

**TOWARDS OPEN SOURCE SCIENCE? AN ANALYSIS OF WEB 2.0 BASED
INITIATIVES IN THE SCIENTIFIC PUBLISHING SECTOR.**

Roberta Cuel

Dipartimento di Informatica e Studi Aziendali

Università di Trento

Diego Ponte

Dipartimento di Informatica e Studi Aziendali

Università di Trento

Abstract

Internet is consistently shaping the way scientific knowledge is produced and disseminated. While initially the advent of the Internet allowed to fasten and to enhance many of the services offered by traditional scientific publishers, a multiplicity of innovative applications inspired by the collaborative potentials of the so called Web 2.0, have recently pushed the sector to a second renewing process. In spite of the burgeoning number of initiatives that are providing traditional and new services in innovative ways, there is no clear understanding on how and to what extent these initiatives might affect the publishing sector as a whole. The aim of this paper is to propose a first non exhaustive overview of services provided by scientific publishers before and after the advent of the Internet and to review the large arena of Web 2.0 based initiatives provided by publishers, scientific communities, and other new competitors. Authors want to demonstrate that the advent of the Internet first, and the development of Web 2.0 applications second, challenge the scientific publishing market and the business models of publishers.

INTRODUCTION

Scientific publishers (thereafter simply publishers) play an important role in the current scientific publishing field by providing services for libraries, research institutions, scientific communities and individual scholars. Although over time the most important services offered by publishers have not significantly changed, the advent of the Internet consistently shaped the way through which they have been provided.

From an historical point of view, the Internet revolution might be divided in two main phases. In the first phase (thereafter simply Web 1.0), the rapid evolution of various Information and Communication Technologies (ICT) dramatically cut the costs of sending information. This opportunity allowed to fasten and enhance many of the services provided by publishers such as, for instance, the accessibility to journals and the review processes. In general, we can say that this phase saw the sector slightly shifting from a paper-based to a digital-based medium.

The second phase of the Internet revolution, saw the birth of the so called Web 2.0. Although the notion of Web 2.0 is still under debate (Anderson, 2006), this second phase generally identifies a shift in a more proactive approach of digital collaboration, interaction and sharing of knowledge. Within the scientific publishing industry, this shift can be seen in the multiplicity of innovative initiatives that are exploring innovative ways of scientific knowledge production and dissemination.

While this burgeoning number of Web 2.0 projects gathers attention from researchers, libraries and publishers who foresee huge potential benefits, their increasing number, their diversity and inconsistencies show that we are still far from a clear evaluation of whether and to what extent these new initiatives might affect the publishing sector as a whole.

Starting from these considerations, the aim of this paper is to propose an overview of the services as traditionally offered by scientific publishers before and after the advent of the Internet, to review the large arena of Web 2.0 based services provided by heterogeneous actors (traditional publishers, scientific communities, new competitors) and to propose a first analysis of how these new services might affect the current configuration of the sector.

This paper is organized as follows. Section 1 introduces the scientific publishing sector and its peculiarities, it describes also actors and their role in that market. Section 2 focuses on services currently offered by the major scientific publishers, and overviews from a historical perspective how the most important services evolved due to the diffusion of the Internet.

Section 3 proposes the description of Web 2.0 phenomenon, some of the most important Web 2.0 initiatives and their effects on the scientific publishing sector. Section 4 and 5 focus on how traditional publishers are reacting to the Web 2.0 revolution, research communities are increasing their presence in the publishing market and new actors are entering in the large publishing arena providing online services for the production and dissemination of scientific contents.

1. THE SCIENTIFIC PUBLISHING MARKET

This paper focuses on the science, technology and medicine (from now on: STM) publishing industry, a small part of the scientific publishing sector which does not include humanities, arts and social sciences. A recent OECD report (Houghton, 2005) states that the core STM publishing market is estimated between USD 7 and 11 billion, which corresponds to a less than 10 per cent of the scientific publishing industry. Although, North America and Europe are the bigger markets of scientific publishing (respectively with 58 per cent and 26 per cent of the total), several Asian countries show double-digits increases per year.

The scientific publishing sector is considered as a quite complex market. The product exchanged is the output of scientific research (scientific knowledge) which has been considered both as a public good and the core element of economic growth. Therefore, it should be shared and disseminated in order to promote innovation and social development.

1.1. The most important actors in the scientific market

There are many actors that interact in the scientific market, having different needs and behaviours. The most important ones are researchers, publishers and libraries (Gudeon, 2001; ECDGR, 2006; Enserink, 2007; House of Commons, 2004; Wellcome Trust, 2003; STM, 2008):

Researchers might act with two identities: authors and editors/reviewers. As authors they produce scientific papers, book chapters or monographs. Publications are considered as the most important artefacts that reify researchers' work. Researchers' career relies mainly on the ability to publish significant results in well recognized journals (usually with high impact factors). As editors/reviewers they act as certifiers of knowledge (often) on behalf of scientific publishers.

Publishers fall in one of the following two categories: commercial/for profit firms and not for profit organizations. While the market is populated by a myriad of small size publishers, the first five commercial publishers (Reed-Elsevier, Springer, Thomson, Wolters Kluwer, and John Wiley) account for more than 50 per cent of the whole market (Madras, 2008) while the whole set of commercial firms account for about 64 per cent of market share. Not for profit publishers may be divided in two different sub-categories: learned society publishers and institutional publishers. Learned societies are prominent networks of scholars and professionals with a common interest in a particular area. Their primary goal is the dissemination of research findings related to their field, usually through journals. The major learned society, the American Chemical Society (ACS), accounts for less than 4 per cent of market share while learned society publishers account of 30 per cent market share. Institutional publishers, such as University Press and Cambridge University Press, are associated with a University or Research Institutions, and account for about 4 per cent of the market share. Although all these actors gather intellectual work from authors and manage it to publish books and periodicals, in the last decades non-profit organizations have pioneered new licensing models (open access and creative commons).

Libraries have the aim to guarantee the access to past, present and future scholarly literature and research documentation, ensure its preservation, assist users in discovery and use knowledge, and offer educational programs to enable users to develop lifelong literacies. They usually buy books and periodicals journals from publishers or, as it happens in the last years, they guarantee access to digitalized libraries. In the publishing market they are considered as the most important final customer.

1.2. The supply and demand sides of the scientific market

The scientific publishing market can be analyzed in both demand and supply sides, but many actors (researchers, research communities, universities, learned societies, editors, publishers, libraries and common readers) play important roles in both sides producing and offering knowledge or demanding it. Some others, publishers in particular, usually play the important role of intermediary among them.

Researchers play their role in both demand and supply side, they respectively offer scientific results usually as author of scientific publications and look for papers in order to explore and exploit their research topics. Researchers might also contribute to the publishers work,

reviewing papers, judging the goodness of works, and selecting the most significant and relevant scientific contributions.

Libraries, single researchers, research-oriented firms and the government demand scientific results in the forms of databases, articles, journals, digital libraries, etc. Currently the most important customers of the scientific market are academic/medical institutions with 60 per cent share followed by corporate customers (25 per cent) and government (5 per cent) (House of Commons, 2004). Individuals and single researchers are not considered as part of the customers composition, since they use the services provided by libraries for free. From the point of view of publishers, although individuals obtain the most important benefit, they are not considered as customers.

Research communities and universities might influence, the success of a journal raising their level of reputation which is often expressed by the so called impact factor (based on the citations number).

Scientific communities might desert some journals, nor reading papers or submitting contribution to more reliable ones.

Publishers play mainly the role of intermediaries. They edit journals, select (thanks to the work of reviewers and scientific editors) the best papers and the most significant scientific contributions, publish and disseminate scientific results in the whole society negotiating mainly with libraries. It is common thought that this industry is controlled and managed by few commercial publishers as in an oligopoly. This is due to several reasons. They do not have a specific research field to foster such as learned societies. Commercial publishers might get in touch with a wide number of authors and research fields, developing as much as journals as research disciplines. Learned society publishers are usually reluctant to endeavour in new research fields if this does not explicitly come from a scientific community request. Publishers do also maintain “soft” relationships with authors and reviewers, none of which materially gain from this. This relationship is essential to the reputation of both publishers and researchers.

2. PUBLISHER SERVICES AND THEIR EVOLUTION

Publishers offer a set of services to the scientific publishing field (STM, 2008), the most important and commonly accepted services supplied by publishers are:

Registration (acknowledgement and defence of the moral rights of authors). This activity aims at establishing the intellectual property of authors' concepts and ideas especially the right of attribution and integrity.

Certification. This service aims at certifying the good quality of scientific knowledge and validating the research findings. Both goals are achieved by means of the review process, which is considered as one of the most important phases of scientific knowledge quality check.

Formalisation. The formalization aims at improving the formal quality of published material. This is done by means of (often internal) services that take care of verifying references and indexing, checking the typesetting on the basis of the internationally supported standards to improve accessibility of research. Some communities have developed a well shared structure of scientific paper, including the way in which state of the art, hypothesis and demonstration are organized and presented in the document.

Dissemination. This service aims at creating strong global distribution systems which save authors from having to publish their research on a country-by-country basis and help readers to find good scientific knowledge.

Preservation. This service aims at preserving over time scientific research which was produced before the Internet and after the Internet.

Use. This service aims at improving the overall quality of whole sector by fostering interoperability among publishers, and helping users in finding knowledge.

In the last decades the scientific market has been influenced by many technological changes, in particular by the advent of the Internet. The above mentioned services and the relations among actors have changed over time. In the following paragraphs some of these changes have been highlighted.

2.1. Publishing services before the advent of the Internet

Before the diffusion of the Internet, the relationship between publishing service providers and scientists took usually place mainly through off-line channels. Almost all communication streams between researchers and publishers were based on snail mail. This was particularly the case of both the submission and the reviewing processes. Researchers usually sent hardcopies of their papers to publishers, who in turn rerouted them to the designated

reviewers. The review process was usually based on blind peer reviews (authors do not know who reviewers are, but reviewers know authors' names), or double blind peer reviews (authors do not know who reviewers are, and reviewers do not know authors' names). Then, reviewers sent evaluations, suggestions for improvement and corrections to publishers, who sent back to researchers.

The dissemination of knowledge was based on hardcopies of journals sent to libraries, institutions, and individuals subscribing the journal. The preservation of journals was done mainly by libraries that stored hardcopies and made them accessible to single users visiting the library and accessing to the collection of documents.

2.2. Publishing services after the advent of the Internet

After the advent of the Internet, a large share of the activities related with the creation, submission and evaluation of papers and scientific contributions began to take place in digital form and most of the digital services started to be offered online. Nowadays, most of the communication between researchers and publishers is based on email or takes place within *ad hoc* websites managing the whole submission and evaluation process (e.g., for example the Elsevier management website and EasyChair conference system). The digitalization of the whole process speeded up many phases of the publishing process: the submission of the paper to publishers and to reviewers, the whole review process and the feedback to authors. Heavy changes can be found in the way knowledge is disseminated because an exponentially growing number of journals is offering online access to their content. Researchers can thus browse much more knowledge than before the advent of the Internet, refining also the research process of scientific artifacts, and reducing time of accessing it.

Likewise, libraries have been jumping on the digital revolution as a way to cope with decreasing purchasing budgets and other costs related to hardcopies management (e.g., storage costs). While this has sometime taken the form of bundling hardcopies to rights of online access of the digital version due to the existence to European questionable tax policies, the underlying trend is clearly into a gradual substitution of physical with digital copies (ECDGR, 2006). This is due to the fact that most of the European Countries' VAT policies foresee a different VAT percentage for printed and digital periodicals. This policies create a bizarre practice within the publishing industry as clearly explained by the European Commission Report “[...] simple computations show that even if holding the print collection

does not bring social benefit, there is scope for the “print+electronic” option to be chosen by the library when the VAT rates are not the same.” (ECDGR, 2006, p. 55). As an example, consider that in Italy the VAT for printed journals is 4 per cent while for the digital version is 20 per cent. In France the situation is even worse as the printed journals’ VAT is 2,1 per cent while for the digital one is 21,6 per cent. As a matter of fact the market is still struggling to definitively jump on the “only digital” dissemination model.

One of the most debated services and an open problem of current mainstream market is the digital preservation of scientific knowledge. Indeed the market is experiencing several situations: some publishers permit to libraries to store the digital copies of their journal on libraries’ owned archives; others do not allow libraries to store the copies and sell an “access right” to their archives.

Table 1. Publishers services before and after the advent of the Internet

<i>Period</i>	<i>Before the advent of the Internet</i>	<i>After the advent of the Internet</i>
<i>Publishers’ services</i>		
<i>Registration</i>	Based on mail and printing dates	The whole process has been moved to a digital medium
<i>Certification</i>	Reviews: peer-review Impact factor and h-index	The general system has not been influenced. Nonetheless now it is faster due to the digital medium
<i>Formalisation</i>	Editors and publishers check papers	Authors, editors and publishers check papers
<i>Dissemination</i>	Based on hardcopies	Based on a mixed print/digital version. The market is slightly moving towards a “digital only” market
<i>Preservation</i>	Libraries store hardcopies in their rooms	Publishers preserve electronic data although the situation is evolving
<i>Use</i>	Libraries rooms	Websites with user-friendly navigation systems

Summarizing, it is possible to say that, in its early stage, the publishing sector was not heavily shaped by the introduction of the Internet, a part of the channel of communication through

which the whole process works (Table 1). The digitalization allowed fastening up several processes such as the registration, the certification and the dissemination and search of papers and publication. One of the most important tricking aspects relies at the preservation services which currently is a major issue under debate.

A less discussed issue is the increasing charge of activities that authors have to carry on in writing and submitting digitalized papers. It is a matter of fact that in the interaction with publishers, authors are asked to produce a final version of their reports (the camera ready), which is very close to the published version of scientific contributions. Therefore, authors claim that the amount of work publisher usually do, in reshaping papers and producing the camera ready versions, is very condensed because a large share of the effort involved is nowadays delegated to authors.

3. WEB 2.0 INSPIRED PUBLISHING INITIATIVES

The second phase of the Internet revolution, saw the birth of the so called Web 2.0 which considers the web as a social platform by adding a new layer of information interactivity based on tagging, social networks, user-created taxonomies (McAfee, 2006). Some relevant, albeit non-exhaustive, examples of Web 2.0 examples are the blogging platform Wordpress, the social media site Digg, the image hosting site Flickr, the social bookmarking site Delicious, the video sharing platform YouTube, the business-oriented social networking site LinkedIn. More in general the services provided by Web 2.0 platforms might be summarized in: Blogs, Wikis, social networking, tagging of knowledge and syndication services. All these services can also be successfully implemented within the scientific publishing industry.

In this sector, a multiplicity of innovative initiatives are exploring more collaborative and proactive ways of scientific knowledge production and dissemination, but the most intriguing aspect is that such services are being offered by a heterogeneous actors such as: traditional publishers, Web-based companies and startups, communities of researchers, learned societies. While traditional publishers are still controlling the mainstream market, the disintermediation as offered by the Internet infrastructure permits old and new actors to offer innovative services.

Therefore, rather than on traditional literature review, the research of innovative services was based on browsing the web, analyzing (digital) journal articles, interviewing researchers and publishers, and reading blogs. The results of our activities, far from being complete and

exhaustive, allowed us to categorize the various Web 2.0 initiatives according to the type of service provided:

- Journals and other publications with open and online peer review processes
- Social bookmarking systems
- Citation indexing initiatives
- Scientific blogs and wikis.

In the following subsections each of these categories will be reviewed. Finally some enhanced initiatives such as general platforms and research projects will be analyzed.

3.1. Journals and other publications with open and online peer review processes

A first example of Web 2.0 inspired initiatives is the growing set of internationally recognized journals and archives that have adopted an open and collaborative process of evaluating scientific papers. To cite but a few, these are arXiv, Nature Precedings and Plos One. By trying different kinds of “online” and collaborative review systems, these latter seek to improve the traditional “offline” blind peer review process. The following examples were chosen on the basis of two aspects: the first is the relevance of each case study. The second relies on the innovativeness of the review system as adopted by each initiative. The review service is the one in which the collaborative features of the Web 2.0 might have an important impact while at the same time is one of the most debated about. Thus it is necessary to analyze how such innovative experiments cope with this aspect.

ArXiv.org is an e-print service offered by Cornell University in the fields of physics, mathematics, non-linear science, computer science, quantitative biology and statistics. ArXiv is a highly-automated electronic archive and distribution server for research articles. The main characteristic of arXiv is that papers are not evaluated on the basis of the peer review system. Instead, to submit a paper to the system, researchers must obtain an endorsement which can be given by other arXiv endorsers only. The idea surrounding such policy is that it allows restricting arXiv submissions to relevant and legitimate research contributions without adding to the administrative cost of managing peer review.

Nature Precedings is “.. a free online service from NPG (Nature Publishing Group) that enables researchers in the life sciences to openly share preliminary findings, solicit community feedback, and claim priority over discoveries by posting preprint manuscripts,

white papers, technical reports, posters, and presentations.”. Nature Precedings is thus an archive for publishing preliminary research and initial findings. The submitted documents are reviewed by the team of Nature that ensure the quality and appropriateness of submitted documents. However, this check is not seen as a substitute of the peer review system. After this check the material is posted online and open to comments and sharing.

PLoS ONE is an international, peer-reviewed, open-access, journal which is experimenting post-publication peer review. The PLoS ONE core principle is that “Scientific progress requires the exchange and discussion of data and ideas. PLoS ONE is a unique publication dedicated to presenting the results of scientific research from any scientific discipline in an open-access environment. At the same time, it provides a forum in which to discuss that scientific research and so provide for each and every paper its maximum possible impact”. PLoS ONE provides several web 2.0 services such as qualitative and quantitative feedback from readers (average rating and annotations) as well as linking services (trackback of the paper and social bookmarking).

Atmospheric Chemistry and Physics (ACP) is an open access journal launched in 2001 by the European Geosciences Union. What characterizes this journal is its two-stage publication process. This is made of a first access peer-review focused on the quality of the paper. After this first quality check, an “Interactive Public Discussion” takes place. This process lasts for about a year. In the first eight months the paper is subject to an open discussion period. After this period the author might counter-comment within 4 months. The journal does offer Web 2.0 tools such as RSS or tagging.

Journal of Interactive Media in Education (JIME) is a journal whose topics are the theoretical and practical aspects of interactive media in education. This journal provides a particular peer review system based on a first “hidden” phase where the author(s) openly interact with selected reviewers and a second “open” phase in which the paper is discussed with the broader scientific community. Finally, the most interesting review comments/exchanges are published with the final version, providing readers with insight into the issues that arose during review, and enabling them to build on those discussions. JIME does not use Web 2.0 tools.

Biology Direct, launched in February 2006, is an open access, peer-reviewed online journal. While it aims at encompassing the full spectrum of biological science, currently three are the covered areas: Genomics, Bioinformatics and Systems Biology, Immunology and

Mathematical Biology. According to the website, the peer review is quite innovative and is based on three dimensions: first, the authors are responsible for obtaining reports from three members of the editorial board; second, the peer-review process is open and not anonymous; third, the reviewers' reports are public. If accepted, papers can be commented by readers. Finally, articles published in Biology Direct can be linked to Web 2.0 tools such as CiteULike, Connotea and Del.icio.us.

Philica is an online academic journal accepting publications on any subject. All articles submitted to *Philica* are immediately published and the review process takes place after the publication. Reviewers, which remain anonymous, can both rate the paper and give qualitative reviews. Any reader can review the article; indeed reviewers are not selected by editors.

3.2. Social bookmarking systems

The exponential growth of available knowledge in the Internet, might become a problem as it raises the inefficiencies of searching for good knowledge. To avoid such problem several websites started to offer social bookmarking tools that allow the management, commenting upon and publishing of scientific knowledge links. The most used social bookmarking tools are Connotea, a “Free online reference management for all researchers, clinicians and scientists” owned by Nature Publishing Group (NPG), CiteULike, is a “free service for managing and discovering scholarly references” recently acquired by Springer, BibSonomy a “social bookmark and publication sharing system”.

These websites offer more or less the same services such as online storage and organization of personal bookmarks. These are often visible to other users. Such shared lists are helpful to develop sophisticated collaborative filtering algorithms that allow making socially constructed recommendations lists (e.g. like Amazon suggested books).

BibSonomy is a social bookmarking and publication-sharing system. It aims to integrate the features of bookmarking systems as well as team-oriented publication management. BibSonomy offers users the ability to store and organize their bookmarks and publication entries and supports the integration of different communities and people by offering a social platform for literature exchange. Both bookmarks and publication entries can be tagged to help structure and re-find information. As the descriptive terms can be freely chosen, the assignment of tags from different users creates a spontaneous, uncontrolled vocabulary: a

folksonomy. In BibSonomy, the folksonomy evolves from the participation of research groups, learning communities and individual users, organizing their information needs.

CiteULike is based on the principle of social bookmarking and is aimed to promote and to develop the sharing of scientific references amongst researchers. In the same way that it is possible to catalog web pages (with Furl and del.icio.us) or photographs (with Flickr), scientists can share information on academic papers with specific tools developed for that purpose. When browsing issues of research journals, small scripts stored in bookmarks (bookmarklets) allow to import articles from repositories like PubMed. Then, the system attempts to determine the article metadata (title, authors, journal name, etc) automatically. Users can organize their libraries with freely chosen tags which produces a folksonomy of academic interests

3.3. Citation indexing initiatives

Some websites offer citation indexes of scientific publications. The most known are CiteBase, Scopus, the Web Citation Index and Scholar Google.

Citebase Search is a semi-autonomous citation index for online research literature. It harvests pre- and post-prints from archives that use the “Open Archives Initiative Protocol for Metadata Harvesting” (OAI-PMH). The website parses and links references and indexes of the found material. Currently the system is under test.

Scopus is an Elsevier owned citation index. Drawing from the website, Scopus is “the largest abstract and citation database of research literature and quality web sources covering over 16,000 peer-reviewed journals from more than 4,000 publishers. It covers over 16,000 peer-reviewed journals from more than 4,000 international publishers. Furthermore, it covers the following subject areas: Life Sciences, Health Sciences, Physical Sciences and Social Sciences. Additionally, Scopus also covers 431 million quality web sources, including 23 million patents. Web sources are searched via Scirus (www.Scirus.com is the Elsevier owned science-specific search), and include author homepages, university sites and resources such as the preprint servers CogPrints and ArXiv.org, and OAI compliant resources.

Web Citation Index is a Thomson Scientific owned multidisciplinary citation index of scholarly content from institutional and subject-based repositories such as preprints, technical reports, dissertations, proceedings, and other gray literature. It was launched in 2005. Thomson Scientific content editors select Web repositories which are deemed scholarly

relevant. In the opinion of the firm, this approach ensures the Web Citation Index delivers only the highest-quality, most-relevant content.

Google Scholar is a citation indexing services that searches among many disciplines and sources: peer-reviewed papers, theses, books, abstracts and articles, from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations.

Mendeley is a free desktop and web solution designed for managing and sharing research papers, discovering research data and collaborating online. It combines Mendeley Desktop, an easy-to-use PDF and reference management application (available for Windows, Mac and Linux) with Mendeley Web, an online social network for researchers.

3.4. Scientific blogs and wikis

The set of initiatives within this category allow researchers to write posts in blogs on specific scientific topics. The most famous initiatives are ScienceBlog, ResearchBlogging.

ScienceBlog Launched in January 2006, ScienceBlogs is an initiatives of the famous journal Science that allows to a limited number of selected researchers (selected by the website team) to open a blog on a specific topic. At the time of writing this article, 78 blogs were open. The indention of the team is quite appealing: “We believe in providing our bloggers with the freedom to exercise their own editorial and creative instincts. We do not edit their work and we do not tell them what to write about. We believe that science literacy is a pre-condition for progress in the 21st century. At a time when public interest in science is high but public understanding of science remains weak, we have set out to create innovative media ventures to improve science literacy and to advance global science culture.”

ResearchBlogging is a weblog platform system which hosts hundreds of blogs providing academic commentaries of peer reviewed research. The initiative is meant to contrast the dominance of general press release and news reports in the communication of science improvements and findings through the alternative take represented by a system of blogs, mostly written by experts in their field, offering a more thoughtful and balanced review of published research. As specifically mentioned by the website, this initiative is intended for those that are “tired of "science by press release".

Sci-mate ... sharing ideas is a wiki system that merely uses wiki technology to assist cooperative writing processes. It is a highly developed pre-publishing environment, that has been further developed to simplify editing, record and display authorship information, assign relevant features for copyright protection, and include built-in processes for peer-review. No training is required because the software itself manages most of these issues without the need to train users.

Rosetta commons user support wiki provides user support for the Rosetta Software Suite for macromolecular modeling. Registered wiki users are encouraged to contribute to the freely available online manual and help support other users through the Rosetta Forums.

3.5. Enhanced and recent initiatives

Some recent initiatives integrate many of the services as presented until now.

2Collab is one of the biggest enhanced initiatives launched in 2007 by Elsevier initiative. The website is an “online collaboration tool that enables to share, connect and discuss relevant research with peers”. According to the website *2collab* combines three key elements in one integrated platform: online bookmarking and reference management, group management (to work with collaborators) and networking (to find, evaluate, and initiate contact with other experts).

Nature Network is a professional networking website for scientists. It is provided by Nature Publishing Group (NPG), the publisher of the famous scientific journal Nature. The website offers various services ranging from organizing personal/group pages and blogs, interacting in forums, managing a personal research network, posting or browsing job market offers, and many more.

Liquidpub is a European funded project that tries to “reinvent” the whole scientific knowledge production and dissemination process on the basis of the idea that scientific knowledge should be considered as an evolutionary, collaborative, multi-faceted object that can be composed and consumed at different levels of detail. While the current approach encourages authors to write many (possibly incremental) papers to get more “tokens of credit”, generating often unnecessary dissemination overhead for themselves and for the community of reviewers the Liquidpub project proposes a paradigm shift in the way scientific knowledge is created, disseminated, evaluated and maintained. This shift is enabled by the notion of Liquid Publications, which are evolutionary, collaborative, and composable scientific contributions.

Many Liquid Publication concepts are based on a parallel between scientific knowledge artifacts and software artifacts, and hence on lessons learned in (agile, collaborative, open source) software development, as well as on lessons learned from Web 2.0 in terms of collaborative evaluation of knowledge artefacts.

4. DISCUSSION: TOWARDS A WEB 2.0 PUBLISHER

The above described Web 2.0 initiatives present a huge variety in terms of services provided, stakeholders involved and technologies implemented. Some initiatives are very similar to standard journals owned by publishers which tend to improve traditional services. Some others are very close to categorization systems (e.g. indexing systems, taxonomies, search engines) managed by actors that apparently have no big roles in the publishing sector.

In this section we aim at exploring how and to what extent these initiatives might challenge the whole sector and the services provided by traditional publishers. We also briefly describe the revenue models of the most important initiatives to shed light on the economic sustainability of such services.

4.1. The impact of Web 2.0 initiative to publisher's services

The analysis of Web 2.0 initiatives allows researchers to state that some of the traditional services offered by publishers (described in paragraph 2) are shaped and improved by these initiatives. As described in Table 2, *registration, formalization and preservation services* are slightly modified by these initiatives. In particular:

Registration services are not really being shaped by the new initiatives, but in the next future blogs and wikis applications, might raise many intellectual property issues. The scientific community is debating on intellectual property rights, creative commons licenses, and other protection mechanism of author/publisher intellectual property.

Certification services are shaped by several open and collaborative peer review systems. As showed in the previous analysis, several Web 2.0 journals are already testing these systems, attempting to improve their impact factor and their certification process by avoiding gender, race and other cultural biases. Sequential (“waterfall” style) review processes, also, might disappear in favor of a more collaborative process of co-writing and co-participation. Such

initiatives are carried on by well recognized journals such as Biology Direct (impact factor: 3.29) and ACP (impact factor 2007: 4.86).

Table 2: Traditional publishing services and Web 2.0 add-ons.

<i>Services</i>	<i>New applications</i>	<i>Real case studies</i>
<i>Registration</i>	---	---
<i>Certification</i>	Open, tier-based, 2 stage and open post publication peer review, blogs, forums	Nature Precedings, Biology Direct, ACP, PLoS One, Philica
<i>Formalisation</i>	---	---
<i>Dissemination</i>	Blog, Social Network, Forum, Wiki, Podcast, RSS, multimedia sharing	Nature Network, ScienceBlog, ResearchBlogging, CiteULike, 2Collab, BibSonomy
<i>Preservation</i>	---	---
<i>Use</i>	Citation managers, taxonomies, semantic search engines	Web of Science, Google Scholar

Formalisation services are slightly being affected by Web 2.0 as current computer based technologies already help in findings errors in documents. Due to the digitalization, authors are asked to produce the camera ready of their reports, rising their discontent with publishing services.

Dissemination services are being shaped by researchers' web pages, blogs, social networks and bookmarking systems. Through their web pages, authors publish their papers on line and enable the download of publications in order to improve the dissemination of their work. Also, focused social networks (Nature Network and 2collab), as well as social bookmarking initiatives (CiteULike, BibSonomy) might help finding journals and scientific contributions. While such initiatives do not have any (at the moment) role in evaluating the careers of researchers, some departments are taking into account the number of citations to distribute funds. Finally initiatives such as ScienceBlog and ResearchBlogging might disseminate scientific content improving the visibility of excellent researchers.

Preservation services seem to be slightly affected by Web 2.0 initiatives. It is still unclear how data managed by private institutions (Nature Network, 2collab) should be preserved over time and with what limitations and accessibility.

Use services and the fruition of a growing amount of knowledge and scientific contributions might be facilitated by social bookmarking systems and *ad hoc* search engines. Indeed, the web is already helping people “jumping” into papers and scientific documents they might be interested in.

4.2. The issue of sustainability

In order to understand the sustainability of Web 2.0 initiatives, their revenue models should be analyzed. In this paragraph the revenue models of both the more traditional services (journals with open peer review) and other initiatives are analyzed (Table 3). The study is quite superficial as the economic impact of the Internet on the publishing sector is still being unexplored (VV.AA., 2006; Wang and Chan, 2004; Weill et al., 2006; Featherstone and Ellis, 2005). As described in paragraph 3.1, most of the proposed journal initiatives are based on public funds, other use open access models, while only one seems to be a full cost for the firm. Note that no big scientific publisher (apart for Nature) funds these experiments, this means perhaps that they are waiting for future evolutions (Milmo, Dan, 2006).

Table 3. Business models of open review journals

<i>Journals</i>	<i>Business model</i>
<i>arXiv</i>	Subsides
<i>Nature Precedings</i>	Covered by means of other initiatives of the firm
<i>PLoS One</i>	Author pays
<i>ACP</i>	Author pays
<i>BiologyDirect</i>	Author pays
<i>JIME</i>	Subsides
<i>Philica</i>	Subsides

The other Web 2.0 initiatives described in paragraph 3 contrasting situations are showed.

Two social bookmarking systems (paragraph 3.2) out of three (Connotea and Citeulike) are recently bought by multinational publishers. It seems they do not embed any revenue model and their total cost of ownership appears to be completely covered by publishers. This is the same situation BibSonomy encountered. In that case, the Knowledge and Data Engineering Group of the University of Kassel is funding the initiative.

The four citation indexing websites under investigation (paragraph 3.3) are private initiatives, but two out of four initiatives are available on subscription only. Google Scholar is for free while Citebase Search is under testing and it is not clear whether users should pay for its services.

The two initiatives which fall in the Blogs and Wiki category (paragraph 3.4) are for free and are funded by private firms. ResearchBlogging sustain itself by means of advertisements while ScienceBlog does not have any clear revenue model.

Finally, only one of the enhanced initiatives described in paragraph 3.5 is working. 2collab is a free service supported by Elsevier and developed by the teams behind ScienceDirect and Scopus. The other two experiments will provide concrete solutions only in few years.

Table 4. Revenue models of Web 2.0 initiatives

<i>Web 2.0 initiatives</i>	<i>Revenue Model</i>
<i>Social Networks and researchers' blogs</i>	Covered by means of other initiatives of the firms (e.g. Nature Network) Public subsidies
<i>Wikis</i>	Users pay (Word) Open software (for free)
<i>Collaborative tagging and social bookmarking</i>	Covered by means of other initiatives of the firms
<i>Citation indexing initiatives</i>	Users pay (e.g. Thompson) Covered by means of other initiatives of the firm (e.g. Google)
<i>Enhanced initiatives</i>	Still at an experimental phase

5. CONCLUSIONS AND FUTURE WORK

As explained in many researches, the Internet, the Web 1.0 and the Web 2.0 inspire a paradigmatic shift in the scientific publishing industry. New initiatives are experimenting innovative services stressing publishers to improve their business models and revenue systems.

From our analysis it emerged that Web 2.0 initiatives are sparse, barely integrated and are provided by actors who might appeared in the scientific publishing market by chance. Although some of the entering actors are not directly involved in the publishing activities, they provide innovative services to both libraries and researchers (e.g. Google and University of Kassel). For instance Google Scholar and the Google Book Search Library Project will enable libraries to make available more significant portions of its extraordinary archival and special collections to scholars and researchers worldwide in ways that will ultimately change the nature of scholarship (Columbia, 2007).

In addition, no player (neither traditional publishers nor new actors) offer a full integrated and revolutionary system of registration, certification, formalization, dissemination, preservation and use of scientific knowledge. In the past, more than 100 society and university not for profit publishers have invested millions of dollars, creating unique and powerful online services (<http://www.dcprinciples.org/statement.pdf>). They might now provide innovative and competitive services that enable an integrated publishing system. In both Europe and USA, two initiatives aimed at integrating publishing services have been carried on: the Elsevier initiative “2collab” and the European Project LiquidpPub.

Finally, according to the row analysis of revenue models it emerges that many initiatives are based on public funds and are not sustainable in the short time.

This analysis, far from being exhaustive, is a first attempt to assess the impact of wide and heterogeneous set of Web 2.0 initiatives in the scientific publishing market. Future work in this direction is needed to strengthen the initial findings described in this paper. A more complete framework of analysis should be provided and a more complete and representative set of case studies should be defined and studied.

Future work might also identify the very innovative services that might radically change the publisher’s business models and systems of revenue.

REFERENCES

- Anderson, N.** 2006, "Tim Berners-Lee on Web 2.0: "nobody even knows what it means"". *Arstechnica.com*. Link: <http://arstechnica.com/news.ars/post/20060901-7650.html>. Retrieved on April, 10 2009
- Columbia,** 2007. "*Columbia University Libraries Becomes Newest Partner in Google Book Search Library Project*". Retrieved on April, 10 2009
Link: <http://www.columbia.edu/cu/lweb/news/libraries/2007/2007-12-13.google.html>
- ECDGR,** 2006, *Study on the economic and technical evolution of the scientific publication markets in Europe*. European Commission Directorate General for Research, Accessed: December 01, 2008, Retrieved on April, 10 2009
Link: http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf
- Enserink, Martin,** 2007, "European Union Steps Back From Open-Access Leap", *Science*, **315**: 1065.
- Featherstone and Ellis,** 2005, "*The Abundance of Simple Business Models on the World Wide Web*". Retrieved on April, 10 2009.
Link: <http://ausweb.scu.edu.au/aw05/papers/refereed/featherstone/paper.html>
- Guédon, Jean Claude,** 2001, In Oldenburg's Long Shadow : Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing, Association of Research Libraries.
- Houghton, John,** 2005, *Digital Broadband Content: Scientific Publishing*. Directorate For Science, Technology And Industry, OECD, Retrieved on December 2, 2008.
Link: <http://www.oecd.org/dataoecd/42/12/35393145.pdf>
- House of Commons,** 2004, *Scientific Publications: Free for All?*, The Science and Technology Committee. Accessed: December 01, 2008, Link: <http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39902.htm>
- Madras, Giridhar,** 2008, "Scientific publishing: Rising cost of monopolies", *Current Science*, **95**(2): 163.
- McAfee, Andrew,** 2006, "Enterprise 2.0: The Dawn of Emergent Collaboration", *MIT Sloan Management Review*. 47(3): 21-28.

Milmo, Dan, 2006, "Publishers watch in fear as a new world comes into view", *The Guardian*, April 19, 2006.

Link: <http://www.guardian.co.uk/technology/2006/apr/19/news.science1>

STM, 2008, *An Overview of Scientific, Technical and Medical Publishing and the Value it adds to Research Outputs*. Position Paper on Scientific, Technical and Medical Publishing, International Association of STM Publishers. Retrieved on April, 10 2009

Link: <http://www.stm-assoc.org/documents-statements-public-co/2008-04%20Overview%20of%20STM%20Publishing%20%20Value%20to%20Research.pdf>

VV.AA. 2006, *Assessing the impact of open access. Preliminary findings from Oxford Journals*, Oxford University Press. Accessed: December 2, 2008.

Link: http://www.oxfordjournals.org/news/oa_report.pdf

Wang, C.-P. and Chan, K. (2003). *Analyzing the taxonomy of Internet business models using graphs*.

Weill et al., 2006, "Do Some Business Models Perform Better than Others? A Study of the 1000 Largest US Firms", *MIT Center for Coordination Science Working Paper No. 226*.

Link: <http://ccs.mit.edu/papers/pdf/wp226.pdf> Retrieved on April, 10 2009

Wellcome Trust, 2003, *Costs and business models in scientific research publishing*.

Accessed: December 2, 2008.

Link: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtd003184.pdf

WEB REFERENCES

Journals with open and online peer review processes

<http://www.ArXiv.org>

<http://precedings.nature.com>

<http://www.plosone.org>

<http://www.atmospheric-chemistry-and-physics.net>

<http://www-jime.open.ac.uk>

<http://www.biology-direct.com>

<http://philica.com>

Social bookmarking systems

<http://www.bibsonomy.org/>

<http://www.citeulike.org>

<http://www.connotea.org/>

Citation indexing initiatives

<http://www.citebase.org>

<http://www.scopus.com>

<http://scientific.thomson.com/press/2005/8298416/>

<http://scholar.google.com>

Blogging and wikis

<http://www.scienceblog.com/>

<http://researchblogging.org/>