Web 2.0 and Scholarly Communication

By Mark Ware

Abstract

We examine the ways in which Web 2.0 tools and services – including blogs, wikis, social bookmarking and tagging, social networking and data interoperability and re-use – are affecting scholarly communication, with examples and usage data where available. We find that many of the tools have yet to live up to their early promise and the expectations that rode on them, and discuss the possible reasons for this.

What is Web 2.0?

This article is about the impact of Web 2.0 on scholarly communication, and the reader might therefore reasonably ask for our definition of Web 2.0. Given the widespread coverage it has had in virtually all media, it feels a bit superfluous to attempt another definition but the conventions of authorship require me to provide something. JISC in a 2007 report defined it thus: "Web 2.0 encompasses a variety of different meanings that include an increased emphasis on user-generated content, data and content sharing and collaborative effort, together with the use of various kinds of social software, new ways of interacting with web-based applications, and the use of the web as a platform for generating, re-purposing and consuming content." This may be accurate and comprehensive but is like many such definitions unhelpful if you do not already have a good idea of what's being talked about. In practice, I am working on the assumption that readers of this article will already their own idea of Web 2.0 (and a pretty good one too) to which we can apply the *elephant test*: it is difficult to describe, but you know it when you see it. [1]

This may become clearer as we progress to specific examples but let me, however, clarify our definition by drawing attention to two (non-exclusive) flavours of Web 2.0: *social* applications (such as blogs, wikis, social bookmarking, social networking) with their emphasis on explicit and implicit user-generated content, and *data* applications such as data mashups in which the web is treated as a computing platform for combining and re-using data in new ways. The distinction is not always clear and furthermore data applications will increasingly tend to overlap with the semantic web (sometimes called Web 3.0, though that label is also unhelpfully applied to other developments, such as 3D or immersive interfaces). No matter. The labels are here to help rather than hinder and we can cheerfully proceed without further theological distinctions.

Web 1.0 and scholarly communication

But before moving on to Web 2.0 it's worth briefly taking stock of how the plain old vanilla web has affected scholarly communication. This could easily be the subject of a whole series of separate articles but let us just highlight the following. Virtually all current (and an increasingly large fraction of historical) journal content (in STM at least) is now published electronically via the web. The combination of this ubiquitous delivery channel with new business models (in particular the big deal and consortia/regional/national licensing) has broadened the access to the literature of the average scientist to historically unprecedented levels. [2] Alongside this access has come powerful free indexing and search tools, notably Google and Google Scholar; it is increasingly the norm for scientists (and not just the younger generations) to start and even end their literature searches on Google, even when they have free (at the point of use) access to superior dedicated tools like Web of Knowledge, Scopus or SciFinder. Unsurprisingly, search-based strategies are of growing importance in keeping abreast of the literature, although browsing remains important. Interestingly in the context of a discussion of Web 2.0 and its social dimension, King & Tenopir found a substantial increase between 1997 and 2003 in the importance of word of mouth in finding articles for chemists (from 3% to 14% of instances).

Relationship with Open Access

It's a rare article on scholarly communication in the late noughties that does not mention open access and we have no intention of making an exception here. There is a clear synergistic relationship between the open access (and related ideas of open data and open science) and Web 2.0. At a trivial level, for instance, a blog commenting on a published paper presupposes access to that paper. Institutional repositories, as a platform for sharing scholarly content, could (perhaps should) be very "Web 2.0" in their design philosophy (although in practice they are not). At a much deeper level, the Web 2.0 culture of content purposing and re-use is much harder to realise within a non-open access environment. One of the arguments in favour of open access [3] is that it facilitates the creation of new services and new knowledge through data mining and data mashups of the published literature. We shall return to this last point, as it will become increasingly important for publishers.

Examples of Web 2.0 in scholarly communication

Blogs

Blogging began during the mid-1990 [4]but did not really take off until the arrival of free, easy-to-use web-based software in 1999. Initially associated with the personal journal and with self-referential commentary on the web itself, the blog in fact can be thought of as a web platform suitable for almost any kind of content. The key features of this platform are: a simple content management system allowing users to create and post content (including rich media such as images, audio and video) to the web without technical knowledge; persistent deep links to individual articles; the ability of readers to leave comments on posted articles; the *trackback,* which automatically appends to the article a link to (and typically a brief extract from) other articles that reference it [5]. Although a few influential blogs do not allow comments (typically because of the costs of removing unwanted comments such as spam or abusive content) it is the last two features that give blogs their social power, converting them from one-way publishing platforms to a web of interlinked conversations.

There are generally thought to be about 100-1500 scientific blogs. For instance, the aggregator site **Postgenomic** covers 750-800 blogs. Its statistics page shows that about 300 of these blogs are active in any given week and the total number of posts averages about 2000 per week[6]. Neither of these figures shows current signs of growth (if anything, the trend may be slightly downwards). Within chemistry, the site **Chemical Blogspace** performs a similar function to Postgenomic. It reports about 60 blogs active per week and an average total of 150-160 new posts per week.

It's also worth noting that there are two commercial blog publishers active in science, Seed Media Group's Scienceblogs and Corante. Both companies rely primarily on an advertising model that means they are focussed on broader audiences, and tend towards a popular science type of coverage.

The most popular blogs on Postgenomic and Chemical Blogspace are a mixture of types, including popular science, news and gossip, personal opinion. Only relatively few have a substantial overlap with scholarly communication such as research-level content, discussion of published articles, etc.

General news and gossip is perhaps unsurprisingly more prevalent. One of the more popular general blogs is **ChemBark**, written by Paul Bracher, a PhD student at Harvard. "Chembark has morphed into the water cooler of chemistry," Bracher was quoted in a draft version of a 2008 Scientific American article [7]. "The conversations are: What should the research agencies be funding? What is the proper way to manage a lab? What types of behavior do you admire in a boss? But instead of having five people around a single water cooler you have hundreds of people around the world." Of course, while this may be conversation between scholars it's not what we usually mean by scholarly communication.

Totally Synthetic

An interesting example of a successful research-oriented blog is **Totally Synthetic**[8]. Started by Paul Docherty in early 2006 when still a PhD student, it covers the synthesis of organic compounds through discussion of recently published papers and has become one of the best known and best read blogs in chemistry, with around 32,000 unique readers per month[9]. Although the number of active commenters is only a small fraction of the total readership, articles receive plenty of comments, with 30-40 comments common and up to 100-150 in some cases. Although it is produced using blog software, Totally Synthetic has close similarities to a virtual journal (selection and highlighting of interesting papers), to a recommendation service like Faculty of 1000 and to post-publication commentary/peer review (e.g. if journals used the same track-back functionality used by blogging software, blogs comments such as on Totally Synthetic could be automatically linked to from the article.)

Social bookmarking

Social bookmarking refers to systems that allow users to store internet bookmarks and categorise them (with "tags") so that as well as being available for the user's own future use, they can be shared, for example with colleagues or with anyone interested in the field represented by the categories used. The field was created by the general-purpose **Del.icio.us**, which launched in 2003. In the academic sector the idea has been expanded so that the systems do not just capture the URL of the bookmarked page but also automatically extract (with the same single click) bibliographic information if the viewer's browser contains an academic article or abstract. The systems also generally allow the captured information to be downloaded to local reference management software such as EndNote, which in turn integrate, with word processing software for authoring purposes.

There are at least three competing services aimed at academics, **CiteULike**, **Connotea** (Nature Publishing Group) and (more recently launched) **2collab** (Elsevier). Most of the leading electronic journals platforms offer clickable icons for at least one of these services, most allowing easy use and promoting the services.

These services potentially offer a number of benefits to academics. One obvious use is to allow a research group to share literature discoveries with each other and to maintain a single shared bibliography. Perhaps more interestingly, it would be possible to use the combined metadata of the user population to identify articles related to a particular article in ways that were not necessarily obvious from the content or keywords. [10]

As of August 2007 Connotea had about 50,000 registered users of which about a third were active. These numbers will have grown since 2007 but are still likely to be small compared to the population of potential users. Surveys during 2007 and 2008 have put the proportion of academics using social bookmarking at about 7-10%. It is not clear why uptake of these services has been so slow, given the apparent utility and ease of use; this aspect is discussed further below.

Social networking

The massive popularity and explosive growth of social networking sites such as Facebook, Bebo and MySpace (and LinkedIn in the business/professional sphere) is well known. These sites allow the building of online communities of shared interest or practice, and provide a variety of means for users to communicate and to share content. Networking is as important to academics as the next professional, whether to find research partners or just to discuss the latest findings – witness the popularity of the academic conference – which has led to academics, publishers and other entrepreneurs to launch social networking sites for academics. For example, **Nature Network** (NPG) is one of the better known sites but there are a host of others including **ResearcherID** (Thomson), **Academia.edu**, **Labmeeting**, **Epernicus**, **ResearchGate**, **Science Advisory Board**, **SciSpace**, **Lalisio**, **SciBog** (sic), **Laboratree**, **SciMeet**, and others.

It has to be said that these sites have not seen the dramatic explosion in use of Facebook and its ilk. There is currently a limited awareness of the potential of web-based social networking among academics. Recent (2007-08) surveys of scientists have found about 10-15% using social networking sites for professional purposes. This is despite the likely high penetration of Facebook among junior researchers (given its high penetration on campuses everywhere). In fact the popularity of Facebook may be an inhibitor – junior researchers may not want to be seen using tools associated with socialising in a professional context (hence the jibe, social networking = social NOT working).

Some more recent sites may be able to avoid the "social as in socialising" connotation. For example sites such as **BioMedExperts**, **pubScholar** and **SciLink** allow users to explore and expand the social network created by the web of literature citations. The American Chemical Society's Member Network has a notably restrained and professional appearance (not dissimilar to LinkedIn) and focuses on professional networking rather than "chat".

Workflows

A newer idea is that of social websites like **myExperiment** [12] ("A Web 2.0 Virtual Research Environment" from the universities of Manchester and Southampton) for the sharing of scientific workflows and experiment plans. Workflows are formal descriptions of processes in specialised computer languages like Taverna, originally mainly used in computational biology and bioinformatics, but with growing use in chemistry, social statistics and even music information retrieval. At the time of writing, myExperiment had over 1250 users and 490 workflows. Although the nature of the content is arcane, the site itself in structure and function is a standard Web 2.0 media-sharing environment, similar to say YouTube (for video) or Slideshare (for presentations).

Podcasts

Podcasts are usually included within the Web 2.0 galaxy as another example of user generated content. Within academic publishing, podcasts are becoming an increasingly common adjunct to online journals and are reported to be popular. The Nature podcast was said by Timo Hannay of Nature Publishing Group to have gained 30,000 downloads within 3 months of its launch in October 2005, an audience level that allowed it to gain sponsorship to cover its costs. Hannay also reported that feedback from podcast users:

> *"indicated that researchers liked hearing the author interviews because it gives them an insight into reports from outside their fields that they would never normally read in the journal. It also allows them to connect with these scientists as people, unfiltered by the formal, passive style of research papers. (Needless to say, authors also love being given a platform to talk about their work in front of tens of thousands of fellow researchers.) More prosaically, the show enables researchers them to make more productive use of their time."*

Podcasts are also said anecdotally to be popular with non-native English speakers as a way to improve their technical English.

Popular or otherwise, however, this kind of podcasting is surely not really a "Web 2.0" phenomenon involving community participation, but more akin to the adoption by media companies of a new distribution channel. There is little evidence of individual researchers creating regular podcasts but it is increasingly common for conferences to offer audio, audio + slide or video recordings of talks, and sites like **SciVee** do seem to be growing.

Wikis

Wikipedia is not just the best known general-purpose user-generated encyclopaedia but for many people defines what a wiki is. Despite initial and continuing scepticism in some quarters about the quality of its content, it is increasingly used by researchers and academics. Although they might not rely on it for critical information (e.g. to support an argument in a grant application or peer-reviewed publication) they do use it for example for quick reference in areas with which they are already competent or for quick overviews of new areas, and see it as a helpful teaching resource.

In addition to the *ad hoc*, article-by-article approach to content generation on Wikipedia, there are also coordinated projects aimed at improving the number and quality of articles within specific scholarly disciplines. There are such WikiProjects in the sciences, technology, engineering as well as other disciplines. For instance the **WikiProject Physics** has about 40 listed participants. It has identified about 9000 physics-related articles on Wikipedia and has set project goals including bringing every physics article as close to

"Featured Article" or "Featured List" status as possible and improving compliance with Wikipedia's style manual.

While of some interest, Wikipedia *per se* is unlikely to have much impact on core areas of scholarly communication. More relevant are specific projects that utilise the core functionality of the wiki platform for research or other scholarly purposes. This core functionality can be seen as a web page that can be created and/or edited by users (either all comers or restricted set of users) or more generally, a web-accessible database of content editable by users.

OpenWetWare (http://openwetware.org/) is now something of a venerable example of community use of a wiki in science. This is aimed at scientists working in biology and biological engineering and focuses on providing a database of protocols and materials for life science laboratory work, plus the facility for research groups and labs to maintain home pages on the site. The OWW statistics show about 11,000 pages. There are 5500 registered users. Use of this site is substantial, with traffic of around 1.5 million pageviews or 275,000 sessions per month.

The recent growth of such wiki-like sites in biology has been such that researchers have begun to joke about "wikiomics". One example is **WikiPathways**, which uses standard wiki software to create a site "dedicated to the curation of biological pathways by and for the scientific community". As of mid-2008 WikiPathways had some 350 registered users, of whom 50 or so had made changes to at least one pathway [13].

Other similar examples include **PDBWiki** (biological molecular structures), **Proteopedia** (proteins & other molecules), **Chemspider** (chemical structures), **Galaxy Zoo** (galaxies), **Zebrafish GenomeWiki** (community annotation of the zebrafish genome), **WikiSpecies** (taxonomy), and **Proteins Wiki** (proteins, especially structure and function; although it contains nearly 50,000 entries, it is currently described as "inactive").

Other projects extract content from Wikipedia, add content and/or improve the quality of the pages, and then either push the improved content back to Wikipedia or publish it independently[14]. A good example is **Gene Wiki**, which used a software robot to extract some 9000 Wikipedia articles on human genes that were then combined with information from NCBI's Entrez Gene database, links to data repositories and to the literature.

One issue that may prevent academics from contributing to wikis is the lack of attribution for their work, which is important both in terms of moral rights but also for career and professional advancement. And of course from the user's perspective, authorship attribution is important to assess the origin, authority and reliability of information. An interesting attempt to address this is **WikiGenes**[15]. This uses newly developed wiki software that allows users to easily identify the author of every word and also allows users to rate other users. WikiGenes also provides editing tools that provide authors with integrated database and ontologies look-up, which both simplifies the authoring process and improves the quality of the result (by facilitating consistency).

A tool closer to the publishing world to **GoPubMed** (a free/paid-for service), which provides enhanced searching of PubMed. First, it text-mines the papers, allowing it automatically to identify keywords and concepts, and to populate a subject taxonomy from this. In addition, the system allows users to curate the content by improving the text-mined categories.

Publisher wikis

Academic publishers have been slow to adopt wikis, most likely because the wiki model relies on open, editable and reusable content that is not easy to monetise. Three experiments are worth mentioning, though none is conclusive. Elsevier's **WiserWiki** was launched in early 2008 with content seeded from an existing (out of print) textbook (Textbook of Primary Care Medicine, Third Edition, by John Noble). Only qualified doctors are allowed to edit or create pages. After about a year of operation, the site had received a total of 600,000 page views, with some 200 valid content pages available.

Elsevier's **SciTopics** was launched in January 2009 (previously soft-launched in beta form as Scirus Topic Pages in June 2007). It allows invited experts to maintain pages on topics of their choice, with summaries of the topics, further reading and web links supplied by the expert, supported by automatically generated links to related articles in Scopus and search results from Scirus. Quality control is provided by the selection of editors and by the moderation of their content by 14 subject editors. At launch Elsevier said it contained 650 live topic pages with many more in draft.

The journal ***RNA Biology*** recently changed its policy to require authors of articles on RNA families also to submit a draft article on the RNA family for publication in Wikipedia. The hope is that the Wikipedia page will become the hub to collect later information about the RNA family. (One suspects the journal will also not be averse to the additional traffic that flows from the page to the journal, given the high position that Wikipedia pages generally have in Google searches.) The move has not been welcomed by all Wikipedia supporters, some of whom argue that such specialised and narrow content is not suitable for a general encyclopaedia.

Data

The importance of improving the ability to use and reuse research data and their integration into the research literature has been covered by many other authors. At a broad level, the *Towards 2020 Science* report [16] concluded "Our findings have significant implications for scientific publishing, where we believe that even near-term developments in the computing infrastructure for science which links data, knowledge and scientists will lead to a transformation of the scientific communication paradigm" and more recent articles [17] have reported on some early strategies for managing scientific data and integrating it with publications.

There is clearly huge potential in the creation of open scientific datasets and in the development of interoperability standards to allow these datasets to be shared and combined in new ways. Historically, scientists in many (although not all) fields have been reluctant to share data, whether for competitive secrecy or just because they were too busy. Increasingly, however, research funders are requiring researchers to deposit a copy of data (usually after an embargo period) in an open repository as a condition of funding. Levels of compliance with existing mandates are not known, as they are not routinely monitored. A paper in PLoS, though, reported that up to 20% of qualifying articles did not have corresponding entries in GenBank as they should have done [18]. Institutional repositories focussing on data rather than publications are starting to emerge (e.g. Oxford, the TU3 Federation, etc.), although the evidence (e.g. from studies like StORe and SPECTRa) is that discipline-specific repositories are required.

Although not primarily a Web 2.0 issue, there are Web 2.0 approaches that could be employed in this field. There is an overlap with the discussion of specialist wikis in the previous section. The **ChemSpider** database, for example, was initially launched as a free chemical structure database but its longer-term ambition is to use the database as the centre of a "Chemical Structure Centric Community for Chemists" in which wiki-style crowd-sourcing, curation and social network features are a core part. Unregistered users can post comments, while registered users ("curators") can upload and actively edit the content.

It is notable that the research projects investigating the linking of data and publications are mainly working with preprint repositories rather than journals, and without the direct participation of publishers. (R4L is one exception, in which ALPSP was a partner.) Publishers surely need to get more involved.

Peer review

Conventionally peer-reviewed literature continues to be the cornerstone of scholarly communication. The community appears to have a conservative approach to change in the way peer review is conducted (e.g. see our survey conducted for the Publishing Research Consortium [19]). For instance, while there is a strong intellectual argument for open peer review, in practice reviewers are reluctant to publish signed reviews, not least because of concerns about the repercussions of giving negative reviews in public.

Post-publication review is an approach that combines the Web 2.0 features of commenting and rating to the scholarly literature. The PRC survey found about a third of researchers said they supported post-publication review despite some obvious weaknesses (e.g. it encourages instant reactions) provided it was a supplement to formal peer review and not a replacement for it. In practice, however, in many trials has proved difficult to persuade researchers to comment or rate articles. For example, the open access publisher BioMed Central has offered a commenting feature since November 2002. Up to July it accumulated a total of 945 comments from 753 different users on 732 unique papers, out of a total of 37,916 papers by BMC over the same period. In other words, only 2% of BMC papers had attracted comments. Furthermore, some 40% of these were author updates (including corrections of errors) and author links to supplementary information.

The open access journal PLoSONE is designed to depend on post-publication review. Its peer review system prior to publication is designed to accept all papers that are judged to be technically sound, with judgements about the importance of any particular paper being made post-publication by the community in the form of comments and ratings (using a 1–5 scale) left on the journal's website. An analysis made in 2008 showed that 23% of published papers (647 out of 2773) had comments, notes or replies and that 13% had ratings.

We built it, why won't they come?

The research community has been surprisingly slow to adopt Web 2.0 solutions to scholarly communications needs.

For example, a number of surveys in 2007 and 2008 reported the proportion of scientists reading blogs on a regular basis to be no more than about 15%. Furthermore the amount of time that even the more active blog users devote to reading blogs is very small in comparison to the estimates of time spent reading the literature [20].

In the CERN & APS surveys reported by Paul Ginsparg [21] less than 10% had tried social bookmarking tools, and of these only 1% found them useful. In the PRC survey mentioned above (January 2008) about 7% of respondents said they used social bookmarking.

We described above the growth in wiki sites (most notably in life sciences). It remains to be seen, however, to what extent the community will support these new wikis: some researchers say that previous attempts to engage the community in supporting earlier biological databases have foundered and wonder why these will be different.

One problem is the proliferation of competing services in the same area and fragmenting their audiences. There are three competing academic social bookmarking services, plus numerous non-specialist ones. We listed above well over a dozen social networking sites: until a clear leader emerges, would-be users (apart from early-adopting enthusiasts) may be actively discouraged from participating by this situation, not just because the likelihood of your real-life network being present on any given network is low, but also because hard-won social capital (content, relationships, etc. hosted on the site) would be at risk if your chosen network were to fold. Similarly, it's easy to see why hard-pressed scientists may feel it's too much trouble to decide which biological wikis are going to flourish and which will fold.

David Crotty, Executive Editor at Cold Spring Harbor Protocols has published a thoughtful account of the current crop of Web tools for biologists and why they are not more successful [22]. He argues that there has been too much "Web 2.0 for the sake of Web 2.0", copying without thought from the consumer sector, and too much emphasis on the social rather than the timesaving aspects. He sees the main reasons for lack of adoption as being lack of time (both in the sense that researchers do not have time to take up new tools without a clear payback, and in the sense that researchers lack the (uncertain amount of) time to wait for the community to respond to a posted query); lack of incentive (you get no credit for commenting on someone's paper, so why do it?); lack of attribution; lack of critical mass; inertia (why learn new tools – "good enough and familiar" is favoured over "better and hard to learn"); and inappropriate tools that do not fit the culture of science ("scientists do not interact like teenagers chatting or rock bands reaching out to their fans").

Web 2.0 proponents have argued that it was not surprising that uptake of Web 2.0 tools would be measured, given the complex mixture of social and psychological barriers to widespread adoption. (Such factors may for instance account for the disparity between the relatively enthusiastic adoption by physicians of online discussion forums and that of biomedical researchers.) Web 2.0 tools are also still at a very early stage of development: clearly many of them of sub-optimal or simply redundant, and most of those around now will not survive the Darwinian struggle for survival. There are certainly generational differences although these are not quite as pronounced as some might think. An interesting finding of the PRC peer review survey (replicated in other work we have done) was that Asian researchers in less developed countries (China, India, etc.) were substantially more likely to use Web 2.0 tools than their Western counterparts. This was partly but not wholly linked to their lower average ages; one might also speculate that it had something to do with the relative youth or weakness of the existing social/professional networks and traditions.

So persuading researchers to adopt Web 2.0 tools will take time. For some people it may be proving harder than initially anticipated: Timo Hannay of Nature Publishing Group talking at the British Library in 2008 said [23]: "But I'm less optimistic about the inevitability of this potential [for the web to greatly improve the productivity of doing science] being fully realised, at least in anything less than a generational timescale. For every scientist who sees it as self-evident that they should be using these tools, or promoting open information-sharing, there are dozens who just don't see the point. For every publisher or librarian who 'gets it' there are many who don't – at least not fully and not yet."

Our view is that Web 2.0 technology offers tremendous potential to enhance scholarly communication. Adoption rates may have been slower than some anticipated but we suspect this is likely to be because the first generations of tools simply have not been good enough either in terms of the additional value they provide or in terms of suitability for a researcher use. Paradoxically, Web 2.0 may become less visible as they become more widely adopted, as they are incorporated into existing platforms or into the underlying platform of the Web itself.

Further reading

The following are recommended for accessible further reading:

Science 2.0—Is Open Access Science the Future? by Mitch Waldrop, *Scientific American* (April 2008) http://www.sciam.com/article.cfm?id=science-2-point-0.

Web 2.0 in Science, by Timo Hannay, in *CTWatch Quarterly* (August 2007) http://www.ctwatch.org/quarterly/articles/2007/08/web-20-in-science/

And for what lies beyond Web 2.0, *The Future of Research (Science & Technology)*, by Carole Goble. Presentation to British Library Board Awayday Sept 2008, http://www.semanticgrid.org/presentations/BritishLibrary2008GOBLEpublished.ppt

Mark Ware

Director, Mark Ware Consulting Ltd
Email: info@markwareconsulting.com
Website: www.markwareconsulting.com

Notes & references

[1] Based on US Supreme Court Justice Potter Stewart's famous ruling on obscenity "I know it when I see it".

[2] Anecdotally, I have been struck by many conversations with scientists who support the idea of open access for moral or political reasons but then say they do not in fact expect it to improve their own access to the literature in practice, because they feel they already have access to what they need.

[3] e.g. see Swan, A. (2006) Open Access: Why should we have it? Cahiers de la Documentation: Bladen voor Documentatie. Available at http://www.keyperspectives.co.uk

[4] Although a respectable case could easily also be made for its historic roots being in the threaded interface of the pre-Web bulletin boards with their similar convention of posted topic and multiple responses

[5] The physics repository, arXiv, began accepting blog trackbacks in 2005

[6] http://www.postgenomic.com/stats.php

[7] Waldrop M (2008) Science 2.0—Is Open Access Science the Future?: Scientific American http://www.sciam.com/article.cfm?id=science-2-point-0

[8] http://totallysynthetic.com/blog/

[9] Personal communication

[10] Nature Publishing Group have demonstrated the potential of this approach by integrating Connotea with the EPrints repository software, see http://blogs.nature.com/wp/nascent/2006/03/tagging_and_bookmarking_in_ins.html

[12] http://wiki.myexperiment.org/index.php/Main_Page

[13] Big data: Wikiomics, by Mitch Waldrop, Nature 455, 22-25 (2008) doi:10.1038/455022a

[14] Such resuse of content is of course explicitly permitted indeed encouraged by the terms of the Wikipedia licence

[15] Hoffmann, R. (2008) A wiki for the life sciences where authorship matters, Nature Genetics 40, 1047 - 1051 (2008) doi:10.1038/ng.f.217

[16] Emmett, R. et al. (2006) Towards 2020 Science, Microsoft. http://research.microsoft.com/en-us/um/cambridge/projects/towards2020science/

[17] e.g. Borgman, C. (2008) Data, disciplines, and scholarly publishing. *Learned Publishing*, 21, 29–38 doi: 10.1087/095315108X254476; Carlson, S. (2007) Lost in a Sea of Science Data, Chronicle of Higher Education, 52 (42), A35; Strickland, P. *et al.* (2008) Integrating research articles and supporting data in crystallography, Learned Publishing, 20, 63–72 doi: 10.1087/095315108X248347

[18] Noor et al. (2006) Data Sharing: How Much Doesn't Get Submitted to GenBank? PLoS Biology 4(7) doi:10.1371/journal.pbio.0040228

[19] Ware, M. & Monkman, M (2008) *Peer review in scholarly journals: Perspective of the scholarly community – an international survey.* A research report for the Publishing Research Consortium.

[20] e.g. see RIN or King/Tenopir data

[21] Ginsparg, P. (2007) Next-Generation Implications of Open Access, CTWatch Quarterly (August 2007) http://www.ctwatch.org/quarterly/print.php?p=80

[22] Crotty, D. (2008) Web 2.0 for Biologists–Are any of the current tools worth using? posting on Bench Marks blog http://www.cshblogs.org/cshprotocols/2008/04/03/web-20-for-biologists-are-any-of-the-current-tools-worth-using/

[23] http://blogs.nature.com/wp/nascent/2008/10/social_not_working.html