used when describing the extent of the increase in usefulness of a telecoms system as more and more users join. When a new telephone user joins the network, not only do they as an individual benefit, but the existing users also benefit indirectly since they can now ring a new number and speak to someone they couldn't speak to before[19]. Such discussions are not confined to telecoms and are, for example, widely referred to in relation to technology products and their markets. There is an obvious parallel with the development of social software technologies such as MySpace—as a new person joins a social networking site, other users of the site also benefit. Once the Network Effect begins to build and people become aware of the increase in a service's popularity, a product often takes off very rapidly in a marketplace.

However, this can also lead to people becoming 'locked in' to a product. A widely cited example is the great commercial success of Microsoft Office. As more and more people made use of Office (because other people did, which meant that they could share documents with an increasingly larger number of people), so it became much harder to switch to another product as this would decrease the number of people one could share a document with.

[19] There are many subtleties to network effects and interested readers are pointed to: http://oz.stern.nyu.edu/io/network.html [last accessed 15/01/07].

One of the implications of the network effect and subsequent lock-in to technology products is that an inferior product can sometimes be widely, or even universally, adopted, and the early momentum that developed behind VHS as a video format (over Betamax) is an example that is often cited. Although economists provide much nuanced argument as to the details of this (Liebowitz and Margolis, 1994) it is a powerful driver within technology marketing as it is believed that a new product is more likely to be successful in the long-term if it gains traction and momentum through early adoption. This has led to intense competition at the early adopter phase of the innovation demand curve (Farrel and Klemperer, 2006) where social phenomena such as 'word of mouth' and 'tipping point' and the human tendency to 'herd' with others play an important role (Klemperer, 2006).

As the Internet is, at heart, a telecommunications network, it is therefore subject to the network effect. In Web 2.0, new software services are being made available which, due to their social nature, rely a great deal on the network effect for their adoption. Indeed, it could be argued that their *raison d'être* is the network effect: why join MySpace unless it is to have access to as many other young people as possible in order to find new friends with shared interests? Educationalists should bear this in mind when reviewing new or proposed Web 2.0 services and their potential role in educational settings. As one lecturer recently found out, it is easier to join with the herd and discuss this week's coursework online within FaceBook (a popular social networking site) than to try and get the students to move across to the institutional VLE. There are also implications for those involved in the framing of technology standards (Farrel and Klemperer, 2006), where the need for interoperability is important in order to avoid forms of lock-in.

3.5.2 How big is the network effect?: the problem with Metcalfe's Law

How big is the network effect? Can we put a finger on the scale of its operation? The scale of the effect is important because this may have a bearing on the way the architectures of Web-based systems are designed and, in part, because discussions over the business models for new technologies that are developed on the basis of Web 2.0 ideas, see these network effects as important.

It is popularly believed that Robert Metcalfe (the inventor of Ethernet) proposed, in the early 1970s, a network effect argument whereby growth in the value of a telecommunications network, such as the Internet, is proportional to n (the number of users) squared (i.e. n^2)[20]. Metcalfe's original idea was simply to conceptualise the notion that although the costs of a telecoms network rise linearly (a straight line on the graph), the 'value' to customers rises by n^2 and therefore at some point there is a cross-over at which value will easily surpass costs, which means that a critical mass has been achieved.

Although this was originally intended as a rough empirical formulation rather than a hard physical law it was subsequently described as such ('Metcalfe's Law') in 1993 by George Gilder, a technology journalist, who was influential during the dot-com boom of the 1990s. However, recent research work has undermined this and subsequent theories that built on top of it. Briscoe *et al.* (2006) argue that these formulations are actually incorrect and that: 'the value of a network of size n grows in proportion to n log(n)' (p. 2). A growth of this scale, whilst large, is much more modest than that attributed to Metcalfe. Briscoe *et al.* further argue that: 'much of the difference between the artificial values of the dot-com era and the genuine value created by the Internet can be explained by the difference between the Metcalfe-fuelled optimism of n^2 and the more sober reality of n log(n)' (p. 2).

[20] A communications network with n users means that each can make (n-1) connections (i.e. place calls to by telephone), therefore the total value, it is argued, is n(n-1), which is roughly n^2.

It is important to appreciate how deeply entrenched Metcalfe's ideas have become. Long after the boom and bust the idea that there are 'special effects' at work on the Internet driven by the scale and topology[21] of the network remains powerful, and indeed the formula is considered by sociologists to be one of the defining characteristics of the information technology revolution or paradigm (Castells, 2000[22]). In terms of Web 2.0 this will matter again if commentators' fears of an emerging technology 'Bubble 2.0' are founded.

So why is the network effect likely to be proportional to n log(n)? The key to understanding this is to be aware that the term 'value' has been identified by Briscoe *et al.* as a rather nebulous term. What does it mean to say that the value (to me) of the telecommunications network has increased when one new person becomes a new subscriber to the telephone system or another website is added to the Web? To understand this we must delve into the shape of the Web and become aware of the role of power laws operating on it.

3.5.3 What shape is the Web?: the role of Power Laws

In addition to the physical network effects of the telecoms-based Internet, there are also Web-specific network effects at work due to the linking that takes place between pieces of Web content: every time users make contributions through blogs or use services that aggregate data, the network effect deepens. This network effect is driving the continual improvement of Web 2.0 services and applications as part of the architecture of participation.

In the previous section we saw how Briscoe *et al.* had made the argument that the size of the Network Effect was proportional to n log(n) rather than Metcalfe's n^2. They argue that this is quantitatively justified by thinking about the role of 'value' in the network: adding a new person to the network does not provide each and every other person on the network with a single unit of additional value. The additional value varies depending on what use an existing individual might make of the new one (as an example, some of your email contacts are many times more useful to you than the rest). As this *relative value* is dictated by a power law distribution, with a long tail, it can be shown mathematically that the network effect is proportional to n log(n) rather than n^2.

A power law distribution is represented by a continuously decreasing curve that is characterised by 'a very small number of very high-yield events (like the number of words that have an enormously high probability of appearing in a randomly chosen sentence, like 'the' or 'to') and a very large number of events that have a very low probability of appearing (like the probability that the word 'probability' or 'blogosphere' will appear in a randomly chosen sentence)' (Benkler, 2006). Such power law distributions have very long 'tails' as the amplitude of a power law approaches, but never quite reaches zero, as the curve stretches out to infinity[23]. This is the Long Tail referred to by Chris Anderson (see below).

[21] the 'shape' and 'connectedness' of the network

[22] Although there is, I believe, an error on page 71, where he describes the formula as n to the power of (n-1).

[23] Formally, a power law is an unequal distribution of the form $y=ax^k$ where a is a constant for large values of x, and k is the power to which x is raised – the exponent. In the graph the k^{th} ranked item will measure a frequency of about $1/k^{th}$ of the first.

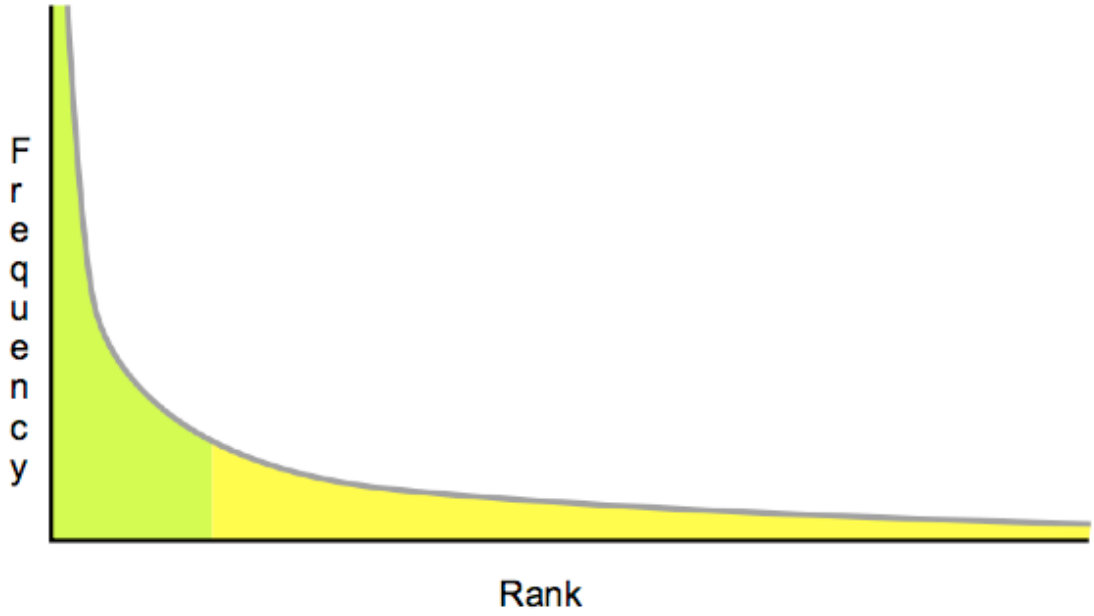

Figure 1: The Long Tail

The history of research on network effects and Web topology shows that the network effect formula is not the only facet of life on the Internet and the Web that follows a power law distribution. In fact, the shape of the Web (the way in which hypertext materials are linked) and the connection patterns of Internet routers themselves also follow a power law distribution.

3.5.4 The Long Tail

The Long Tail is the title of a book by *Wired* Editor, Chris Anderson (2006). In it, Anderson sets out to demonstrate the economic and social implications of the fact that the distribution of many facets of life on the Web is unequal and follows a power law. It transpires that not only do the physical interconnectedness of the Internet and the virtual interconnectedness of hypertext links follow a power law distribution, but, also, that many facets of the actual interaction that comes about through using tools that utilise these, also follows such a distribution pattern.

To help understand this concept, Anderson provides an example from the process of selling music albums to explain this process in the context of retailing on the Web. If one maps the number of albums sold in a particular week – the frequency – against the name of the album, it will be possible to see that the left hand side of the graph is dominated by huge sales of the popular, chart-listed albums receiving radio air-play. Often, but not always, these will be the newest albums. As one moves towards the right of the graph sales drop off dramatically, roughly according to the power law curve described above (i.e. the second highest seller will sell half the number of albums of the first). The curve continues falling away to the right, following the 1/n rule, but, and this is the crucial point outlined by Chris Anderson, only *if there is no artificial barrier to people buying less popular albums*. Artificial barriers include things like physical shelf space, which is limited and expensive, which means that only the most popular albums, or those receiving the most promotion, are stocked in shops. In a digital environment, there is no real limit to 'virtual' shelf space, so there is also no real limit to the number of albums that can be 'stocked'. Up until now, the presence of artificial barriers has cloaked the extent of the long tail.

Towards the end of the long tail the sales become smaller and smaller, in fact, tending towards zero. However, what economists have noticed is that for sales of albums, books and other artefacts, even the most unpopular items do have some sales. These are the niches at the far end of the tail. What has excited economists and business analysts is that the total sales at the lower reaches of the tail, although the items are individually unpopular, add up to a substantial amount (the area under the graph). According to Anderson, in traditional retail, new albums account for 63% of sales [in 2005], but online that percentage is reversed (36% of sales). It is therefore obvious how Amazon has used the long tail to astonishing effect. Wikipedia, too, is an excellent demonstrator of the concept as it contains tens of thousands more entries than any published, book-based encyclopaedia could ever hope to collate.

3.5.5 The Implications of Web topology

Why does this matter? What are the implications of these two topological 'rules' with regard to the developing Web 2.0 agenda? Understanding the shape of the Web and the implications of power law distribution has important implications in general for making use of the Web and the development of Internet-based technologies. It also has ramifications for debates about the role and direction of Web 2.0 technologies, in which social connections between people are a key part of the mix.

Firstly, there are implications from the development of the long tail. Chris Anderson argues that we are moving towards a culture and economy where the huge number of people participating in the niches in the tail really matters. Specialism and niche interests, personalisation and fragmentation are all potentially driven by the march rightwards on the graph. One of the forces driving this is the 'democratization' of the tools of production—the number of albums released in 2005 increased by 36% but 300,000 free tracks, many of which were produced by amateurs, were uploaded to MySpace, demonstrating the fact that 'We are starting to shift from being passive consumers to active producers' (Anderson, 2006, p. 63) and developing towards a culture which writer Doc Searls[24] calls *producerism*.

Secondly, what does topology tell us about the shape of what might be called our 'information environment'? How does this impact on the diffusion of new knowledge and the sociology of new content creation? In the Web 2.0 era in which blogs and wikis are an important part of the mix, much is made of the Internet 'conversation' afforded, particularly by the rise of the blogosphere. What does our emerging knowledge on the shape of the Web (its topology) tell us about the state of this conversation? Does the blogosphere actually work as a coherent Internet-based cultural conversation? Or is it, as some fear, a case of when everyone can speak, no-one can be heard[25], in which an uncontrolled mish-mash of conversations reduces the Web to mush.

These are the kinds of questions that Yochai Benkler attempts to tackle in his book, *The Wealth of Networks* (2006). He argues that we need an analysis of the blogosphere because it is an increasingly important tool in the dissemination of new ideas and because blogs form powerful social community-building tools. To some, this may sound like history repeating itself with echoes, for example, of past debates about Web portals concentrating power and debate in much the same way as 'old' media. But in fact, it is quite different.

Benkler's point is that the topology of the Web and the links and connections that form the conversation within the blogosphere is such that the system forms a kind of active filtration process. This means that although individually most blogs should be taken with a pinch of salt, collectively, they provide a mechanism 'for topically related and interest-based clusters

[24] Doc Searls blog: http://doc.weblogs.com/2006/01/15 [last accessed 14/02/07].
[25] What Benkler (2006) calls the *Babel objection* (p.10)

to form a peer-reviewed system of filtering, accreditation, and salience generation' (p. 252). He believes that this is proving more than an equal to mainstream media and that that while the Internet, Web and blogosphere may not be a communications utopia, it is a considerable improvement, from the point of view of political, cultural and public engagement and understanding, than traditional mass media.

Such an analysis has been made possible through a deepening understanding of the structure of information on the Web. Although the deeper subtleties of Benkler's arguments are beyond the scope of this report, and whilst you might not agree with the conclusions of his analysis as summarised here, it is wise to be aware of the context of these debates and the importance of the Web's topology to their discussion.

3.6 Openness

The development of the Web has seen a wide range of legal, regulatory, political and cultural developments surrounding the control, access and rights of digital content. However, the Web has also always had a strong tradition of working in an open fashion and this is also a powerful force in Web 2.0: working with open standards, using open source software, making use of free data, re-using data and working in a spirit of open innovation. An important technology in the development of Web 2.0 has been the open source Firefox browser and its system of extensible plug-ins which allow experimentation. Readers with an interest in exploring open source in general are referred to the JISC-funded OSSWatch service hosted at the University of Oxford[26].

3.6.1 Expose the Data

In general, Web 2.0 places an emphasis on making use of the information in the vast databases that the services help to populate. There is a parallel trend towards opening the stores of data that have been collected by public sector agencies using taxpayers' money. Readers will no doubt be aware of the wide-ranging debate within the academic and publishing communities over open access to scientific and humanities research and the role of journals in this regard, and this is not unconnected to moves within Higher Education and the research community to expose experimental data (Frey, 2006).

However, the apparent drive towards openness has to be tempered by the 'epic scale of data' that is being collected and aggregated, in non-standard ways, by commercial companies. There needs to be continual focus on open data exchange and the adoption of open standards. As Tim O'Reilly said when speaking to the Open Business forum (2006a): 'The real lesson is that the power may not actually be in the data itself but rather in the control of access to that data. Google doesn't have any raw data that the Web itself doesn't have, but they have added intelligence to that data which makes it easier to find things.'

The sharing of data is an issue within Web 2.0. Lawence Lessig recently noted the difference between 'true' sharing and 'fake' sharing, using YouTube (now Google) as an example: 'But never does the system give users an easy way to actually get the content someone else has uploaded' (Lessig, 2006). Other services are more forgiving, for example, Backpack and Wordpress both allow user data to be exported as an XML text file.

3.6.2 Open APIs.

For this discussion see the technology section.

[26] http://www.oss-watch.ac.uk [last accessed 14/02/07].

3.6.3 IPR

Web 2.0, like open source software, is starting to have an effect on intellectual property rights (IPR) and how they are perceived. One obvious example is the role of copyright. As Chris Anderson points out, the influx of 'creators' at the far end of the tail, who do not rely on being paid for their content, are choosing to give up some of their copyright protections. At the same time the scale and reach of Web 2.0 aggregators means that such systems may be republishing material for which the process of assigning the rights has been obscured: the Times Higher recently reported how UK academics had unwittingly stumbled across their own scholarly outputs available for sale on Amazon for a few dollars. Other examples include the uploading of copyright protected material to YouTube and other services.

4. Technology and standards

'The goal? To help us more easily develop the next generation of Web applications that are every bit as good as or better than desktop PC applications.'

Dion Hinchcliffe, blog post, 11th Sept. 2006.

One of the key drivers of the development of Web 2.0 is the emergence of a new generation of Web-related technologies and standards. This has been underpinned by the powerful, though not particularly new, idea of the **Web as platform**[27]. Whereas in the past, software applications ran on the user's machine, handled by a desktop operating system such as MacOS, Windows or Linux, under the Web as platform, umbrella software services are run within the actual window of the browser, communicating with the network and remote servers.

One consequence of the Web as platform is that there is less emphasis on the software (as a package: licensed and distributed) and far more on an application providing a service. The corollary of this is that there is much less emphasis on the release of software and, indeed, many well known Web 2.0 services remain in a kind of 'perpetual beta'.

So why has the idea of the Web as platform become more feasible now? The answer is that browser technology has moved on to a new stage in its development with the introduction of what are known as Rich Internet Applications (RIA)[28]. Currently the main technology for delivering RIAs is Ajax, but there are some alternatives which are mainly based on Flash technology.

N.B Tim O'Reilly's conceptualisation of Web technology with respect to Web 2.0 has since moved on to the idea of the network as platform. This is especially important for another one of his key ideas: software above the level of a single device. O'Reilly cites iTunes and TiVo as exemplars of this approach as, although not Web applications themselves, they leverage it as part of their infrastructure.

4.1 Ajax

The delivery of Web 2.0 applications and services has been driven by the widespread adoption of one particular group of technologies which are referred to as Ajax – Asynchronous Javascript + XML – a term first coined by Jesse James Garrett (Johnson, 2005; Garrett, 2005). As a term, Ajax attempts to capture both an approach to working with the Web and the use of a specific range of technologies.

One of the big frustrations for users of traditional HTML-based websites is the time spent waiting for pages to reload and refresh after the user has chosen an option or clicked on a hypertext link. Several attempts have been made over the years to improve the dynamism of webpages through individual techniques such as Javascript, hidden frames, Dynamic HTML (DHTML), CSS and Microsoft's XMLHttpRequest ActiveX tool. However, it is really only

[27] This idea was pioneered by Netscape, the company that developed one of the first successful Web browsers back in the 1990s, but eventually succumbed to competition from Microsoft, who had a vested interest in maintaining the *status quo*. This 'competition' was not without considerable controversy (see, for example, Auletta, 2001, for further details). O'Reilly (2005a) argues that the next phase will be between Windows/the desktop paradigm 'the pinnacle of proprietary control' and the open platform of the Web, and that 'battle is no longer unequal, a platform versus a single application, but platform versus platform, with the question being which platform, and more profoundly, which architecture, and which business model, is better suited to the opportunity ahead' (p. 2).

[28] For an example of the sophistication and power of these types of interfaces see the Flex demo at: http://examples.adobe.com/flex2/inproduct/sdk/dashboard/dashboard.html [last accessed 14/02/07].

with the introduction of Ajax that this has come together successfully. With Ajax, only small amounts of information pass to and from the server once the page has first been loaded. This allows a *portion* of a webpage to be *dynamically* reloaded in real-time and creates the impression of richer, more 'natural' applications with the kind of responsive interfaces that are commonly found in desktop applications (Google calendar is a good example of this).

Although Ajax is a group of technologies (see sidebar), the core is the Ajax engine, which acts as an intermediary, sitting within the client's browser and facilitating asynchronous communication with the server of smaller items of information. So, if a webpage contains a lot of text, plus, as a side-bar, a graph of the current stock price of the company being written about, this graph can be asynchronously updated in real-time without the whole page being reloaded every few seconds. The Ajax engine processes every action that would normally result in a trip back to the server for a page reload, before making any really necessary referrals back to the server. Ajax relies heavily on JavaScript and XML being accurately and efficiently handled by the browser. The need for browsers to adhere to existing standards is therefore becoming an important issue (Johnson, 2005). There is also an emerging debate with regard to the adoption of emerging standards. For example there is a debate over standards for the user interface for Ajax-style applications. Mozilla, for example, is committed to the XML User Interface (XUL) standard[29] whereas Microsoft are standing by their Extensible Application Markup Language (XAML)[30].	The Ajax technologies: – HTML/XHTML (a standards-based way of presenting information within the browser) – CSS – Document Object Model (DOM) (a way of dynamically controlling the document) – XML (data interchange and manipulation) – XSLT (data interchange and manipulation) – XMLHttpRequest (asynchronous data retrieval from the server)[31] – Javascript (or ECMA script)

A detailed overview of Ajax and its application in Web 2.0 services is provided by the Open Ajax group: http://www.openajax.org/whitepaper.html [last accessed 14/02/07].

4.2 Alternatives to Ajax

There are alternatives to Ajax, the most important of which make use of Flash—the ubiquitous graphics plug-in from Macromedia (now Adobe) that first appeared in the 1990s. It allowed sophisticated, but quick-to-download, vector graphics and animation to be displayed in the browser window. Flash requires a browser plug-in to work, although within only a few years of its launch 99% of computers had the necessary addition to support it.

Flash is still being used to deliver compelling content within the browser (in fact the Flash video player is beginning to take off because YouTube have adopted it). It has been used as

[29] A mark-up language for user interface graphics. See: http://www.xulplanet.com/
[30] http://msdn2.microsoft.com/en-us/library/ms752059.aspx [last accessed 14/02/07].
[31] XMLHttpRequest object is implemented in most popular Web browsers and presents a simple interface that allows data to be transferred from the client to the server and vice versa, while the user continues to interact with the webpage. See: http://www.javaworld.com/javaworld/jw-10-2005/jw-1017-ajax.html?page=2 [last accessed 14/02/07].

the basis of other RIA development tools, including Adobe's Flex and OpenLaszlo. Developers in HE/FE might be particularly keen on OpenLaszlo as it uses an open source model: OpenLaszlo programs are written in XML and JavaScript and then transparently compiled to both Flash and non-proprietary Dynamic HTML.

As well as these Flash-based systems there are several emerging technologies which focus on displaying rich graphics within the browser window. These include Microsoft's WPF/E[32], XBAP, and the related XAML[33] (all of which feature heavily in the Vista operating system); Mozilla's XUL; and Ethan Nicholas's proposed, minimalist Java Browser Edition (Hinchcliffe, 2006).

The introduction of these alternative RIA technologies is not without controversy and debate amongst developers. Some of these solutions require the addition of a plug-in to the browsers and make use of core technology that is proprietary. There is also some concern that the approach taken by these products is 'breaking the model of the web' (Hinchcliffe, 2006 p. 1).

4.3 SOAP vs REST: A Web architecture debate

'At the heart of REST is the idea that the web works precisely because it uses a small number of verbs applied to a large number of nouns.'

McGrath, 2006.

A further strand in the development of Web technology is the use of what are called *lightweight* or *simplified programming models,* which facilitate the creation of loosely coupled[34] systems. This flexibility is a source of debate since, the lightweight 'ideal' is often viewed in contrast to the production of more robust Web Services which use what are seen as the 'heavyweight' and rather formal techniques of SOAP and WS-*. This debate is focused as much on issues of genre and style of programming practice and development techniques as it is on the mandating of any particular technology, although the use of scripting languages such as Perl, Python, PHP and Ruby, along with technologies such as RSS, Atom and JSON[35] is one of the favourite ways of (lightweight) working.

Without going into this in too much depth, readers should be aware that these discussions about style within the Web development community are crystallising around two main approaches: REST and SOAP. This can be seen in a wider context of a generalised, on-going debate within technology circles over simplicity vs. sophistication. REST stands for Representational State Transfer, an *architectural* idea and set of principles first introduced by Roy Fielding (Costello, 2005). It is not a standard, but describes an approach for a client/server, stateless architecture whose most obvious manifestation is the Web and which provides a simple communications interface using XML and HTTP. Every resource is identified by a URI and the use of HTTP lets you communicate your intentions through GET, POST, PUT, and DELETE command requests. SOAP and WS-*, on the other hand, are more formal and use messaging, complex protocols and Web Services Description Language (WSDL).

One way of visualising the ensuing debate is provided by Sean McGrath. He describes the Web as an enormous information space, littered with nouns (that can be located with URIs) and a small number of verbs (GET, POST etc). Where SOAP is more of a Verb Noun system,

[32] Windows Presentation Foundation is the graphical subsystem feature of .NET Framework 3.0
[33] Extensible Application Markup Language (XAML: pronounced "Zammel")
[34] loosely coupled entities make few assumptions about each other, limit dependencies and employ communications techniques that allow for flexibility and for one end to change without affecting the other.
[35] JavaScript Object Notation, see Johnson (2005) for details

he argues that SOAP/WSDL allows the creation of too many (irregular) verbs (McGrath, 2006). There is considerable debate between communities of developers over these issues.

4.4 Microformats

Microformats are widely used by Web developers to embed semi-structured semantic information (i.e. some level of 'meaning') within an XHTML webpage (Khare, 2006). Information based on open data formats (a microformat) is buried within certain XHTML tags (such as 'class' or 'div') or attributes (such as 'rel' or 'rev'). The information is not used by the browser for display or layout purposes but it can be picked up by applications such as search engines[36].

An example of a microformat is the hCard format which allows personal or organisational contact information based on the vCard standard to be embedded in a webpage[37]. Proponents argue that microformats will have significant benefits for the development of the Web because they will allow bloggers or website owners to embed information that services and applications can make use of without the need to go and visit the application's website and add the data.

Of course, to a certain extent, Web search engines already do this when they crawl a website or blog and index the content for other people to locate. Microformats provide additional information for these kinds of services. As an example, provision of information in the hListing microformat (which is for small ads) on a blog would allow a small ads service (such as Craigslist) to automatically find your listing. Future versions of the Firefox browser (possibly version 3) are likely to incorporate functionality that makes use of microformats in order to automatically move such data into one's chosen applications or online services (for example moving any contact information buried in a webpage into Gmail contacts list)–a process described as being more 'information broker' than browsing (Wagner, 2007). An illustration from Mozilla shows clearly how this vision fits with the Web as Platform idea[38]:

The use of microformats is not without its detractors and debates around this subject tend to be centred around whether they a) help or hinder the process of moving Web content towards

[36] See: http://microformats.org/about/ [last accessed 14/02/07].
[37] See: http://microformats.org/wiki/hcard. For those interested in the detail of an implementation of a hCard in a webpage, see the tutorial at: http://usabletype.com/weblog/2005/usable-microformats/ [both last accessed 14/02/07].
[38] http://people.mozilla.com/~faaborg/files/20061213-fundamentalTypes/informationBroker.jpg_large.jpg [last accessed 14/02/07].

the Semantic Web vision (they are sometimes referred to as the 'lowercase semantic web'[39]) (Khare and Celik, 2006) and b) have bearing on the on-going and wide-ranging discussions over the merits or otherwise of the use of lightweight (REST etc.) or heavyweight (SOA etc.) approaches and solutions.

4.5 Open APIs.

"When I hear the word open used for services and APIs, I cringe, Just because something's available on the Internet, is it 'open'?"

Brian Behlendorf[40], in: Prodromou, 2006, p. 4.

An Application Programming Interface (API) provides a mechanism for programmers to make use of the functionality of a set of modules without having access to the source code. An API that doesn't require the programmer to license or pay royalties is often described as *open*. Such 'open' APIs have helped Web 2.0 services develop rapidly and have facilitated the creation of mash-ups of data from various sources.

One way of finding out what APIs are available is to look at the Programmable Web website (http://programmableweb.com/), which keeps track of the number of APIs and what people are doing with them (it recently registered over three hundred). One of the key examples is the Google Maps API, which allows Web developers to embed maps within their own sites (http://www.google.com/apis/maps/). Programmable Web claims that over 50% of data mash-ups use Google Maps. Amazon has also started to allow access to its database through Amazon Web Services (AWS[41]) API.

However, there has been considerable debate over what constitutes 'openness'. Increasingly the discussions have moved beyond the parameters of open source software *per se* and into discussing what open means in the context of a Web-based service like Google (O'Reilly, 2006b). Some argue that for a service it is the data rather than the software that needs to be open and there are those that hold that to be truly open the user's data should be able to be moved or taken back by the user at will. Tim Bray, an inventor of XML, argues that a service claiming to be open must agree that: 'Any data that you give us, we'll let you take away again, without withholding anything, or encoding it in a proprietary format, or claiming any intellectual-property [*sic*] rights whatsoever.' [42]

[39] For more on this debate see Brian Kelly: http://www.ariadne.ac.uk/issue44/web-focus/#8 [last accessed 14/02/07].
[40] One of the founding members of the Apache Group, which became the Apache Software Foundation.
[41] http://www.amazon.com/AWS-home-page-Money/b/ref=sc_fe_l_1_3435361_1/002-3264884-9188051?ie=UTF8&node=3435361&no=3435361&me=A36L942TSJ2AJA [last accessed 14/02/07].
[42] http://www.tbray.org/ongoing/When/200x/2006/07/28/Open-Data [last accessed 14/02/07].

5. Educational and Institutional Issues

There is significant debate over the alleged advantages and disadvantages of incorporating social software into mainstream education. This is compounded by the fact that there is very little reliable, original pedagogic research and evaluation evidence and that to date, much of the actual experimentation using social software within higher education has focused on particular specialist subject areas or research domains (Fountain, 2005). Indeed, JISC recently announced an open call to investigate the ways that this technology is being used by staff and students and identify opportunities for integration with existing institutional IT systems[43]. In this section we review some examples of preliminary activity in four areas: learning and teaching, scholarly research, academic publishing, and libraries.

5.1 Teaching and learning

One of the most in-depth reviews undertaken in the UK of the potential impact of social software on education has been carried out by the Nesta-funded FutureLab. Their recent report, *Social Software and Learning* (Owen *et al*., 2006), reviews the emerging technologies and discusses them in the context of parallel, developing trends in education. These trends tend towards more open, personalised approaches in which the formal nature of human knowledge is under debate and where, within schools and colleges, there is a greater emphasis on lifelong learning and supporting the development of young people's skills in creativity and innovation.

Within higher education, wikis have been used at the University of Arizona's Learning Technologies Centre to help students on an information studies course who were enrolled remotely from across the USA. These students worked together to build a wiki-based glossary of technical terms they learned while on the course (Glogoff, 2006). At the State University of New York, the Geneseo Collaborative Writing Project deploys wikis for students to work together to interpret texts, author articles and essays, share ideas, and improve their research and communication skills collectively[44]. Using wikis in this way provides the opportunity for students to reflect and comment on either their work or others. Wiki-style technology has also been used in a tool developed at Oxford University to support teachers with 'design for learning'[45].

Bryan Alexander (2006) describes social bookmarking experiments in some American educational research establishments and cites Harvard's H2O as an exemplar project[46]. Alexander also believes that wikis can be useful writing tools that aid composition practice, and that blogs are particularly useful for allowing students to follow stories over a period of time and reviewing the changing nature of how they are commented on by various voices. In these scenarios, education is more like a conversation and learning content is something you perform some kind of operation on rather than 'just' reading it.

In the UK, Warwick University has provided easy to use blogging facilities to allow staff and students to create their own personal pages. The intention is that the system will have a

[43] http://www.jisc.ac.uk/fundingopportunities/funding_calls/2007/01/web_2_use.aspx [last accessed 14/02/07].
[44] http://node51.cit.geneseo.edu/WIKKI_TEST/mediawiki/index.php/Main_Page [last accessed 14/02/07].
[45] http://phoebe-app.conted.ox.ac.uk/cgi-bin/trac.cgi/wiki/WikiStart [last accessed 14/02/07].
[46] H20 provides for shared playlists (shared lists of readings, blog postings, podcasts and other content), which can be tagged and subscribed to as RSS feeds. Playlists can be compiled by anyone and are published under the Creative Commons. See: http://h2o.law.harvard.edu/index.jsp [last accessed 14/01/07].

variety of education-related uses such as developing essay plans, creating photo galleries and recording personal development[47].

But these developments are not without debate. Apart from concerns around learner attention (in an 'always-on' environment), identity, the emerging digital divide between those with access to the necessary equipment and skills and those who do not, there are other, specific, tensions. While some experts focus on the idea of 'self production' to argue that learners find the process of learning more compelling when they are producers as much as consumers[48], others argue that the majority of learners are not interested in accessing, manipulating and broadcasting material. Indeed, there is serious concern that 'techno-centric' assumptions will obscure the fact that many young people are so lacking in motivation to engage with education that once these new technologies are integrated into the education environment, they will lose their initial attraction.

It is beyond the limited scope of a TechWatch report to do real justice to the wide-ranging debate over of the pedagogical issues but it is perhaps important to point out some of the implications that these issues will have for education in the same way as other sectors:

- there is a lack of understanding of students' different learning modes as well as the 'social dimension' of social software. In particular, more work is required in order to understand the social dimension and this will require us to really 'get inside the heads of people who are using these new environments for social interaction' (Kukulska-Hulme, 2006, 16:50).
- Web 2.0 both provides tools to solve technical problems and presents issues that raise questions. If students arrive at colleges and universities steeped in a more socially networked Web, perhaps firmly entrenched in their own peer and mentoring communities through systems like MySpace, how will education handle challenges to established ideas about hierarchy and the production and authentication of knowledge?
- How will this affect education's own efforts to work in a more collaborative fashion and provide institutional tools to do so? How will it handle issues such as privacy and plagiarism when students are developing new social ways of interacting and working? How will it deal with debates over shared authorship and assessment, the need to always forge some kind of online consensus, and issues around students' skills in this kind of shared and often non-linear manner of working, especially amongst science/engineering students (Fountain, 2005).

One area where this is already having an impact is the development of Virtual Learning Environments (VLEs). Proponents of institutional VLEs argue that they have the advantage of any corporate system in that they reflect the organisational reality. In the educational environment this means that the VLE connects the user to university resources, regulations, help, and individual, specific content such as modules and assessment. The argument is that as the system holds this kind of data there is the potential to tailor the interface and the learning environment (such as type of learning resources, complexity of material etc.) to the individual, particularly where e-learning is taking place, although so far relatively little use has been made of, for example, usage statistics of VLEs or tailored content to substantiate these claims.

However, others now question whether the idea of a Virtual Learning Environment (VLE) even makes sense in the Web 2.0 world. One Humanities lecturer is reported as having said: "I found out all my students were looking at the material in the VLE but going straight to

[47] http://www2.warwick.ac.uk/services/its/elab/services/webtools/blogs/about/

[48] Cych (2006) cites the work of Steven Heppel and his ideas on *symmetry* and *participation* to argue that social software technologies help to develop this kind of collaborative production.

Facebook[49] to use the discussion tools and discuss the material and the lectures. I thought I might as well join them and ask them questions in their preferred space."[50]

Partly in response to these concerns, there has been research and discussion devoted to the development of a more personalised version of the VLE concept – PLEs – to make use of the technologies being developed in order to bring in social software and e-portfolios (Wilson, 2006).

5.2 Scholarly Research

Tim Berners-Lee's original work to develop the Web was in the context of creating a collaborative environment for his fellow scientists at CERN and in an age when inter-disciplinary research, cutting across institutional and geographical boundaries, is of increasing relevance, simple Web tools that provide collaborative working environments are starting to be used. The open nature of Web 2.0, its easy-to-use support for collaboration and communities of practice, its ability to handle metadata in a lightweight manner and the non-linear nature of some of the technology (what Ted Nelson once called *intertwingled*[51]) are all attractive in the research environment (Rzepa, 2006) and there are four specific technology areas which have seen uptake and development:

Firstly, folksonomies are starting to be used in scientific research environments. One example is the CombeChem work at Southampton University which involved the development of a formal ontology for laboratory work which was derived from a folksonomy based on established working practices within the laboratory[52]. However, there is, to put it mildly, some debate about the role and applicability of folksonomies within formal knowledge management environments, not least because of the lack of semantic distinction between the use of tags. A recent JISC report *Terminology services and technology* (Tudhope *et al*., 2006) reviewed some of the characteristics of 'social tagging' systems and the report notes that 'Few evaluative, systematic studies from professional circles in knowledge organisations, information science or semantic web communities have appeared to date' (p. 39). Issues raised by the JISC report include the obvious lack of any control over the vocabulary at even the most basic level (for example, word forms – plural or singular – and use of numbers and transliteration) and goes on to highlight shortcomings related to the absence of rules in the tagging process, for example, on the granularity or specificity of tags. The main recommendation of the report is that social tagging should not replace indexing and other knowledge organisation efforts within HE/FE. There are also specific recommendations (see pages 40–43) which are beyond the scope of this report.

Some researchers are, however, beginning to investigate whether it could be fruitful to combine socially created tags with existing, formal ontologies (Al-Khalifa and Davis, 2006). Tagging does provide for the marking up of objects in environments where controlled indexing is not taking place, and as the tagging process is strongly 'user-centric', such tagging can reflect topicality and change very quickly. We are also now starting to see folksonomies being developed alongside expert vocabularies as a way of enabling comparative study e.g. of the meaning-making process around artworks[53]. We are also beginning to see compromise solutions known as *collabulary* in which a group of domain users and experts collaborate on a shared vocabulary with help of classification specialists.

---

[49] http://www.facebook.com/ A popular social networking site
[50] Comment by attendee at ALT-C, 2006 (anonymous). Taken with thanks from private notes made by Lawrie Phipps at JISC ALT-C stand.
[51] to express the deep inter-connectedness and complexity of knowledge.
[52] see: http://www.combechem.org/tour.php?tourpage=onto.html [last accessed 14/01/07].
[53] see the Steve museum project: http://www.steve.museum/ [last accessed 12/01/07].

Secondly, although evidence is only anecdotal, blogging seems to be becoming more popular with researchers of all disciplines in order to engage in peer debate, share early results or seek help on experimental issues (Skipper, 2006). However, it has had no serious review of its use in higher education (Placing, 2005). Butler (2005) argues that blogging tends to be used by younger researchers and that many of these make use of anonymous names to avoid being tracked back to their institutions. Some disciplines are so fast-moving, or of sufficient public interest, that this kind of quick publishing is required (Butler cites climate change as one example).

There has also been a trend towards collective blogs (Varmazis, 2006) such as ScienceBlogs[54] and RealClimate[55], in which working scientists communicate with each other and the public, as well as blog-like, peer-reviewed sites such as Nature Protocols[56]. These tools provide considerable scope to widen the audience for scientific papers and to assist in the process of public understanding of science and research (Amsen, 2006). Indeed, Alison Ashlin and Richard Ladle (2006), argue that scientists need to get involved in the debates that are generated across the blogosphere where science discussions take place. These tools also have the potential to facilitate communication between researchers and practitioners who have left the university environment.

Thirdly, social tagging and bookmarking have also found a role in science (Lund, 2006). An example of this approach is CiteULike[57] a free service to help academics share, store, and organise the academic papers they are reading.

Finally, there have also been developments in scientific data mash-ups and the use of Web Services to link together different collections of experimental data (Swan, 2006). Examples include AntBase[58] and AntWeb, which use Web Services to bring together data on 12,000 ant species, and the USA-based water and environmental observatories project (Liu *et al*., 2007). This corresponds to moves in recent years to open up experimental data and provide it to other researchers as part of the process of publication (Frey, 2006) and the Murray-Rust Research Group is particularly well known for this[59]. The E-bank project is also looking at integrating research experiment datasets into digital libraries[60].

However, opinion is divided over the extent to which social software tools are being used by the research community. Declan Butler, for a recent article in Nature (2005), conducted interviews with researchers working across science disciplines and concluded that social software applications are not being used as widely as they should in research, and that too many researchers see the formal publication of journal and other papers as the main means of communication with each other.

5.3 Academic publishing

Speed of communication in fast-moving disciplines is also a benefit offered to academic publishing, where social software technologies increasingly 'form a part of the spectrum of legitimate, accepted and trusted communication mechanisms' (Swan, 2006, p. 10). Indeed, in the long run, the Web may become the first stage to publish work, with only the best and most durable material being published in paper books and journals, and some of this may introduce a beneficial informality to research (Swan, 2006).

[54] http://www.scienceblogs.com/channel/about.php [last accessed 14/01/07].
[55] http://www.realclimate.org/index.php/archives/2004/12/about/ [last accessed 14/01/07].
[56] http://www.nature.com/nprot/prelaunch/index.html [last accessed 14/01/07].
[57] http://www.citeulike.org/
[58] http://www.antbase.org/
[59] See: http://wwmm.ch.cam.ac.uk/wikis/wwmm/index.php/Main_Page [last accessed 14/02/07].
[60] http://www.ukoln.ac.uk/projects/ebank-uk/ [last accessed 14/02/07].

Such developments are obviously closely tied up with the Open Access debate and the need to free data in order to provide other researchers with access to that data: these datasets will need to be open access before they can be mashed. Those involved in the more formal publishing of research information are actively working on projects that make use of Web 2.0 technologies and ideas. For example, Nature is working on two developments: Open Text Mining Interface (OTMI) and Connotea, a system which helps researchers organize and share their references[61].

Some publishers are also experimenting with new methods of a more open peer reviewing process (Rogers, 2006). Once again, Nature is devoting resources to a system where authors can choose a 'pre-print' option that posts a paper on the site for anyone to comment on, whilst in the meantime the usual peer-reviewing processes are going on behind the scenes. Another website, arXiv[62], has also been providing pre-publication papers for colleagues to comment on. In addition, the SPIRE project[63] provides a peer-to-peer system for research dissemination.

5.4 Libraries, repositories and archiving

As with other aspects of university life the library has not escaped considerable discussion about the potential change afforded by the introduction of Web 2.0 and social media (Stanley, 2006). Berube (2007) provides a very readable summary of some of the implications for libraries and there have been debates about how these technologies may change the library, a process sometimes referred to as 'Library 2.0' a term coined by Mike Casey (Miller, 2006).

Proponents argue that new technologies will allow libraries to serve their users in better ways, emphasise user participation and creativity, and allow them to reach out to new audiences and to make more efficient use of existing resources. Perhaps the library can also become a place for the production of knowledge, allowing users to produce as well as consume? Others worry that the label is a diversion from the age-old task of librarianship.

However, what is interesting about many of these debates is that they are very broad, sometimes contradictory, and much of the discussion can often be seen in the context of the wider public debate concerning the operation of public services in a modern, technology-rich environment in which user expectations have rapidly changed (Crawford, 2006), rather than Web 2.0 *per se*. For example, comparison has been made between Amazon's book delivery mechanisms and the inter-library loans process (Dempsey, 2006). People worry that library users expect the level of customer service for inter-library loans to be comparable to Amazon's, and while this is obviously an important aspect of what Amazon provides, it is not one of its Web 2.0 features.

This is not to say that there is no genuinely Web 2.0-style thinking going on within the Library 2.0 debate (for example, in the USA, the Ann Arbor public library online catalogue utilises borrowers' data to produce an Amazon-style, 'readers who borrowed this book, also borrowed' display feature[64] and John Blyberg's Go Go Google Gadget[65], which uses data mash-ups to provide a personalised Google homepage with library data streams showing

[61] See: http://blogs.nature.com/wp/nascent/2006/04/web_20_in_science.html for further details [last accessed 14/02/07].
[62] http://arxiv.org/
[63] http://spire.conted.ox.ac.uk/cgi-bin/trac.cgi [last accessed 28/01/07].
[64] available at: http://www.aadl.org/cat/seek/record=1028781 [last accessed 28/01/07] (you will need to scroll to the bottom of the page). See also LibraryThing: http://www.librarything.com/ [last accessed 28/01/07].
[65] available at: http://www.blyberg.net/2006/08/18/go-go-google-gadget/ [last accessed 28/01/07].

popular lendings, items you have checked out, etc.), only that it might be helpful for librarians, in terms of thinking about the future of libraries, to separate out the Web 2.0 ideas, services and applications from the technology and more general concerns about 'user-centred change'. How, for example, might libraries take part of the ethos of the long tail (everything has a value that goes beyond how many times it is requested) and not only learn from the way Amazon has applied it, but perhaps even better it?

This idea is not without precedent, especially in areas where traditional library skills and processes can be mapped to the development of Web 2.0-style applications and services, and information retrieval (IR) is an interesting case in point. Mark Hepworth (2007) argues that tagging is a form of indexing, blog trackbacking is similar to citation analysis, blog-rolling echoes chaining and RSS syndication feeds can be considered a form of 'alerting'—all recognised concepts within discussions of IR. This is not to say that they are necessarily the same: whereas traditional IR normally works with an index based on a closed collection of documents, Web searching involves a different type of problem with an enormous scale of documents/pages, a dynamic document base, huge variety of subject domains and other factors (Levene, 2006). However, we can say that the thinking and discussion that has taken place within IR both in traditional systems and more recently in the context of the Web in general (Gudiva, 1997) will have some bearing on an understanding of Web 2.0 services and applications. It may even be the case that Web 2.0 ideas and applications can contribute solutions to some of the recognised existing problems within IR with regard to user behaviour and usability issues (Hepworth, 2007), and even that the newer Web technologies such as RIA may be harnessed to help the user or learner to organise and view data or information more effectively.

Another reason why it may be important to think about the ideas behind Web 2.0 is in the issue of the archiving and preservation of content generated by Web 2.0-style applications and services.

5.4.1 Collecting and preserving the Web

'The goal of a digital preservation system is that the information it contains remains accessible to users over a long period of time.'

Rosenthal, 2005, section 2.

'The most threatened documents in modern archives are usually not the oldest, but the newest.'

Brown and Duguid, 2000 p. 200

The Web is an increasingly important part of our cultural space and for this reason the archiving of material and the provision of a 'cultural memory' is seen as a fundamental component of library work (Tuck, 2007), and there has been considerable discussion, debate and research work undertaken in this area (Tuck, 2005a; Lyman, 2002). At the British Library it is the policy that 'the longer term aim is to consider web-sites [*sic*] as just another format to collect within an overall collection development policy' (Tuck, 2005a). However, there are many issues to consider with regard to the archiving and preservation of digital information and artefacts in general, and there are also issues which are particularly pertinent to the archiving and preservation of the Web (Mesanès, 2006). Currently, the only large-scale preservation effort for the open Web is the Internet Archive[66], although there are a number of small-scale initiatives that focus on particular areas of content (e.g. the UK Web Archive

[66] http://www.archive.org/index.php

Consortium, which focuses on medical, Welsh, cultural and political materials of significance[67]).

Within the UK, the UK Web Archiving Consortium (UKWAC) is engaging with the technical, standards and IPR related issues for collection and archiving of large scale parts of the UK Web infrastructure (Tuck, 2005b). This work has included the initial use of archiving software developed in Australia (Pandas), the development of a Web harvesting management system (Web Curator Tool) and investigation work into the longer-term adoption new standards, such as the emerging WARC storage format for Web archiving (Beresford, 2007).

There have also been a number of reports considering the issue of preservation of the Web. In 2003, for example, JISC and the Wellcome Trust prepared a report on general technical and legal issues (Day, 2003) and UKOLN recently developed a general roadmap for the development of digital repositories, which should be considered when reviewing the difficulties of preserving newer Web material (Heery, 2006).

The Day report (2003) outlined two phases to the process of preserving Web content: collection and archiving. Collection encompasses automatic harvesting (using crawler technologies); selective preservation, which uses mirror-sites to replicate complete websites periodically; and asking content owners to deposit their material on a regular basis. Secondly, there is the process of archiving where a respected institution creates a record of the material collected and provides access for future users.

However, part of the problem for the process of preservation is that the Web has a number of issues associated with it which make it a non-trivial problem to develop archiving solutions (Masanès, 2006; Day, 2003; Lyman, 2002; Kelly 2002). For example:

5.4.1.1 The Web is transient.

The Web is growing very rapidly, is highly distributed but also tightly interconnected (by hyperlinks) and on a global scale. This makes the overall topology of the Web transient and it becomes extremely difficult to know what's 'out there'—its true scope. In addition, the average life span of webpages is short: 44 days in Lyman (2002, p. 38) and 75 days in Day (2006, p. 177). Dealing with this ephemerality is difficult, especially when combined with the fact that the Web can be considered an active publishing system (Masanès, 2006) in that content changes frequently and can be combined and aggregated with content from other information systems.

5.4.1.2 Web technologies are not always conducive to traditional archiving practices.

Problems with archiving the Web are inherently caught up with technology issues. At a very basic level, as with all digital content, Web content is deeply entangled with or dependent on technology, protocols and formats. For example, the average page contains links to five sourced objects such as embedded images or sound files with various formats: GIF, JPEG, PNG, MPEG etc. (Lyman, 2002). These protocols and formats evolve rapidly and content that doesn't migrate will quickly become obsolete. In addition, information is always presented within the context of a graphical look and feel which 'evokes' a user experience (Lyman, 2002) and content may even be said to exhibit a 'behaviour' (Day, 2006). This varies according, in part, to the particular browser/plug-in versions in use and it is often argued that preservation should attempt to retain this context. It is the difference between what Clay

[67] A consortium of Wellcome, British Library and National Library of Wales http://info.webarchive.org.uk/index.html

Shirky calls 'preserving the bits' and 'preserving essence'[68]. With this in mind, how do we go about migrating not only the data but also the manner in which it was presented?

However, technology issues also go much deeper. Web content's *cardinality*[69] (an important concept in preservation) is not simple. A webpage's cardinality might be considered to be one, as it is served by a single Web server and its location is provided by the unique identifier, the URL. Masanès (2006) argues this means that, in archiving terms, it is more like a work of art than a book and is subject to similar vulnerabilities, as the server can be removed or updated at any time. However, this is further complicated by the fact that a webpage's cardinality can be considered one and it can be many, at the same time. A large, perhaps almost unlimited, number of visitors can obtain a 'copy' of the page for display within their browser (an *instantiation*) and the actual details of the page that is served may well vary each time[70]. This complex cardinality is an issue for preservation in that it means that a webpage permanently depends on its unique source (i.e. the publisher's server) to exist.

In addition, they way HTTP works poses problems for archiving as it provides information on a request-by-request basis, file by file. It cannot, unlike FTP, be asked to provide a list of the whole set of files on a server or directory. This means that there is an extra layer of effort involved as the extent of a website has to be uncovered before it can be archived. This problem can be extrapolated to the whole of the Web.

The main method for gathering this information about the extent of a website, either for search engine indexing or for archiving, is to follow the paths of links from one page to another (so-called 'crawling') and there are two main issues with this:

- Websites can issue 'politeness' notices (in robots.txt files on the server) using the Robots Exclusion Protocol (Levene, 2006). These notices issue instructions about the manner in which crawling can be carried out and might, for example, restrict which parts of a site can be visited or impose conditions as to how often a crawl can be carried out.
- Robot crawlers may not actually reach all parts of the Web and this leaves some pages or even whole websites un-archived. There are two main reasons for this:
    - some websites are never linked to anything else
    - a large proportion of the Web cannot be reached by crawling as the content is kept behind password-protected front-ends or is buried in databases in what is known as the 'deep', 'hidden' or 'invisible' Web (Levene, 2006). Levene estimates that the size of this hidden Web is perhaps 400 to 550 times the extent of standard webpages.

Content in the 'hidden Web' needs a specific set of user interactions in order to access it and such access is difficult to automate. Some, limited, headway has been made with this problem by attempting to replicate these human actions with software agents that can detect HTML forms and learn how to fill them in, using what are known as hidden Web agents (Masanès,

---

[68] See: http://discuss.longnow.org/viewtopic.php?t=39 and http://video.google.com/videoplay?docid=4000153761832846346&q=longnow.org&pl=true [last accessed 28/01/07].

[69] In simple terms the number of instances (or copies) of each work that are available to deal/work with. In the traditional case of a book, a number of copies, maybe 2,000 of each edition are published, printed and distributed (each of which is the same in terms of content). There is no need for an archive to use a particular one of these copies in order to preserve a representation of that edition. In this instance, the book's cardinality would be 2,000.

[70] A simple example: Many website homepages graphically display the current time and date. If we take a copy of that page then it is unique on the date and at the time shown, but will not be the same on the next visit.

2006). One alternative requires direct collaboration with a site's owner, who agrees to expose the full list of files to an archive process through a protocol such as OAI-MHP[71]. Another alternative, which saves the site's owner from setting up a protocol and which is useful for websites that offer a database gateway which holds metadata about a document collection, is to extract (*deep mine*) the metadata directly from the database and archive it, together with the documents, in an open format. In effect, the database has been replaced, at the archive, by an XML file. This is the approach being facilitated by the deepArc tool that is being developed by the Bibliothèque Nationale de France as part of the International Internet Preservation Consortium (IIPC)[72].

5.4.1.3 Legal issues pertaining to preservation and archiving are complex

Day (2003) argues that another major problem that relates to Web archiving is its legal basis. In particular, there are considerable intellectual property issues involved in preserving databases (as opposed to documents) which are compounded by general legal issues surrounding copyright, lack of legal deposit mechanisms, liability issues relating to data protection, content liability and defamation that pose problems for the collection and archiving of content.

5.4.2 Preserving content produced through Web 2.0 services and applications.

As we have seen, there are considerable issues around the long-term preservation of the Web, but how do these issues change with the introduction of Web 2.0 ideas and services?

Material produced through Web 2.0 services and applications is clearly dynamic, consisting of blog postings, data mash-ups, ever-changing wiki pages and personal data that have been uploaded to social networking sites. Some would argue that much of this content is of limited value and does not warrant significant preservation efforts. On the other hand, Web 2.0 material is still part of the Web and others argue that since the Web is playing a major role in academic research, scientific outputs and learning resources there is a strong case for preserving at least some of it (Day, 2003) and a clear argument is now developing for the preservation of blogs and wikis (Swan, 2006). Blogs in particular clearly form part of a conversation that is increasingly part of our culture. From the point of view of education, increasingly, published academic research will make reference to Web 2.0-type material, for example, a peer group wiki focused on an experiment.

There are two key questions one can ask of Web 2.0 with regard to preservation. Firstly, to what extent does Web 2.0 content form part of the hidden Web? Most Web-based archiving tools make use of crawler technology and the issue here is whether the Web is evolving towards an information architecture that 'resists traditional crawling techniques' (Masanès, 2006, p. 128). Getting at the underlying data that is being used in a wide variety of Web 2.0 applications is a major problem: many Web 2.0 services and mash-ups use layered APIs which sit on top of very large dynamic databases. Unfortunately, technology to allow the preservation of data from a dynamic database is only just beginning to be developed[73]. This might involve the development of some kind of 'wayback machine' that reconstructs a database's state at a specific time (Rosenthal, 2006).

In addition, the APIs used by many of the Web 2.0 systems are often described as open, but they are, in fact, proprietary and subject to change; much of Web 2.0 is in perpetual beta and

[71] Open Archive Initiative Metadata Harvesting Protocol
[72] http://netpreserve.org/about/index.php [last accessed 14/02/07].
[73] Peter Buneman at the University of Edinburgh has begun to develop the basic concept. See the references section for a selection of his work.

preservation mechanisms that make use of these interfaces would need to be able handle this kind of change.

Secondly, how important is it to capture the graphical essence of Web 2.0 content and is this technically possible? Many Web 2.0 services utilise a strong graphical look and feel in order to create a powerful user experience and this is often more substantial than the constituent raw data[74]. There have been discussions within the repositories community about the problems inherent in capturing this in an archive[75].

5.4.2.1 Web 2.0 ideas and preservation issues

In the following section we review and discuss the particular characteristics of content produced by Web 2.0-type services and their implications for preservation and archiving in the context of the six ideas that have been developed elsewhere in this report. Secondly, we look at the individual categories of Web 2.0 service and the characteristics that may inform debate about the manner in which they could be preserved. This is very much a work-in-progress and should be seen as a springboard for discussion and further development within the higher education community.

The key questions with regard to Web 2.0 are: is the content produced by Web 2.0 services sufficiently or fundamentally different to that of previous Web content and, in particular, do its characteristics make it harder to preserve and archive? Are there areas where further work is needed by researchers and library specialists? Firstly, the six ideas that underpin Web 2.0 can be examined and reviewed with regard to their impact on preservation:

Table 2: Impact of each of the six ideas of Web 2.0 on preservation

	Key Concept	Initial thoughts on impact on preservation
1	Individual production	A great deal of content is being produced by individuals and stored in central services often owned by American corporate companies. It is unclear who has ultimate responsibility for archiving this content and introduces considerable legal issues. It could be argued that, in a sense, these services provide a kind of archive: certainly many people consider Flickr, for example to be their photograph 'repository'. As these services are owned by private companies there are questions that need to be asked about what would happen to these 'repositories' if the companies removed the service or changed it significantly[76]. As these services are owned by private companies the cardinality of the content is also subject to significant change or removal of the service.
2	Harness the power of the crowd	An archive might obtain/collect all the underlying data but not be able to reproduce the 'intelligence' that is created by the service, as this relies on proprietary algorithms for aggregating and processing the collective content—this is the service being provided, and it belongs to the company. For example, Cloudmark's Advanced Fingerprinting algorithms for automatically detecting email messaging threats.
3	Data on an epic scale	The scale of data being collected and aggregated into new services means that the process of collecting an archiving it will probably have

[74] See, for example, Google Maps data mash-up, www.housingmaps.com [last accessed 02/02/07].
[75] See, for example, Andy Powell's blog at http://efoundations.typepad.com/efoundations/2006/11/flash_is_the_ne.html [last accessed 02/02/07].
[76] Recently, for example, Google removed the SOAP interface to its Google Maps service

		to be automatic and will require huge processing and storage capacities[77]. It is also interesting to think about what can be done with this data as an aggregated whole. Google, for example, mines it to provide meta-information such as its 'zeitgeist' service – showing how the popularity of various search terms changes across time. This information is of cultural relevance and historians, in particular, will be interested in reviewing it.
4	Architecture of Participation	Services that get better the more people contribute to them will be difficult to capture in a way that recreates the full service at a later date. Often, the 'cool' factor, which is closely tied to the graphical look and feel and ease of use of a tool, is part of the mechanism for encouraging participation, and this is something that may be hard to capture in a repository.
5	Network Effects	Services that make use of the power of the network effect, for example, social networking sites, often combine data from a number of sources in a dynamic fashion and this is hard to recreate. In addition, the content has less meaning without the connectivity that is implied by the social links between users. The scale of the network effect throws into sharp relief the 'importance' and, arguably, the 'collectibility' of these types of Web 2.0 content: i.e. as indicators of types of social and cultural activity rather than as a collection of content.
6	Openness	Despite the underlying assumption that Web 2.0 makes increasing use of more open ways of working there are many complex legal issues emerging. Tim Bray, for example, argues that a service can not be considered open unless the user's data can be moved or taken back by the user at will, without the service provider withholding anything, encoding it in a proprietary format, or claiming any IPR. This is clearly not the case with many Web 2.0-based services, but adopting such a policy would make the job of collecting and archiving much easier. It would also alleviate the problem of how users could preserve their data in the case of a corporate service provider removing or significantly changing their service. However, the requirement for service providers to not withhold user data undermines the principle of data on an epic scale: the Web 2.0 business model depends on the idea of colossal amounts of data, held in hard-to-recreate databases, to create collective 'value' in its services. This is clearly directly in conflict with Bray's definition of openness.

Secondly, we can consider the common categories of Web 2.0 service and their particular implications for archiving (see table 3).

[77] A recent Wired magazine article highlighted the enormous hardware resources that Google requires in order to crawl and index the existing Web (Gilder, 2006).

Table 3: Web 2.0 services and characteristics with respect to archiving

Web 2.0 category of service	Some characteristics of concern with respect to archiving
Blogs	• Part of the topology of the Web and its rapid growth. However, as blogs frequently contain discussion about content elsewhere on the Web, commentary on linked objects that no longer exist or that are no longer identifiable creates a kind of 'link rot'. This hampers the integrity of the Web in general, but as links are an integral part of blogs, link rot can be expected to have an especially severe impact on blog archives (Entlich, 2004). • Time-sensitive content: bloggers are not usually concerned about persistence but we may consider it important to preserve as part of the cultural conversation and for the historical record. • Issues with scope—what is the complete blog? Does this include comments added by others, track-back links etc? • Are blogs part of the hidden Web? Blogs are hosted by a server system (the blog CMS tool) and actual content is usually held in a database. • The blog provides its own internal archive system through the blog software which should allow an accurate and full harvest by some form of crawl although in reality there are issues with permanence guarantees and user agreements (Entlich, 2004). • Different hosting services provide different functionality. • Blogs tend to be individual rather than organisational. This could be an issue for archivists keen to make sure that a preserved domain is representative[78].
Wiki	• Hidden Web: Actual content is held as text in a flat-file system or database and served by wiki script software (Ebersbach *et al.*, 2005). • Provides history function for versions of pages.
Media Sharing (YouTube etc.)	• Content is part of the hidden Web • Proprietary technology, therefore may be access/permission issues • Provides storage, so is it already a repository? Users often consider it so. • Individuals can create their own personal catalogues for trading and social networking.
Data mash-up	• Hidden Web: services like Google Maps use layered APIs which rest on large-scale database systems. If we move to personalised content feeds, who has responsibility for preserving this combination? In practical terms, Web, and particularly hidden Web, archiving relies on collaboration with the site's owner. • Look and feel/user experience is an integral part of service and is difficult to capture in an archive.
Podcasts	• Another example of a personal catalogue: widespread use of iTunes, with option to back-up to Apple's .mac repository. • Less of an issue with which version is being archived • Work has begun within HE to store educational podcasts (see: http://ed-cast.org)
Social tagging	• Users create their own collections of bookmarks etc. and share (i.e. a personal catalogue). • Layers of proprietary API and hidden Web
Social networks	• Look and feel/user experience is integral part of service • Creation of a personal 'space' – who is responsible for archiving? • Usually provided by corporate entities who possibly create own archives, which means there are potential IPR issues – who owns the content in your space?

[78] To date national archiving work in the UK has focused on devolving 'what' to archive to domain area specialists e.g. UK National Web Archive consortium member, the Wellcome Library, will focus on collecting medical sites.

5.4.2.2 Ingest

Just as most Web 2.0 users do not usually concern themselves with preserving their content, they do not usually concern themselves with ingest, either. However, there are certain implications for ingest that can be teased out from these discussions and some of these implications have particular relevance for HE. These have been compiled with the assistance of David Rosenthal[79].

Who will be responsible for ingest of content? Day (2003) proposes a range of organisations and discusses this in the context of public records and the role of national institutions and historical archives. There may be a particular problem for JISC-related Web projects as there is a reliance on archiving by host institutions where short-term funding means that staff turnover is high.

- How will we handle the boundaries between a student's Web 2.0 material and that of the institution? What about e-portfolios? Who is responsible, for example, for a student's Friend-Of-A-Friend (FOAF) record or their MySpace area? Can this be administered?

- Rosenthal *et al.* (2005) points out that normally, speed of ingest is not a factor in digital preservation. However, this may or may not be an issue for Web 2.0 material. For example, he notes that blog entries do tend to become fixed after a certain length of time (i.e. there is a point after which nobody adds any more comments to an 'old' blog entry). Experimental work has been undertaken at Stanford on archiving the Ariana Huffington political blog through the Cellar project within the LOCKSS initiative[80] and there are also plans to develop archiving software that will allow a feed directly into the Cellar store from a blog's RSS (rather than using Web crawling) which will make for easier ingest. However, will these conditions apply for other types of Web 2.0 content? In a dynamically changing Web 2.0 environment, the answer is probably not.

- There are no clear guidelines as to what kind of API is needed to deposit different kinds of resources into various types of repository. For example, Flickr and Fedora have published (proprietary) APIs that anyone can write to, however there are no comparable APIs for DSpace or ePrints, for example (Barker and Campbell, 2005). This is on-going issue for the JISC DRP Support Team[81].

All of these factors, when taken together, have one very obvious conclusion: that the characteristics of the Web and the way it has developed are not conducive to traditional collection and archiving methods and that this situation is unlikely to change. It therefore becomes necessary to think about how the traditional skills and expertise of professional library and information staff could be harnessed in order to rise to these challenges. However, it is not only in the area of skills where libraries and librarians are able to respond to the challenges of Web 2.0. They also have a long tradition of maintaining and developing a public service ethic that will become increasingly important in negotiating a Web 2.0 world where an individual's personal information, even identity, is in danger of becoming corporate property.

---

[79] The author is indebted to David Rosenthal (http://www.lockss.org/lockss/David_S.H._Rosenthal) for assistance with and discussion of this section.
[80] http://www.lockss.org/lockss/Home [last accessed 14/02/07].
[81] see: http://www.ukoln.ac.uk/repositories/digirep/index/Deposit_API [last accessed 14/02/07].

Further reading

Marieke Guy (2006) provides a discussion of some of the existing uses as part of a wider review of public sector wikis as well as concrete examples of wiki use. These include JISC's OSS Watch service, the DigiRepwiki (intended for all those working on the JISC Digital Repositories Programme) and, at Manchester University, the wikispectus, an alternative student prospectus.

Preservation of Digital Information: Issues and Current Status by Alison Bullock is a very readable introduction to some of the issues facing digital preservation in general. Available online at: http://epe.lac-bac.gc.ca/100/202/301/netnotes/netnotes-h/notes60.htm [last accessed 30/01/07].

Brian Kelly and the team at UKOLN have highlighted a number of HE/FE uses and examples, including the work at Warwick University on blogs[82] and the use of Google Map mash-ups for campus mapping[83] and for conference organisation and event planning (Kelly, 2006).

Jenny Levine and Michael Stephens have created a reading list on these issues for the Library 2.0 course for the American library Association at Squidoo: http://www.squidoo.com/library20 [last accessed 02/02/07].

Readers may also be interested in a forthcoming report (Spring 07) from OCLC concerning social software and its future role in libraries (Sharing, Privacy and Trust in the age of the network community) http://www.oclc.org/reports/privacyandtrust/default.htm [last accessed 02/02/07].

Blog- and wiki-based commentary on libraries and Web 2.0 [last accessed 14/02/07]:
http://www.ukoln.ac.uk/repositories/digirep/index/JISC_Digital_Repository_Wiki
http://ukwebfocus.wordpress.com
http://litablog.org/2006/10/31/web-20-becoming-library-20/
http://www.blyberg.net/2006/03/12/library-20-websites-where-to-begin/
http://litablog.org/2006/11/02/wikis-when-are-they-the-right-answer/

[82] http://blogs.warwick.ac.uk/ [last accessed 14/02/07].
[83] See, for example, Northumbria University's near-to-campus attractions map at: http://northumbria.ac.uk/browse/radius5/

6. Looking ahead - the Future of Web 2.0

Within 15 years the Web has grown from a group work tool for scientists at CERN into a global information space with more than a billion users. Currently, it is both returning to its roots as a read/write tool and also entering, through the power of the six big ideas, a new, more social, community and participatory phase. But where will it go next? Although Web 2.0 is barely off the ground, some are already beginning to ask: What will Web 3.0 look like?

Firstly, it is important to say a little about the overall direction of development. The large-scale collection of user data and creation of user generated content, aggregated by Web applications, will continue and no doubt deepen as people explore new ideas. The scale of this will grow through the network effect as more people come online and existing users increase their use of Web 2.0 services. Just how great this growth will be should be tempered by a consideration for what we have already learned about the topology of networks and the need for a less techno-centric view of the number of people who actually have the time and inclination to participate—witness the large number of blogs that are set-up and then abandoned[84]. The production processes to generate such online content will become more sophisticated with the advent of increasingly powerful and easy-to-use software (Cerf, 2007) and digital devices, and the use of mash-ups will grow.

This will, however, pose considerable problems for intellectual property protection and information overload may start to have a noticeable effect on many people. With so many different ways of accessing information (blogs, wikis, RSS feeds etc.) there may also be a sense in which people worry that they do not understand or use all of these forms and a sense of anxiety may even develop as to whether they are as fully connected as they should be.

A developing trend will be the growth of people's *personal catalogues*—digital collections of music, photographs, videos, lists of books, places visited etc. Some of the material will be self-generated, much of it will have been collected (either downloaded or linked to) from a growing range of services. It is likely that individuals will want to manipulate the content in these catalogues or archives, cutting, pasting, copying and editing within a personal digital space and potentially carrying out a process of 'innovation' (Borgman, 2003). Such collections will be considered manifestations of a person's persona and the contents will be shared and exchanged (Beagrie, 2005; Borgman, 2003). These collections will become extremely important to people, developing into a form of *personal archive* of a lifetime. They may well contain content from a person's educational experience and have direct links with Personal Learning Environments. Increasingly, as the amount of available online information grows and network effects increasingly take hold, a person's path through the information space will become profoundly important. This path might include a record of the *history* of interaction with information sources, the setting up and continual modification of personal filtering mechanisms, records of group interactions with an information source and the use of other people's filters and knowledge (the power of the crowd). This information path could be become part of our personal catalogue and used by others to make judgements about us—how credit-worthy we are, for example. Careful readers of Tim Berners-Lee's blog may have spotted an oblique reference to Garlik, a monitoring service that tracks subscribers' online personal information to help identify potential security threats.

Alongside this trend of information paths, 'digital objects' such as Word documents or personal photographs, may themselves become 'history-enriched' (Morville, 2006, p. 150) with the digital equivalents of the properties that physical artefacts like books gather through time e.g. scrawled marginalia, becoming dog-eared etc. Although Morville's discussion is

[84] Mann, 2006 indicated that more than 10 million of the 12.9 million blogs profiled on the Blogger service were inactive.

conceptual rather than practical, it is certainly possible to see how Web 2.0 services will try to facilitate these types of activities as their business models will depend on this kind of information to fuel their services.

The Web, or more precisely the network, as platform and the idea of software above the level of a single device is becoming firmly entrenched as a concept and it is likely that over the next few years we will start to perceive personal computing more as a process of interacting with networked services rather than using a particular computing device. This trend can only be exacerbated by the move towards ubiquitous computing.

Finally, in general terms we may also begin to see a change in the way in which we interact with other people: what Nigel Shadbolt refers to as 'the fabric of people being connected' through these new technologies and the formation of new social communities in which we share information and carry out collective endeavours (Shadbolt, 2006). The social aspects of the Web's topological interconnectedness are becoming increasingly important and indeed this may be the most important long-term trend. As one example, a survey by Oxford University's Internet Institute, as long ago as 2005, found that one in five people in the survey had met a new person or made friends online (Dutton *et al*., 2005).

6.1 Web 2.0 and Semantic Web

At the beginning of this report I discussed the two Tims (Tim Berners-Lee and Tim O'Reilly) as a way in to understanding the difference between Web technologies and Web 2.0 ideas. There is, however, another contentious issue for the future development of the Web: the relationship between Web 2.0 ideas and the Semantic Web.

In the original exposition of the idea of the Semantic Web for an article in Scientific American, Tim Berners-Lee's vision included scenarios in which autonomous agents and machine processing units will carry out actions on our behalf[85]. There is still some confusion over what precisely the Semantic Web really is and where it is heading, not least from business and commerce. For Tim Berners-Lee it is in essence about the shift from documents to data—the transformation of a space consisting largely of human-readable, text-oriented documents, to an information space in which machine-readable data, imbued with some sense of 'meaning', is being exchanged and acted upon. However, to date, even its proponents argue that this vision is largely unrealised (Shadbolt *et al*., 2006) although technologies and applications are now beginning to appear, as opposed to just being researched.

There is a potential split between the Web 2.0, social software enthusiasts, and proponents of the Semantic Web (Morville, 2006). As we have seen in our discussion of folksonomies (see section 3.2.3) there has been considerable and at times heated debate between those who favour the formality of controlled vocabularies and ontologies and those who prefer the more informal nature of social tagging. An issue that has dogged the development of the Semantic Web is the need to develop ontologies for a multitude of domains, which could have considerable resource costs. Some would like to see the role of folksonomies and collabularies informing this debate and the idea of the social context in which ontologies operate is being discussed (Mika, 2006). Morville argues that these communities need to work together more closely, perhaps in a layered approach. Indeed, Mika argues that although the Semantic Web is envisioned as a machine-to-machine system, the process of creating and maintaining it is a *social* one, acting within a social context, particularly with regard to the creation of ontologies. For example, Nickles (2006) argues for formal inclusion of information about social attitudes ('sociality') and controversial opinions within the Web in

[85] See the JISC TechWatch report: Matthews, 2005. Semantic Web Technologies (TSW0502).

order to help its development. Such work builds on Seely Brown and Duiguid's previous discussion of the social life of information (2000).

As part of this process there are several areas where developments in Semantic Web and those within social software are beginning to be explored in consort:

Semantic Wikis

This is a developing research area, but in essence, researchers are looking at ways to annotate wiki content with semantic information[86]. A Semantic wiki allows users to make formal descriptions of things in a manner similar to Wikipedia, and also annotate these pages with semantic information using formal languages such as RDF and OWL (Oren *et al*., 2006). A number of engines are being developed to support this concept including Platypus and SemperWiki[87]. An alternative, OntoWiki[88], harnesses the architecture of participation to allow users to work collaboratively on information maps (Auer *et al*., 2006).

Semantic Blogging

Blogs can be more than an easy-to-use publishing tool. Their ability to also generate machine-readable RSS and Atom feeds means that they can also be used to distribute machine-readable summaries of their content and thus facilitate the aggregation of similar information from a number of sources (Cayzer, 2004). Traditionally, these feeds are used for the headlines from blog postings, but by combining the ideas behind the Semantic Web with blogging software – Semantic Blogging – it may be possible to develop new information management systems[89]. For example, RDF semantic data can be used to represent and export blog metadata, which can then be processed by another machine. In the long run the inclusion of this semantic information, by instilling some level of meaning, will allow queries such as 'Who in the blogosphere agrees/disagrees with me on this point?'

Semantic Desktop

It is envisaged that combining the ideas of the Semantic Web and Web 2.0 services with traditional desktop applications and the data they hold (such as word processor files, emails and photos) on your local computing device will facilitate a more personalised way of working. In theory, this should create a more focused information and knowledge management environment, helping to find a way through personal 'data swamps'[90]. Research work is at an early stage, but IBM is working on QEDWiki, a wiki-based application framework for collaboration working which enables the creation of enterprise mash-ups[91].

Working with ontologies and folksonomies

There are several people working in this area: Patrick Schmitz has presented research into a model that works with both folksonomies and ontologies by leveraging statistical natural language processing. His goal is to develop a system that retains the flexibility of free tagging for annotation but make uses of ontology in the search and browse interface (Schmitz, 2006). Another proposal, from Dave Beckett (2006), is to make more use of the social context within

[86] http://www.semwiki.org/
[87] see: SemperWiki http://www.semperwiki.org/ and http://platypuswiki.sourceforge.net/whatis/index.html [last accessed 14/02/07].
[88] A prototype can be viewed at: http://3ba.se [last accessed 14/02/07].
[89] More details at: http://www.semanticblogging.org/semblog/whatisit.html [last accessed 14/02/07].
[90] Further notes on the idea of the semantic desktop can be found at: http://www.semanticdesktop.org/xwiki/bin/view/Main/ and http://www.gnowsis.org/ [last accessed 28/01/07].
[91] A short video from IBM showing their vision of using mash-up ideas can be seen at: http://www.youtube.com/watch?v=ckGfhlZW0BY [last accessed 28/01/07].

which tags are created[92] by separating the tool that creates the tags from the tool with which they are used. He also proposes that wiki pages should be created for individual tags which users could then add to/edit so that the wiki page, in effect, becomes the tag. The on-going process of refinement for each separate tag would form a kind of consensus as to the meaning of that tag and would also record the processes (the semantic path) by which the end result is being reached. This would, to take just one simple example, allow direct links to other language versions of the same tag.

In terms of bookmarking services such as Del.icio.us and the open source SiteBar (www.sitebar.org), one of the key problems is how best to classify the growing list of URLs. At the WWW2006 conference in Edinburgh, Dominic Benz *et al*, from the University of Freiburg, put forward an idea for automatically classifying bookmarks. The authors proposed an automated system which takes account of how the user has classified bookmarks in the past and how other people with similar interests have also classified their bookmarks. In other words find a similar user who has already classified and stored a bookmark and derive a recommendation based on what they did[93].

6.2 The emerging field of Web Science

Web science is an emerging discipline, recently proposed by Tim Berners-Lee and his colleagues at the University of Southampton and MIT. Its goal is to understand the growth of the Web, its emerging topology, trends and patterns and to develop new scientific approaches to studying it (Berners-Lee *et al.,* 2006). Increasingly, given the importance of the Web as a social tool, there will be more research into the social and legal relationships behind information.

6.3 The continued development of the Web as platform

Computing software architecture tends to go in phases, paradigms even, and the Web or network as platform is one such paradigm. In coming years an increasing number of tools and operating system-like software will emerge to further this process. An example of this is Parakey[94], which is currently being developed by the co-founder of the Mozilla Firefox project, Blake Ross (Kushner, 2006). It will provide a browser-based way to access and manipulate the contents of your desktop PC and also allow others, with your permission, to do the same. In effect, it provides software that essentially turns your computer into a local server.

6.4 Trust, privacy, security and social networks

A great deal of discussion is taking place around provenance, reputation, privacy and security of Web and email data. The sheer scale of material that people are prepared to post, often the most intimate details and photos that a generation ago would only have been seen and known by a handful of friends is changing the nature of privacy (George, 2006). There is also a growing awareness that as the volume of information available from the Web grows, the ability to determine what is accurate and from a trusted source becomes ever more difficult. Increasingly, there is concern about some of the more dubious aspects of search engine optimisation (in which search engines are manipulated so that certain websites appear higher in the rankings), weblink spam (groups of pages that are linked together with the sole purpose

[92] Seely Brown's book The Social Life of Information makes a powerful case for taking account and care of the social context in which information exists

[93] See: http://www.informatik.uni-freiburg.de/cgnm/software/caribo/index_en.html [last accessed 01/02/07].

[94] http://www.parakey.com

of obtaining an undeservedly high score in search engine rankings) (Mann, 2006) and the potential for Semantic Web spam, in which deliberately falsified information is published. It is no coincidence that trust is at the highest levels of the Semantic Web 'layer cake' model (see Matthews, 2005).

There are large numbers of spam and email filters on the market and despite best efforts they are still not regarded as fully adequate. Brondsema and Schamp (2006) argue that such filters should make more use of trust ratings determined from social networks and their Konfidi system[95] attempts to do this. Another proposal, from Jean Camp (Indiana University) is that computer trust models should be more grounded in human behaviour and take account of work in the social sciences in this regard (for example game theory). Her Net Trust system[96] uses social networks to re-embed social information online. A tool bar inserted into the Web browser provides information on the trustworthiness of the website being viewed based on knowledge and ratings obtained both from a social network of friends and colleagues and trusted third parties (such as Consumer Unions and PayPal).

6.5 Web 2.0 and SOA

Service-Oriented Architecture (SOA) is an architectural approach in which highly independent, loosely-coupled, component-based software services are made interoperable, and there is now some discussion around a potential synergy between Web technologies and SOA.

In particular, some argue that bringing together the rich front-end user experience provided by the latest Web technologies such as RIA with SOA-enabled technologies at the back end could provide improved reliability, better scalability, and better governance (Snyder, 2006). Both have openness, data re-use and interoperability at their core. In fact, Web 2.0 data mash-ups could be considered similar to the composite applications of SOA (see diagram below). There are, of course, differences: SOA relies heavily on governance, which Web 2.0 lacks, and on a technical level there is an issue with the on-going SOAP versus REST debate, since SOA implementations make greater use of SOAP and WS-*)[97].

[95] http://konfidi.org/ [last accessed 12/02/07].
[96] See: http://www.ljean.com/netTrust.html for information and pictures [last accessed 12/02/07].
[97] For more on this emerging discussion see:
http://web2.wsj2.com/continuing_an_industry_discussion_the_coevolution_of_soa_and.htm
http://blogs.zdnet.com/Hinchcliffe/?p=72
http://www.soaeosconference.sys-con.com/read/174718.htm

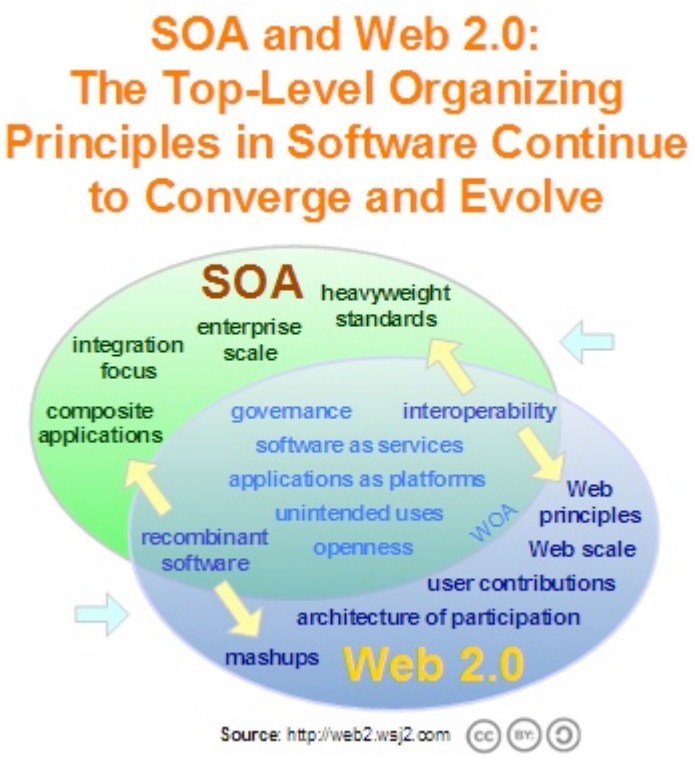

Figure 2: Web 2.0 and SOA

6.6 Technology 'Bubble 2.0'?

"I am not so sure that we're not seeing another bubble"

Howard Rheingold[98]

"When people say to me it's a Web 2.0 application, I want to puke"

Guy Kawasaki, venture capitalist, in Levy and Stone, 2006, p. 5.

When a respected future watcher like Howard Rheingold worries about whether we are witnessing another technology bubble and potential pop it is worth taking note. Indeed, no explanation as to what the Web 2.0 moniker means would be complete without some reference to the surge of investment interest in a new generation of dot-com entrepreneurs and young start-up companies with ideas for social software (Boutin, 2006). Stabilo Boss has prepared an image which shows the large number of brands in the already-saturated world of social software companies and tools[99].

Does this matter to education? The answer is yes, if too much time, resources and data are invested in new and untested applications which are not subsequently supported adequately or are backed by companies that eventually fail. A great many of the new applications are not open source, but small start-ups seeking corporate backing and this means there are justifiable concerns over their sustainability.

[98] text transcription of video conference conversation. Available online at: http://www.masternewmedia.org/news/2006/10/04/web_20_meets_smartmobs_howard.htm [last accessed 21/02/07].

[99] http://flickr.com/photos/stabilo-boss/93136022/in/set-72057594060779001/ [last accessed 19/02/07].

6.7 And Web 3.0?

At the WWW2006 conference in Edinburgh, when asked by TechWatch about the likely characteristics of 'Web 3.0', Tim Berners-Lee stated that he believes that the next steps are likely to involve the integration of high-powered graphics (Scalable Vector Graphics, or SVG) and that underlying these graphics will be semantic data, obtained from the RDF Web, that 'huge data space'. A focus on visualisation is also evident elsewhere: Ted Nelson, the inventor of hypertext, is working on FloatingWorld: a system for displaying documents, including the links between them, in three dimensions. He recently spoke of the idea of translating this concept to a 3-dimensional social networking system. In addition, IBM recently announced the winning ideas in an international search for technology developments that it would fund to the tune of $100Million over the next couple of years[100]. One of the winners was the '3D Internet' which will take the best of virtual worlds such as Second Life and gaming environments, and merge them with the Web.

However, it could be argued that this, once again, is focusing on Web technologies and not looking at the big ideas. For this we should maybe go back to the fundamental idea of the topology of the Web and take a look at what kind of a legacy Web 2.0 may have left us with. If some of the more negative effects of Web 2.0 have taken hold to a demonstrably detrimental effect, it is quite possible to envisage a situation where 'Web 3.0' would become a backlash to Web 2.0: where software that 'cleans up' after you, erasing your digital path through the information space, and identity management services, are at a premium. Where you sell your valuable attention span in blocks of anything from minutes to several hours rather than giving it away for free. Services such as Garlik and AttentionTrust[101] are the first green shoots of these developments – as much essential protections as opportunities to capitalise on the value of your attention and your trust.

[100] see: http://www-03.ibm.com/press/us/en/pressrelease/20605.wss [last accessed 14/02/07].
[101] http://www.attentiontrust.org/ [last accessed 14/02/07].

Conclusion

This report has covered a lot of ground. It has looked at Web 2.0, tried to separate out some of the sense from the sensational, reviewed the technologies involved and highlighted some of the issues and challenges that this poses to higher education in the UK (see appendix A for a summary of these and some tentative recommendations). This is a complex and rapidly evolving area and this report can, perhaps inevitably, seem to raise as many questions as it answers.

I believe, however, that there are a few core points that we should hold on to when thinking about Web 2.0 and how it might impact on education: firstly, that Web 2.0 is more than a set of 'cool' and new technologies and services, important though some of these are. It is actually a series of at least six powerful ideas or drivers that are changing the way some people interact. Secondly, it is also important to acknowledge that these ideas are not necessarily the preserve of 'Web 2.0', but are, in fact, direct or indirect reflections of the power of the network: the strange effects and topologies at the micro and macro level that a billion Internet users produce. This might well be why Sir Tim Berners-Lee maintains that Web 2.0 is really just an extension of the original ideals of the Web which does not warrant a special moniker; but the fact that business concerns are starting to shape the way in which we are being led to think and potentially act on it means that we need to at least be more aware of these influences. For example, many of the Web 2.0 services are provided by private, often American companies. Start-up companies tend to either fail or be bought out by one of a triumvirate of corporates: Google, Yahoo and Microsoft. This raises questions about the ownership of the user data collected. The UK HE sector should debate whether this is a long-term issue. Maybe delineating Web from Web 2.0 will help us to do that.

Finally, it is important to look at the implications of Web 2.0. The changes that are taking place are likely, I think, to provide three significant challenges for education: Firstly, the crowd, and its power, will become more important as the Web facilitates new communities and groups. A corollary to this is that online identity and privacy will become a source of tension. Secondly, the growth in user or self-generated content, the rise of the amateur and a culture of DIY will challenge conventional thinking on who exactly does things, who has knowledge, what it means to have élites, status and hierarchy. These challenges may not be as profound as some of the more ardent proponents of Web 2.0 indicate, but there will be serious challenges none the less (ask any academic for his/her views on Wikipedia as a research tool). And finally, there are profound intellectual property debates ahead as individuals, the public realm and corporations clash over ownership of the huge amounts of data that Web 2.0 is generating and the new ways of aggregating and processing it.

About the Author

Paul Anderson is a technology writer and Technical Director of Intelligent Content Ltd. He is currently acting as Technical Editor of JISC TechWatch. A graduate of Leeds University (computer science), he has over twenty years' experience of working in computing in both academia and industry. He has undertaken research into AI-based recognition systems and worked in industrial software development, project management and technology transfer. Paul is a member of the NUJ, British Computer Society, Association of British Science Writers and an associate member of both the ACM and IEEE.

During his lunch break he blogs about technology at http://techlun.ch and can be contacted by email at: p.anderson [AT] intelligentcontent.co.uk.

Appendix A

One of the purposes of this JISC TechWatch report was to stimulate debate within the HE/FE community on the challenges posed by the development of Web 2.0. I conclude this report, therefore, will some debating points and recommendations. At the ALT-C conference in September 2006, conference attendees were asked their thoughts and ideas about Web 2.0 and this section includes some of that feedback as well as learning points gleaned from elsewhere in the report.

Educational Recommendations

- The education community needs to reflect further on the implications for institutional VLEs. The integration of VLEs and Web 2.0 technologies might make use of their combined strengths and further exploration of how this might be achieved and the implications of doing so, should take place, if it isn't already. How to utilise the visual power of Web 2.0 services should be an especial consideration.

- Assessment and grading in a Web 2.0 world, in which collaboration, knowledge sharing and more constructivist approaches are more common, will need further review. Is, for example, a data mash-up created by a student in some ways equivalent to an essay? Web 2.0 will pose new challenges to the issue of plagiarism and these need to be explored.

- We need to further explore, research and analyse the uses, benefits and limitations of Web 2.0 learning solutions (see, for example, the discussion in Boulos *et al.,* 2006). Do we know enough about the ways in which young people and students are currently using blogs and other tools?[102]. There is a role for JISC to facilitate and fund demonstrators for these types of services in academic settings, in line with the recent call for projects under the Users and Innovation programme[103].

- Further work is required on understanding the pedagogy implications of these services. This will include the need to explore further the social aspects of the learning (Kukulska-Hulme, 2006) that takes place and the many issues concerning participation. We cannot, for example, assume everyone is happy working in the 'self-publish' mode.

Libraries

- Libraries have skilled staff with professional expertise that can be leveraged to rise to the challenge of Web 2.0, not only in collection and preservation, but also in user-centred services. They are also the guardians of a long tradition of a public service ethic which will increasingly be needed to deal with the privacy and legal issues raised by Web 2.0. Library staff should be encouraged to think and act pro-actively about how they can bring this to bear on the development of new, library and information service-based technologies.

- Should libraries take a lead in the introduction of such technologies into the learning and academic workplace, driving the collaboration between academics, administrators and central information services? A recent article in Health Information Library proposed a kind of informal technology lab or test-bed to allow HE experimentation with Web 2.0 services and technologies (Whitsed, 2006). This

[102] For example, Jin Tan, a PHD student at University of Sheffield is undertaking research in this area
[103] http://www.jisc.ac.uk/events/2006/10/event_capital_1006.aspx

proposal should be considered, with a view perhaps to being hosted within collaborating groups of libraries, possibly on a regional basis.

Research

- There seems to be more scope for the use of blogs and wikis in research-based peer-to-peer communication and experimentation but there are questions as to why this is not happening as much as it might. Are there justifiable concerns that this may be being held back by institutional and managerial issues? How engaged are Information Services departments with these new technologies? A review of the current situation with regard to use by researchers of blogs, wikis and other Web 2.0 services and a way forward should be commissioned.

- All the leading open API data mash-ups use corporate data taken from Google, Yahoo etc. Where are the leading examples from education and the public sector? We should actively encourage the development of prototype research data mash-ups, that harness the power of sophisticated visual interfaces, to show the power of this technique.

Technical

- Further research is required into whether institutions should try and utilise the services that power existing social software or find ways to incorporate them into existing IS systems[104] Should we be creating new, potentially even better services that build on the ideas behind existing software? How will we respond to the need to develop compelling user interfaces?

General, administration and Third Stream

- The education community should worry that much of Web 2.0 data is 'hosted externally to academia' (Alexander, p. 42). JISC should take a position on the right to extract a user's data from Web 2.0 services.

- Web 2.0 development is rapid. This poses a problem for those in education who are trying to keep a handle on all these. There are also risks associated with using services that are in 'perpetual beta' and very fluid (for example, Google recently withdrew a SOAP interface to its map service). JISC should consider an online resource for keeping track of emerging new services and tools and their APIs/interfaces. Perhaps this could be in the form of wiki which anyone in the JISC community could contribute to?

- There are profound IPR issues. Do students (even staff) understand that simply 'copying and pasting', uploading commercial video, copying photos etc is not always a legal activity? What are the commercialisation issues with regard to 'free the data', who 'owns' a student group coursework mash-up or a PhD student's peer-contributed experimental data that both sit on a Californian server farm? These important questions need to be formally reviewed and commercialisation staff within university administration departments should be made more aware of these difficulties.

- Staff involved in PR, marketing and the promotion of universities and colleges should be aware of the development of blogging and the blog-based PR tactics that are being adopted by corporate entities, and should try to learn from them.

[104] Almost a quarter of the EU Internet population use such at least site once a month (Guardian, 29th Nov 2006, page 26)

- There are legal implications for student and staff blogging. Is this a form of journalism and therefore subject to the same laws (e.g. libel)? There should be a review of the legal issues at play in this area and the corresponding implications for university and college administrators.

Points for further debate

- Are children as digitally 'native' as we think? It may be necessary to review the skills, attitudes even, that are needed in the new world of Web 2.0. There are information literacy issues and we need to education children and students in how to make best use of these new, collaborative technologies (Boulos, 2006).

- We will need to educate young people more deeply about privacy, trust and the social Web. Those who participate often don't seem to appreciate that the reach of the network means that their profile could potentially be viewed by millions of people and that there could be long-term implications to this (George, 2006). As one example, in autumn 2006, the University of California required students to attend classes in social networking.

- Should libraries become more involved in the *production* of content in a user-generated world? Should they provide the digital and even physical space for this activity? E.g. podcast recording facilities. Is there a role for the libraries in training people in the use of these new technologies and services, facilitating use and encouraging good (and ethical) practice (Hepworth, 2007).

- How could libraries utilise their expertise in niches to take advantage of the 'long tail' effect?

- Are there ways to integrate Web 2.0 services and technologies with more traditional information retrieval technologies such as online databases, gateways and portals to help facilitate research?

- Are there lessons for the UK HE software development community concerning the style and ethos of the development of Web 2.0? For example, the notion of 'always beta'; lightweight programming methods etc.

- What are the challenges and issues with regard to user identity on the network e.g. Federated ID, SxiP, SAML, Identity 2.0?

- How does Web 2.0 connect technically with the developing agenda of m-learning, mobile devices and ubiquitous computing?

- Are there institutional barriers to the adoption of Web 2.0 services? This is an important question as, until it is resolved, it means it is currently difficult to understand the implications of the seemingly low uptake of social software technologies within HE.

- Is there an innovation chasm with regard to the uptake of these technologies within the education community? Has it only been 'early adopters' so far? Do we know what percentage of online users actually engage with and use tools such as blogs and wikis. Should we undertake research into who is using these systems in HE/FE?

REFERENCES

AL-KHALIFA, H. S., DAVIS, H. C. 2006. *Harnessing the wisdom of crowds: how to semantically annotate Web resource using folksonomies*. In: Proceedings of IADIS Web Applications and Research 2006 (WAR2006). Available online at: http://eprints.ecs.soton.ac.uk/13158/ [last accessed 14/02/07].

ALEXANDER, B. 2006. *Web 2.0: A new wave of innovation for teaching and learning*. **EDUCAUSE Review**. Vol. 41, No. 2, March/April 2006, pp. 32–44. EDUCAUSE: Boulder, USA. Updated version available online at: http://www.educause.edu/apps/er/erm06/erm0621.asp [last accessed 14/01/07].

AMSEN, E. 2006. *Who Benefits from Science Blogging?* **Hypothesis Journal.** Vol. 4, No. 2. University of Toronto. Available online at: http://medbiograd.sa.utoronto.ca/pdfs/vol4num2/10.pdf [last accessed 21/02/07].

ANDERSON, C. 2006. **The Long Tail: How endless choice is creating unlimited demand.** Random House Business Books: London, UK.

ASHLIN, A., LADLE, R. 2006. *Environmental Science Adrift in the Blogosphere*. **Science**. April 14, 2006: Vol. 312. No. 5771, p. 201. Requires login: http://www.sciencemag.org/cgi/content/summary/312/5771/201 [last accessed 14/01/07].

AUER, S., Dietzold, S., Riechert, T. 2006. *OntoWiki – a tool for social, semantic collaboration.* **The 5th International Semantic Web Conference**, Athens, GA, USA, November 5-9, 2006, LNCS 4273. http://iswc2006.semanticweb.org/items/in_use_5.php [last accessed 14/01/07].

AULETTA, K. 2001. **World War 3.0: Microsoft and its enemies**. Profile Books: London, England.

von BAEYER, H. C. 2003. **Information: The New Language of Science**. Weidenfeld & Nicolson: London.

BARKER, P., CAMPBELL, L. 2005. *The eFramework Priorities and Challenges for 2006: Repositories Theme Strand*. **Report from the JISC CETIS Conference 2005**, Edinburgh. Available online at: http://www.e-framework.org/Default.aspx?tabid=753 [last accessed 14/01/07].

BEAGRIE, N. 2005. *Plenty of room at the bottom? Personal digital libraries and collections*. **D-Lib magazine**. Iss. 11, No. 6 (June 2005). http://www.dlib.org/dlib/june05/beagrie/06beagrie.html [last accessed 12/02/07].

BECKETT, D. 2006. *Semantics Through the Tag.* **XTech 2006: Building Web 2.0**, 16-19 May 2006, Amsterdam, Netherlands. Available at: http://xtech06.usefulinc.com/schedule/paper/135 [last accessed 12/02/07].

BENKLER, Y. 2006. **The Wealth of Networks: how social production transforms markets and freedom**. Yale University Press: USA.

BENZ, D., TZO, K., SCHMIDT-THIEME, L. 2006. Automatic bookmark classification: a collaborative approach. **WWW2006 Conference**, May 22–26, 2006, Edinburgh, UK. Available online at: http://www.wmin.ac.uk/~courtes/iwi2006/benz_automatic.pdf [last accessed 15/01/07].

BERESFORD, P. 2007. *Web Curator Tool*. **Ariadne**. Iss. 50 (Jan 2007). Available online at: http://www.ariadne.ac.uk/issue50/beresford/ [last accessed 12/02/07].

BERNERS-LEE, T. 1999. **Weaving the Web.** Orion Business Books.

BERNERS-LEE, T., HALL, W., HENDLER, J., SHADBOLT, N., WEITZNER, D. 2006. *Creating a science of the Web*. **Science.** Aug 11, 2006. Vol. 313, No. 5788 pp.769-771.

BERUBE, L. 2007. *On the Road Again: The next e-innovations for public libraries?* Available at: http://www.bl.uk/about/cooperation/pdf/einnovations.pdf [last accessed 12/02/07].

BORGMAN, C. 2003. *Personal digital libraries: creating individual spaces for innovation.* **NSF/JISC Post Digital Library Futures Workshop**. June 15-17, 2003, Cape Cod, Massachusetts. http://www.sis.pitt.edu/~dlwkshop/paper_borgman.html [last accessed 12/02/07].

BOULOS, M., MARAMBA, I., WHEELER, S., *Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education*. **BMC Medical Education**. 15th August 2006, 6:41. Available online at: http://www.biomedcentral.com/1472-6920/6/41 [last accessed 12/02/07].

BOUTIN, P. 2006. *Web 2.0: the new Internet 'boom' doesn't live up to its name*. **Slate** (online). March 29th 2006. Available online at: http://www.slate.com/id/2138951/ [last accessed 14/02/07].

BRISCOE, B., ODLYZKO, A., Tilly, B. 2006. *Metcalfe's Law is wrong*. **IEEE Spectrum**. July 2006. Available online at: http://spectrum.ieee.org/jul06/4109 [last accessed 14/02/07].

BRITTAIN, S., GLOWACKI, P., VAN ITTERSUM, J., JOHNSON, L. 2006. *Podcasting Lectures*. **Educause Quarterly**, Vol. 29, No. 3. EDUCAUSE: Boulder, USA. Available online at: http://www.educause.edu/apps/eq/eqm06/eqm0634.asp [last accessed 15/01/07].

BRONDSEMA, D., SCHAMP, A. 2006. *Konfidi: trust networks using PGP and RDF*. Models of trust of the Web (MTW 06). **WWW2006 Conference**, May 22–26, 2006, Edinburgh, UK. Available online at: http://www.ra.ethz.ch/CDstore/www2006/www.l3s.de/~olmedilla/events/MTW06_papers/paper04.pdf [last accessed 15/01/07].

BROWN, John Seely, DUGUID P. 2000. **The Social Life of Information**. Harvard Business School Press: USA.

BUNEMAN, P., KHANNA, S., TAJIMA, K., TAN, W. 2004. *Archiving Scientific Data*. **ACM Transactions on Database Systems**, 27(1) pp.2–42.

BUTLER, D. 2005. *Science in the web age: Joint efforts*. **Nature**. Nature 438 (1 December 2005), pp. 548-549.

BUTLER, D. 2006. *The scientific Web as Tim originally envisaged.* Tutorial session on Web 2.0 in Science. **Bio-IT world Conference**. March 14, 2006**.** Available online at: http://www.blogs.nature.com/wp/nascent/DeclanButler_BioITWeb2.ppt [last accessed 12/02/07].

CASTELLS, M. 2000. *The Rise of the Network Society*. Volume 1 of **The Information Age: Economy, Society and Culture.** Blackwell Publishing.

CAYZER, S. 2004. *Semantic Blogging and Decentralized knowledge Management.* **Communications of the ACM**. Vol. 47, No. 12, Dec 2004, pp. 47-52. ACM Press.

COSTELLO R., KEHOE. T. 2005. *Five minute intro to REST*. **xFront.com**. PowerPoint presentation available at: http://www.xfront.com/5-minute-intro-to-REST.ppt [last accessed 14/02/07].

CERF, V. 2007 *An Information Avalanche*. **IEEE Computer**. Vol, 40, No. 1 (Jan 2007).

CRAWFORD, W. 2006. *Library 2.0 and "Library 2.0"*. **Cites & Insights**. Vol. 6, No. 2 (Midwinter 2006). Available at: http://cites.boisestate.edu/civ6i2.pdf [last accessed 14/02/07].

CYCH, L. 2006. *Social Networks*. In: **Emerging Technologies for Education**, BECTA (ed.). Becta ICT Research: Coventry, UK.

DAY, M. 2003. Collecting and Preserving the World Wide Web. Version 1.0, 25th Feb, 2003. JISC: Bristol, UK. Available online at: http://www.jisc.ac.uk/uploaded_documents/archiving_feasibility.pdf [last accessed 14/02/07].

DEMPSEY, L. 2006. *Libraries and the Long Tail: Some Thoughts about Libraries in a Network Age.* **D-Lib Magazine**. Vol. 12, No. 4, April 2006. Available online at: http://www.dlib.org/dlib/april06/dempsey/04dempsey.html [last accessed 14/02/07].

DOCTOROW, C., DORNFEST, F., JOHNSON, J. Scott, POWERS, S. 2002. **Essential Blogging**. O'Reilly.

DOWNES, S. 2004. *Educational Blogging.* **EduCause Review**. Vol. 39, no. 5, Sept/Oct 2004, pp. 14–26. Also available online at: http://www.educause.edu/pub/er/erm04/erm0450.asp [last accessed 14/02/07].

DUTTON, W. H., di GENNARO, C., MILLWOOD HARGRAVE, A. 2005. **Oxford Internet Report: The Internet in Britain.** Oxford Internet Internet (OxIS). May 2005.

EBERSBACH, A., GLASER, M., HEIGL, R. 2006. **Wiki: Web Collaboration**. Springer-Verlag: Germany.

ENTLICH, R. 2004. *Blog Today, Gone Tomorrow? Preservation of Weblogs*. **RLG DigiNews** (online). Vol. 8, No. 4 (August 2004). Available at: http://www.rlg.org/en/page.php?Page_ID=19481 [last accessed 12/02/07].

FARREL, J., KLEMPERER, P. 2006. *Coordination and Lock-In: Competition with Switching Costs and Network Effects*. **Working Paper series**. Social Science Research Network. May 2006. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=917785 [last accessed 12/02/07].

FELIX, L., STOLARZ, D. 2006. **Hands-On Guide to Video Blogging and Podcasting: Emerging Media Tools for Business Communication.** Focal Press: Massachusetts, USA.

FOUNTAIN, R. 2005. *Wiki Pedagogy*. **Dossiers Pratiques**. Profetic. Available at: http://www.profetic.org:16080/dossiers/dossier_imprimer.php3?id_rubrique=110 [last accessed 12/02/07].

FREY, J. G. 2006. *Free The Data.* **WWW 2006 Panel Discussion**, Edinburgh, UK, March 25, 2006. Available online at: http://eprints.soton.ac.uk/38009/ [last accessed 12/02/07].

GARRETT, J. 2005. *Ajax: A New Approach to Web Applications.* **Adaptive Path website**, Feb 18th. Available at: http://www.adaptivepath.com/publications/essays/archives/000385.php [last accessed 12/02/07].

GEORGE, A. 2006. *Things you wouldn't tell your mother*. **New Scientist**. Sept. 16th 2006, pp. 50-51.

GILLMOR, D. 2004. **We the media.** O'Reilly.

GILDER, G. 2006. *The Information Factories*. **Wired**. 14.10 (October 2006). Available online at: http://www.wired.com/wired/archive/14.10/cloudware.html [last accessed 12/02/07].

GLOGOFF, S. 2006. *The LTC wiki: experiences with integrating a wiki in instruction*. IN: Mader, Stewart L. (ed.) 2006. **Using Wiki in Education**. Available online at: http://wikiineducation.com/display/ikiw/The+LTC+Wiki+-+Experiences+with+Integrating+a+Wiki+in+Instruction [last accessed 14/02/07].

GUDIVA, V. RAGHAVAN, V., GROSKY, V., KASANAGOTTU, R.1997. *Information Retrieval on the World Wide Web.* **IEEE Internet Computing**. Vol. 1, No. 5, pp. 58-68.

GUY, M. 2006. *Wiki or Won't He? A Tale of Public Sector Wikis.* **Ariadne.** Issue 49. http://www.ariadne.ac.uk/issue49/guy/ [last accessed 14/02/07].

HERRY, R., POWELL, A. 2006. **Digital Repositories Roadmap: looking forward**. UKOLN. Available online at: http://www.ukoln.ac.uk/repositories/publications/roadmap-200604/rep-roadmap-v15.pdf [last accessed 12/02/07].

HEPWORTH, M. 2007. Private e-mail conversation. Jan 2007.

HERTZFELD, A. 2005. **Revolution in the valley: The insanely great story of How the Mac was made.** O'Reilly.

HINCHLIFFE, D. 2006. *The coming RIA wars: a roundup of the Web's new face*. **Enterprise Web 2.0** (blog), Sept. 11th, 2006. ZDNet.com. Available online at: http://blogs.zdnet.com/Hinchcliffe/?p=65 [last accessed 14/02/07].

JOHNSON, D. 2005. AJAX*: Dawn of a new developer: The latest tools and technologies for AJAX developers*. **JavaWorld.com**, Oct 17th 2005. Available online at: http://www.javaworld.com/javaworld/jw-10-2005/jw-1017-ajax.html [last accessed 14/02/07].

KELLY, B. 2002. *Archiving the UK domain and UK websites*. **Proceedings of Web-archiving: managing and archiving online documents and records**. London, March 25, 2002. Available online at: http://www.dpconline.org/graphics/events/webforum.html [last accessed 14/02/07].

Kelly, B. 2006. *Web 2.0: Opportunities and challenges.* **EMUIT meeting**, Nottingham Trent University, Nov 17, 2006. http://www.ukoln.ac.uk/web-focus/events/seminars/emuit-2006-11/ [last accessed 14/02/07].

KHARE, R. 2006. *Microformats: The Next (Small) Thing on the Semantic Web?* **IEEE Internet Computing**, Vol. 10, No. 1, pp. 68-75 (Jan/Feb 2006).

KHARE, R., CELIK, T. 2006. *Microformats: a Pragmatic Path to the Semantic Web***. Proceedings of WWW2006***,* Edinburgh, UK.

KLEMPERER, P. 2006. *Network Effects and Switching Costs: Two Short Essays for the New Palgrave*. **Working Paper** series. Social Science Research Network. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=907502 [last accessed 15/01/07].

KUKULSKA-HULME, A. 2006. *Learning activities on the move*. Podcast, **Handheld learning conference**, 12th Oct 2006, London. http://www.handheldlearning.co.uk. Available online at: http://dtn.ultralab.net/stage/projects/Handheld_Learning_Podcast/ [last accessed 15/01/07].

KUSHNER, D. 2006. *The Firefox Kid.* **IEEE Spectrum**. Nov 2006. Available online at: http://www.spectrum.ieee.org/nov06/4696 [last accessed 12/02/07].

LAMB, B. 2004. *Wide Open Spaces: Wikis, Ready or Not*. **Educause Review** Vol. 39, No. 5 (Sep/Oct 2004), pp. 36–48. Available online at: http://www.educause.edu/pub/er/erm04/erm0452.asp [last accessed 15/01/07].

LANINGHAM, S (ed.) 2006. *Tim Berners-Lee*. Podcast, developerWorks Interviews, 22nd August, IBM website. Available online at: http://www-128.ibm.com/developerworks/podcast/ [last accessed 17/01/07].

LESSIG, L. 2006. *The Ethics of Web 2.0: YouTube vs. Flickr, Revver, Eyespot, blip.tv, and even Google*. **Lessig blog**, Oct 20th 2006. Available online at: http://lessig.org/blog/archives/003570.shtml [last accessed 17/01/07].

LEVENE, M. 2006. **An Introduction to Search Engines and Web Navigation**. Pearson Education Ltd: England.

LEVY, S., STONE, B. 2006. *The New Wisdom of the Web*. **Newsweek**. Available online at: http://www.msnbc.msn.com/id/12015774/site/newsweek/page/5/ [last accessed 21/02/07].

LIEBOWITZ, S. J., MARGOLIS, S. 1994. *Network Externality: An Uncommon Tragedy*. **Journal of Economic Perspectives,** vol. 8, no. 2, Spring 1994. American Economic Association: USA. Also available online at: http://www.utdallas.edu/~liebowit/jep.html [last accessed 14/02/07].

LIU, Y., MYERS, J., MINSKER, B., FUTRELLE, J. 2007. *Leveraging Web 2.0 technologies in a Cyberenvironmnt for observatory-centric environmental research*. Presented at The 19th Open Grid Forum (OGF19), Jan 29th – Feb 2nd 2007, North Carolina, USA. Available online at: http://www.semanticgrid.org/OGF/ogf19/Liu.pdf [last accessed 14/02/07].

LUND, B. 2006. *Social Bookmarking For Scientists - The Best of Both Worlds*. **Data Webs: new visions for research data on the Web.** 28th June, 2006, Imperial College, London. Also available online at: http://xtech06.usefulinc.com/schedule/detail/75 [last accessed 14/02/07].

LYMAN, P. 2002. *Archiving the World Wide Web*. In: **Building a National Strategy for Digital Preservation: issues in digital media archiving**. April 2002. Council on Library and Information Resources, (Washington D.C.) and Library of Congress.

MANN, C. 2006. *Spam + Blogs=Trouble*. **Wired,** Iss. 14.09, Sept 2006, pp. 104–116. The Condé Nast Publications: San Francisco, USA. Also available online at: http://www.wired.com/wired/archive/14.09/splogs.html [last accessed 14/01/07].

MATTHEWS, B. 2005. Semantic Web Technologies. JISC Technology and Standards Watch. April 2005. Available online at: http://www.jisc.ac.uk/whatwedo/services/services_techwatch/techwatch/techwatch_ic_reports2005_published.aspx [last accessed 14/02/07].

McGRATH, S. 2006. *The Web Service amd SOA proposition*. **Personal notes**. XML Summer School, Web Service and Service-Oriented Architecture track, July 2006, Oxford.

MASANES, 2006. **Web Archiving**. Springer-Verlag: Germany.

MIKA, P. 2005. *Ontologies are us: a unified model of social networks and semantics.* **Proceedings of 4th International Semantic Web Conference** (ISWC 2005), held in Galway, Ireland. pp. 522-536. Springer. Available online at: http://www.cs.vu.nl/~pmika/research/papers/ISWC-folksonomy.pdf [last accessed 12/02/07].

MILLEN, D., FEINBERG, J., KERR, B. 2005. *Social Bookmarking in the enterprise*. **ACM Queue**, Nov 2005. Available online at: http://www.acmqueue.com/modules.php?name=Content&pa=showpage&pid=344 [last accessed 12/02/07].

MILLER, P. 2005. *Web 2.0: Building the New Library.* **Ariadne.** Issue 45 (October 2005). UKOLN. Available online at: http://www.ariadne.ac.uk/issue45/miller/#14 [last accessed 14/01/07].

MILLER, P. 2006. *Introducing the Library 2.0 gang*. Recorded telephone conference as part of the **Talking with Talis** podcast series. Jan 31, 2006. Available online at: http://talk.talis.com/archives/2006/02/introducing_the.html [last accessed 14/02/07].

MORVILLE, P. 2006. **Ambient Findability**. O'Reilly.

NARDI, B., SCHIANO, D., GUMBRECHT, M., SWARTZ, L. 2004. *Why We Blog*. **Communications of the ACM.** Vol 47, No 12 (Dec 2004) pp. 41–46.

NICKLES, M. 2006. *Modelling Social Attitudes on the Web*. **5th International Semantic Web Conference**, Athens, GA, USA, November 5-9, 2006, LNCS 4273. Available online at: http://iswc2006.semanticweb.org/items/ [last accessed 14/02/07].

O'REILLY, T. 2003. *The Architecture of Participation*. ONLamp.com. April 6, 2003. Available online at: http://www.oreillynet.com/pub/wlg/3017 [last accessed 14/02/07].

O'REILLY, T. 2005a. *What is Web 2.0: Design Patterns and Business Models for the next generation of software*. O'Reilly website, 30th September 2005. O'Reilly Media Inc. Available online at: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html [last accessed 17/01/07].

O'REILLY, T. 2005b. *Web 2.0: Compact Definition*. O'Reilly Radar (blog), 1st October 2005. O'Reilly Media Inc. Available online at: http://radar.oreilly.com/archives/2005/10/web_20_compact_definition.html [last accessed 17/01/07].

O'REILLY, T. 2006a. *People Inside & Web 2.0: An interview with Tim O'Reilly*. **OpenBusiness website**, April 25th 2006. Available online at: http://www.openbusiness.cc/category/partners-feature/ [last accessed 14/02/07].

O'REILLY, T. 2006b. *Open Source Licences are Obsolete*. **O'Reilly Radar** (blog). Aug 1st 2006. Available online at: http://radar.oreilly.com/archives/2006/08/open_source_licenses_are_obsol.html [last accessed 14/02/07].

O'REILLY, T. 2006c. *Web 2.0 Compact Definition: Trying Again*. **O'Reilly Radar** (blog). Dec 10th 2006. Available online at: http://radar.oreilly.com/archives/2006/12/web_20_compact.html [last accessed 14/02/07].

OREN, E., BRESLIN, J., DECKER, S. 2006. *How Semantics Make Better Wikis*. **Proceedings of WWW2006**, May 23-26, 2006, Edinburgh, Scotland. ACM Press. Also available online at: http://www2006.org/programme/files/xhtml/p171/pp171-oren/pp171-oren-xhtml.html [last accessed 12/02/07].

OWEN, M., GRANT, L., SAYERS, S., FACER, K. 2006. **Social Software and Learning**. FutureLab: Bristol, UK. Available online at: http://www.futurelab.org.uk/research/opening_education/social_software_01.htm [last accessed 15/01/07].

PATTERSON, L. 2006. *The Technology Underlying Podcasts.* **Computer**. Vol. 39, no. 10 (October 2006). IEEE Computer Society.

PLACING, K., WARD, M., PEAT, M., TEIXEIRA, P. 2005. *Blogging Science and science education.* **Proceedings of the 2005 National UniServe Conference, Blended Learning: design and improvisation Symposium,** 28th 30th Sept 2005, University of Sydney, Australia. Available online at: http://science.uniserve.edu.au/pubs/procs/wshop10/2005Placing.pdf [last accessed 12/02/07].

PRODROMOU, E. 2006. *Web 2.0 summit: Open Source software borrows back from the open API Web*. **LinuxWorld.com**, Nov 13th, 2006. Available online at: http://linuxworldmag.com/news/2006/111306-web20-summit.html?page=1 [last accessed 14/02/07].

RACTHAM, P., ZHANG, X. 2006. P*odcasting in academia: a new knowledge management paradigm within academic settings*. In: **Proceedings of the 2006 ACM SIGMIS CPR Conference** (SIGMIS CPR '06) on Computer Personnel Research, Claremont, California, USA, April 13-15, 2006. ACM Press, New York, NY, pp. 314-317.

REISS, S. 2006. *His Space.* **Wired**, Iss. 14.07, July 2006, pp. 143–147. The Condé Nast Publications: San Francisco, USA. Also available online at: http://www.wired.com/wired/archive/14.07/murdoch.html [last accessed 16/01/07].

ROGERS, A. 2006. *Get Wiki with it.* **Wired**. 14.09 (September 2006), pp.30-32.

ROSENTHAL, D., ROBERTSON, T., LIPKIS, T., REICH, V., MORABITO, S. 2005. *Requirements for Digital Preservation Systems.* **D-Lib magazine.** November 2005. Available online at: http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html [last accessed 14/02/07].

ROSENTHAL, D. 2006. Private e-mail conversation. November 2006. Dr. David Rosenthal is Senior Scientist of the LOCKSS program at Stanford University Libraries.

RZEPA, C. 2006. *Wikis and (Meta)data Rich Environments: a Model for Scholarly Publishing.* Talk given at **Exploiting The Potential Of Wikis**, UKOLN Workshop. Available online (as a wiki): http://www.ch.ic.ac.uk/wiki2/index.php/Main_Page [last accessed 14/02/07].

SCARDAMALIA, M. 2002. *Collective cognitive responsibility for the advancement of knowledge*. In: SMITH, B. (ed.) **Liberal education in a knowledge society**. pp. 67–98. Open Court Publishing Company: Chicago, USA. Also available online at: http://ikit.org/fulltext/inpressCollectiveCog.pdf [last accessed 15/01/07].

SCHMITZ, P. 2006. *Inducing ontology from Flickr tags*. **WWW2006 Conference**, May 22–26, 2006, Edinburgh, UK. Available online at: http://www.rawsugar.com/www2006/22.pdf [last accessed 15/01/07].

SHADBOLT, N. 2006. **Private conversation** at *Memories for Life: the future of our pasts* event. British Library, London, Dec 12th 2006. Event details at: http://www.memoriesforlife.org/events.php

SHADBOLT, N., BERNERS-LEE, T., HALL, W. 2006. *The Semantic Web Revisited.* **IEEE Intelligent Systems.** May/June 2006. IEEE Computer Society. Also available online at: http://eprints.ecs.soton.ac.uk/12614/01/Semantic_Web_Revisted.pdf [last accessed 16/01/07].

SKIPPER, M. 2006. *Would Mendel have been a blogger?* **Nature Reviews Genetics**. 7, 664 (September 2006). Available online at: http://www.nature.com/nrg/journal/v7/n9/full/nrg1957.html [REQURIES REGISTRATION]

SNYDER, B. 2006. **Service Oriented Architecture meets Web 2.0**. Podcast. Nov 1, 2006. Available at: http://searchwebservices.techtarget.com/webcast/0,295011,sid26_gci1228292,00.html [last accessed 14/02/07].

STVILIA, B., TWIDALE, M. B., GASSER, L., SMITH, L. C. 2005. **Information quality discussions in Wikipedia.** Technical Report, Florida State University. Available online at: http://mailer.fsu.edu/~bstvilia/ [last accessed 16/01/07].

STANLEY, T. 2006. *Web 2.0: Supporting Library Users.* **QA Focus.** UKOLN. Available online at: www.ukoln.ac.uk/qa-focus/documents/ briefings/briefing-102/briefing-102-A5.doc [last accessed 14/02/07].

SWAN, A. 2006. *Overview of scholarly communication*. In Jacobs, N. Ed. **Open Access: Key Strategic, Technical and Economic Aspects**. Chandos Publishing: Oxford, UK.

TUCK, J. 2005a. *Collection Development and Web Publications at the British Library***. PowerPoint presentation at: Digital Memory,** Tallin. November 24, 2005. Available online at: http://www.nlib.ee/html/yritus/digital_mem/24-2-tuck.ppt [last accessed 15/01/07].

TUCK, J. 2005b. *Creating Web Archiving Services in the British Library.* **DLF Fall Forum**, Nov 9, 2005. Available online at: www.diglib.org/forums/fall2005/ presentations/tuck-2005-11.pdf [last accessed 15/01/07].

TUCK, J. 2007. Author's notes, *Memories for Life: the future of our pasts* event. British Library, London, Dec 12th 2006. Event details at: http://www.memoriesforlife.org/events.php

TUDHOPE, D., KOCH, T., HEERY, R. 2006. **Terminology Services and Technology: JISC state of the art review.** Sept 2006. UKOLN/JISC: Bristol, UK. Available online at: http://www.ukoln.ac.uk/terminology/JISC-review2006.html [last accessed 15/01/07].

VANDER WAL, T. 2005. *Folksonomy definition and Wikipedia*. Blog entry: **www.vanderwal.net**, 2nd Nov 2005. Available online at: http://www.vanderwal.net/random/entrysel.php?blog=1750 [last accessed 15/01/07].

VARMAZIS, C. 2006. *Web 2.0: Scientists Need to Mash It Up.* **BIO-IT World.com.** April 6th, 2006. Available at: http://www.bio-itworld.com/newsitems/2006/april/04-06-06-news-web2 [last accessed 15/01/07].

WAGNER, M. 2007. *Firefox 3: From Html Renderer To Information Broker.* **Information Week.** Jan 3, 2007. Available online at: http://www.informationweek.com/blog/main/archives/2007/01/firefox_3_from.html [last accessed 20/02/07].

WALKER, J. 2005. *Feral hypertext: when hypertext literature escapes control.* In: **Proceedings of the sixteenth ACM conference on hypertext and hypermedia**, 6th – 9th Sept, 2005, Salzburg, Austria, pp. 46 – 53. ACM Press: New York, USA. Also available online at: http://delivery.acm.org/10.1145/1090000/1083366/p46-walker.pdf?key1=1083366&key2=9121469611&coll=&dl=ACM&CFID=15151515&CFTOKEN=6184618 [last accessed 15/01/07].

WHITSED, N. 2006. *Learning and Teaching*. **Health Information and Libraries Journal.** 23:1 (March 2006), pp. 73-75.

WILSON, S. 2006. *Personal Learning Environment*. Presentation at: **Pushing the boundaries of the VLE II**, Sept. 28th 2006. SURF: Utrecht, Netherlands. PowerPoint slides available at: http://www.cetis.ac.uk/members/scott/resources/utrecht.ppt [last accessed 15/01/07].