Scholarly communication and open access: research communities and their publishing patterns.

Valentina Comba¹, Head of the Digital Library Project

MariaLaura Vignocchi¹, Head Librarian

¹ Centro Inter-Bibliotecario, Università di Bologna viale Filopanti, 7 40126 Bologna, Italy

Valentina Comba: comba@cib.unibo.it

MariaLaura Vignocchi: vignocchi@cib.unibo.it
Abstract

**Introduction and background:** At the time of the Budapest Manifesto, self-archiving supporters looked like a revolutionary, "anti-commercial publishers" movement. After some years debate and technological innovation in scientific e-publishing, antagonist positions are able to compromise and consider the tradeoffs. In this context a new attention is being given to authors and their publishing practices.

**Objectives:** Determine what is changing in the authors' attitude towards institutional or disciplinary repositories, and peer-reviewed open access journals, and how authors of different research communities consider OA.

**Data Sources:** Review of current literature.

**Results.** Biomedical authors behave differently from astronomers, computer scientists and mathematicians, who have been using open archives for such a long time. Moreover, other factors affect the publishing trend: the role of new OA journals in evaluation processes and the evolution of bibliometric indicators.

**Conclusions:** Librarians should gain an understanding of differences across scholar community communication patterns to innovate and promote new services effectively. Librarians may also contribute to promote OA journals and reform evaluation criteria while still playing a key role in institutional repository management and users’ training.
INTRODUCTION AND BACKGROUND

The Budapest Manifesto has been the first of a series of seminal statements which contributed to formalize the open access movement, stimulate the debate and obtain public consensus. Those who took part in the event (notably Stevan Harnad, Peter Suber, Fred Friend, Jean Claude Guedon) in December 2001 had been involved in many of the ongoing initiatives which looked at an open access to scholarly communication as a new opportunity for humanity made possible thanks to technological innovations. The Manifesto recommended the strategies to attain an open access to literature by self-archiving and publishing on OA journals [1]. The Budapest Initiative was followed by the Bethesda Statement at the Howard Hughes Medical Institute in April 2003 [2]. In October 2003, the Berlin Declaration on Open Access to Knowledge in the Sciences and the Humanities came to include the humanities in the debate which had been so far dominated by STM exponents [3].

Librarians had been denouncing the dysfunctions of the scholarly publication system since the early 1970s. As Guedon has observed: “Libraries were the first to feel the financial pinch of the new business applied to scholarly journals; they were also the first to document its existence with care” [4]. First, the skyrocketing prices of print serial subscriptions growing much faster than dropping or static library budget provisions barely sufficient to renew core titles, then the licensing system for access to e-published journals and the big deal bundling policy of commercial giants have put the library mission in jeopardy. It is becoming more and more difficult – even to well-funded institutions - to provide their users with all the relevant literature they need for their work. Moreover licences to digital resources generally impose access restrictions which exclude non-institutional readers.

The programs and the activities of Sparc <http://www.arl.org/sparc/> launched by the Association of Research Libraries in 1998 and its European twin initiative, Sparc Europe <http://www.sparceurope.org/> best embody and represent the librarians’ efforts to contribute to change the present commercial publishing system, by supporting sustainable and more efficient alternatives of communication. One of the most convincing arguments with institutional and funding stakeholders against the dysfunctions of the commercial publishing system has been that of the triple payment. The House of Commons Committee on Science and Technology describes the phenomenon as follows: “Public money is used at three stages in the publishing process: to fund the research project, to pay the salaries of academics who carry out peer review for no extra payment, to fund libraries to purchase scientific publications” [5]. The Committee also adds that “Research Councils UK were concerned that the output from publicly funded research is handed free of charge to commercial organizations that appear to
make it more difficult to gain access to publications derived from the same research” [5]. The same concern has encouraged funding agencies such as NIH to adopt their pro-open access policies [6].

Until recently the debate was very harsh in tone especially when opposing the open access advocates to commercial publishers. The main topics under discussion were: the quality and the certification of research outputs disseminated through authors’ self-archiving in disciplinary or institutional repositories; the dubious economic sustainability of open access models; copyright policies which obliged authors to transfer all their rights to the commercial publishers and denied self-archiving.

However, the first analyses and evaluation of actual open access projects and achievements have contributed to what can be defined an assessment of open access impact, allowing a better definition of strengths and weaknesses and preparing for the way forward. As Peter Suber has acutely observed there has been “a shift in tone” [7].

First of all, open access has started to prove its positive influence on research impact. Indeed, free on-line availability of scientific articles seems to determine higher citation and usage rates.

Another argument which is contributing to a positive evaluation of open access comes from recent analysis of viable and sustainable economic models to open access publishing. Economic sustainability has always been a primary concern of the Budapest OA Initiative. The Open Society Institute has recently published a guide that provides practical evidence that learned societies and, generally, small publishers can consider the open access options (both the “straightforward” solution and the hybrid transitional form) as a possible future for their journals without significant financial risks [8]. Despite criticism the author-pay models may prove to become a solid business practice allowing OA journals to overcome the precarious dimension of voluntary work and establish credibility in terms of economic sustainability and quality.

Finally, open access is being considered a measure against increasing digital and information divide which create a chasm between wealthier countries and poorer ones. Subscriptions and licenses to scientific publications are too expensive for institutions in developing countries. Open access to scientific information may become a solution in the long-term. International aid projects like HINARI <http://www.who.int/hinari/en/>, AGORA <http://www.aginternetwork.org/en/> and PERI <http://www.inasp.info/peri/index.shtml> represent what can be done now by seeking the collaboration of commercial publishers.

Yet, despite worldwide advocacy and the recommendations of Councils and Committees, despite the fact that many publishers have already “gone green” [9], authors seem to linger, dwelling upon old and well-established practices. They usually adhere to OA on a principle basis but many of them
move objections and harbor fear and doubts. This fact calls for a new awareness of scholarly communities social aspects.

LITERATURE REVIEW

The scholars’ and scientists’ perspective has been brilliantly voiced by Guedon who has revealed their schizophrenic and time-morphing nature. As readers during the information gathering and research phase they may take advantage of a series of sources and media. As authors they need both for themselves and the literature they quote the “brand” of a quality refereed journal. Moreover, as they advance in their academic career, they may be asked to sit in editorial boards of commercial IF journals. Their role of gatekeeper and guarantor of certified quality research for subscription-based journals may eventually conflict with an unbiased evaluation of open access options [4].
The necessity of gathering as much information as possible upon potential markets and customers stands at the origin of any efficacious marketing campaign and always precedes the launch of a new product. It is not surprising that both open access supporters and publishers have begun commissioning market research [10] [11]. These surveys provide important aggregated data on scholarly authors’ attitude toward the journal system and alternative communication options.
Authors of different research communities, however, consider OA in different ways. The necessity to understand the peculiar differences across the various disciplinary groups has also emerged from the latest author surveys carried out by the Deutsche Forschungsgemeinschaft and by Swan and Brown that show data distribution according to disciplinary communities [12] [13].
The librarians involved in the promotion of new e-publishing services are becoming aware of these social differences. Many librarians working in universities and research institutions which have recently set up an institutional repository have probably experienced the patchy success of their informative and promotional campaigns supporting institutional open archives. Setting up good products and providing good services does not seem to be enough to induce all academic authors to use them.
A key factor for marketing the new e-publishing services and promote open access within our institutions may be to approach the different communities in the right way - that is trying to speak their jargons, promoting the options that most suit them, focusing the issues that most matter to them, such as research impact and evaluation, stressing the different tradeoffs. A more effective planning of a promotional initiative is likely to start from a deeper understanding of the different publishing and communication patterns in the different scientific communities.
We have carried out an extensive literature review about trends in scholarly communication in communities as diverse as astronomers, computer
scientists, mathematicians and biomedical authors in order to determine their publishing patterns and rates of acceptance of open access options.

Astronomers are quite a small community presenting few sub-disciplinary and sub-cultural differences. Studies of demographic trends shows that articles, especially observational or laboratory ones, have more than one or two authors [14]. Astronomers normally use a twofold channel of communication, posting the pre-print to a disciplinary open archive - Astro-ph the astronomy subset of ArXiv preprint server and then, submitting it for publication to peer-reviewed journals.

They rely on an extremely efficacious information system. About 50 or 60 refereed journals certify the quality of published research that is strictly controlled by professional associations and research community. Access to research literature is mostly achieved through NASA ADS (Astrophysics Data System) a free on-line bibliographic database that offers abstracts, lists of references, article citations, citations rates, and a powerful interlinking system that connects each bibliographic record to all other relevant resources included full-text versions of the article - even its pre-print on Astro-ph.

Studies on the reading patterns of astronomers members of the American Astronomical Society carried out by Carol Tenopir and Donald King show that astronomers like other scientists “continue to read scholarly journal extensively - perhaps are actually increasing their amount of reading” and still need the certification of peer-reviewed journals [15]. Yet, new patterns are emerging and more preprints or eprints are being cited in primary research works [16]. Recent articles have documented the major impact of papers available as preprints in Astro-ph server compared to the impact of published articles not freely available on-line. Schwarz and Kennicutt carried out a study of publication, preprint posting, and citation patterns of the Astrophysical Journal (ApJ), one of the major journals in the field [14]. Statistics shows that ApJ articles available in preprint server are cited more than twice as often as those that are not. Metcalfe comes roughly to the same conclusion for papers published in a wider selection of astronomical journals suggesting that preprint server Astro-ph has become an extremely efficacious means for rapid dissemination of research and current awareness in the field, greatly enhancing single articles or journals impact and visibility. Data regarding conference proceedings even though posted to preprint servers show a very low citation rate showing that publication in a refereed journal with an IF remains the primary determinant of the impact of a paper for researchers in astronomical field [17].

Computer scientists are a larger community than astronomers. Due to the way work is conducted in this specific discipline, articles are usually single-authored. Computer scientists normally present their papers in a conference and later publish them in the proceedings or in high quality peer reviewed journals (i.e. ACM, IEEE). As Lawrence points out “in computer science
conference articles are typically formal publications and are often more prestigious than journal articles with acceptance rate at some conferences below 10 %” [18]. A substantial percentage of literature is free on-line, available through disciplinary preprints servers such as NCSTRL, CoRR (Computing Research Repository) or the authors’ personal web pages. Google or other specialized search engines such as ResearchIndex (formerly Citeseer) help locate the papers on the web.

Research evaluation takes into account number of citations provided by services such as ResearchIndex and OpCit rather than IF. Lawrence’s seminal paper shows the results of a statistical study on conference articles in computer sciences that reveals a positive correlation between high citation rates of articles freely available on-line [18]. Mathematicians are a small community. As many have pointed out research in mathematics is often the work of a single genius independent from socio-economic, cultural and political context so papers are generally single-authored. Mathematicians were among the first to take advantage of the Internet for disseminating research results early on via Ftp sites and e-mail, later, through e-journals and e-books. The development of TeX, a computer program for typesetting documents, greatly contributed to the electronic conversion of mathematical and scientific literature in general.

Specialized journals and serials in mathematical fields are about 2000 and approximately 30% is open access. More traditional sub-fields that have developed efficient ways of communication in the early days of the Internet still use mailing lists and newsgroups.

Rate of acceptance of open access options is high among mathematicians [19]. They normally deposit their papers in preprints servers that are located at different sites (departmental or subject archives). Subject-oriented meta-search interfaces, “umbrella servers” such as Front End for the Mathematics and Mpress-MathNet.preprints, have been developed to achieve an aggregation of these distributed archives [20]. Mathematicians rely on a number of important on-line bibliographic databases: MathSciNet of the American Mathematical Society; MATH Zentralblatt, the most complete and longest running abstracting and reviewing service in pure and applied Mathematics in the world, containing more than 1.8 million entries dating back to 1931; MATHDI covers literature in mathematical education [20].

As regards research evaluation, ISI Impact Factor does not seem to be an efficient way of assessing the value of articles since citations may refer to very old works.

Biomedical authors are a very large international community that divides into many sub-disciplinary groups. They have to keep up with a great amount of professional and scholarly literature mostly in the form of journal articles. A
survey conducted by Tenopir and King on the reading patterns of medical faculty at the University of Tennessee Health Science Center shows that medical faculty read a great deal but spend less time on average on articles in comparison to scientists in other work-fields. “They value currency, but also need information to be digested and verified in a way to save them time” [21]. Biomedical authors seek to publish in peer-reviewed journals with high impact factor that certify the quality of articles.

Open archives and open access journals are quite a new, even if lively and rather paradigmatic reality. Pub Med Central <http://www.pubmedcentral.nih.gov/>, the free on-line digital archive of the NIH launched in 2000, hosts and provides free access to the full-text contents of about 187 peer-reviewed journals. The most notable and promising open access journals in biomedical area are private commercial ventures that use the author-pays economic model such Biomed Central <http://www.biomedcentral.com/> - an independent publishing house that provides immediate open access to 130 peer-reviewed biomedical research journals - or nonprofit publishing initiatives such as Plos Publications of the Public Library of Science <http://www.plos.org/>.

Funding agencies such as NIH [6] and Wellcome Trust [22] have recently announced policies in support to open access to biomedical research by requesting their grantees to deposit a copy of their manuscripts, once accepted for publication, to PubMed Central.

The results of a survey conducted on a sample of biomedical authors who have submitted to the BMJ shows that even if they are aware of and support the principles of open access publishing they nevertheless value journal quality and reputation when deciding where to submit their papers. The authors of the study conclude that new open access journals should do more to reassure authors of the quality of their journal [23].

DISCUSSION: THE RESEARCH EVALUATION FACTOR

The community profiles show different rates of acceptance and adoption of open access publishing options which reflect the way work is carried out in the different disciplinary fields and the priorities and the validation systems within the various scholarly communities. Preprints self-archiving in open access repositories or personal web pages may serve well for rapid dissemination of research and current awareness within the communities of astronomers, mathematicians and computer scientists. In the larger biomedical community quality certification of articles is more important than keeping up with cutting edge-research results.

The way research is validated is a major issue in biomedical field. PubMed Central is a free digital archive of journal articles which have been submitted to peer review process and meet high scientific and editorial standards.
Biomedical authors choose high quality impact-factored journals for their publications in order to obtain grants and funds. Although much criticized, peer review and impact factor remain the most widely used methods to certify quality research. Seglen points out: “Ideally, published scientific results should be scrutinized by true experts in the field and given scores for quality and quantity according to established rules” [24]. This may be possible on paper submission for publication, or, hopefully, when deciding scholars’ career advancement, even though more and more evidence is being collected showing that peer-review is far from being always objective and unbiased [25]. Yet, it cannot be a sustainable method when evaluating for example a department or a whole university in order to allocate funds and grants.

In the slippery fields of quality assessment it is not surprising that Eugene Garfield’s IF with its allure of objectivity has gained pervasive and perverse application. Journal Impact Factor, that is the average number of citations received by the articles in a journal within a limited period of time, has been used for evaluating research groups and individual scientists as a “career management tool” [4]. The limits of IF were brilliantly discussed by Seglen in 1997 - just to mention a few: articles citation rates determine the journal impact factor not the contrary; IF depends on research fields; IF applied to single research articles conceal their true value, that is, IF is not statistically representative of individual journal articles, the coverage of the database is not complete and appear to be dominated by American publications, self citation are not corrected…[24]. Yet, it is still nowadays one of the criteria indicated in guidelines for national research assessment.

In Italy, research evaluation in biomedical disciplines is even more complicated by the different evaluation methods used by the Ministries of Education (MIUR) and Health (Ministero della Salute). The latter relies heavily on IF to evaluate research quality in Health Institutions and Hospitals while impact factor is just one of the criteria indicated by CIVR (the Consultant Committee for Research Evaluation) in the Research Evaluation guidelines for 2004-2006 academic assessment exercise [26] [27].

Indeed, there are other measures that could and should be used besides IF. As Harnad has remarked: “The journal impact factor is just one of many potential predictive factors, each with its own weight, and each adding a certain amount to the accuracy of the prediction/evaluation” [28].

The Open Access Citation Index Group was founded at the 3rd OAI Workshop at CERN, Geneva, February 2004 with the aim to review existing evaluation methods, take into account linguistic areas neglected by ISI, and analyze OAI PMH applications on evaluation, by collecting metadata about citations and usage from all OAI compliant journals and archives [29]. The new measures made available automatically by scientometric engines such as Citebase <http://citebase.eprints.org/>, (a Web-based citation linking and impact-rank search service for OAI archives developed within the Open
Citation Project at Southampton) break new paths to evaluation and assessment. Harnad provides a list: citation and download counts for author, article, and journal, early days citation/download correlations, co-citation and co-download counts (who is jointly cited with whom and what is being downloaded with what), page rank algorithm (recursively weighing citation by the weight of the citing work) co-text semantic patterns (what and whose text patterns resemble the cited work), time-series analysis, hub/authority analysis (much-cited vs much citing works) and much more [28].

The group intends also to collect evidence of the higher impact of open access publications compared with access-tolls ones. The statistics on open access literature usage and impact show a correlation between high articles citation rates and their free availability online and have begun to prove how research may benefit from being open access [30]. Kristin Antelman writes: “since Steve Lawrence circulated his study on the impact of free online availability of computer science conference documents under the catchy title of “Online or Invisible” the notion that freely available papers have a greater research impact has taken hold” [31]. Antelman’s article offers a review of the existing studies on OA articles research impact and provides more data to the comprehension of a somehow still incomplete picture. Taking a sample of articles from high impact core ISI journals for four different disciplinary fields: mathematics, electrical and electronic engineering, political sciences and philosophy and comparing the citation rates of articles which resulted to have free online versions with those of non available online ones, she manages to demonstrate that “across a variety of disciplines open access articles have a greater research impact than articles that are not freely available” [31]. These examples, account for the vital role open-archives (OAI-PMH compliant) do already play at least in certain disciplinary fields, by enhancing visibility and impact of research literature and documentation even when they are published in access-tolls journals.

RESULTS
Recently, biomedical authors have been encouraged to choose open access by a series on incentives. Notably the policies in support of open access to public funded scientific research adopted by important funding agencies such as the National Institutes of Health and The Wellcome Trust. The Howard Hughes Medical Institute < http://www.hhmi.org/> provides financial support for the authors who decide to publish in an open access journal. JISC has announced major financial help to institutional repositories [32]. Another incentive may be considered recent evidence of major impact of open access published research [30] [31].

However, there are still obstacles. Generally, open access journals are new born if compared with well-known and established commercial core titles. They may be less than three years old and so they cannot possibly have an ISI Impact Factor. Given the importance that IF has in the research evaluation
system, although much contested and criticized, it is only natural that authors
do not want to take the chance and prefer submitting their works to high
profile commercial titles. Furthermore, in certain cases institutional pressure
may act as a barrier to open access publishing options. Leading and
influential senior scholars presiding over scientific editorial boards of
prestigious commercial journals may well act as gatekeepers and influence
the publishing choices of whole research groups. Finally, lack of awareness
often makes authors stick to the old tested publication practices that is
subscription-based, well-known, high impact journals.
What has not been discussed extensively - though a number of papers at the
9 ICML were on this subject [33] - are the advantages for the biomedical
community as a whole that open access would carry on. “Cultural apartheid”
is happening here and now in our so-called developed countries, where
different degrees of access to informative resources dig a gap between
researchers in well funded institutions which can afford to pay for the wide
portfolios of commercial publications and those whose right to access the
resources has been eroded by prohibitive prices and license limitations. A
typical example regards medical university researches and medical
practitioners serving in the National Health Service in periphery and rural
areas who cannot use the same informative resources.
Authors feel that their articles should be widely read, not only by cutting-edge
academic and research communities but also by medical practitioners who
have no access to a number of licensed databases and electronic journals.
One of the most important consequences of Evidence Based Medicine and
Evidence Based Health Care is the stringent necessity to access the most
relevant literature to make strategic choices: authors are getting aware of the
practitioners’ need to read leading articles - and not only to rely on literature
provided by pharmaceutical companies.

CONCLUSIONS
As many have written, libraries’ role has greatly changed in the digital age.
But, while, on one side, the information professionals’ mission is always to
provide high value information (and therefore it is mandatory to keep on
negotiating licenses and commercial electronic resources access), on the
other side, a new role has clearly emerged in these last years.
Librarians should gain a deeper understanding of communication patterns
across the different scholarly communities and offer a clear picture of the e-
publishing evolution through the different economic models and the new
opportunities that are emerging to provide greater impact to valuable
research publications. Therefore, authors should be informed by librarians
about new publishing opportunities in open access journals and policies of
commercial publishers about copyright and self-archiving [9]; new licenses
models like Creative Commons should be explained, compared with
traditional commercial copyright management and advantages highlighted. As
an active support to the free flow of scientific communication, libraries should create and manage institutional repositories and advocate an institutional policy to support the archives. Moreover, OAI compliant technology may be used at low cost to provide effective measures of the institutional research impact in order to give research agencies more concrete and unbiased data about biomedical research. Some minor tasks should not be neglected, like catalog open access peer reviewed journals, give visibility to institutional publications using OAI [34] and provide a metadata revision service when necessary. Finally, the cost of new services should be carefully monitored and consortia cooperation promoted, in order to avoid a conflict between the promotion of open access and the purchase of high value commercial literature.

ACKNOWLEDGMENTS

We are indebted to Prof. Daniele Dallacasa (Astronomy Dept. University of Bologna), Prof. Fabio Vitali (Computer Science Dept. University of Bologna) and Antonella De Robbio (Head of the Digital Library Project, University of Padova) for precious suggestions.

FOOTNOTES

* Based on a presentation at Open access: the option for the future!?, IFLA Satellite meeting no 17, Oslo, August 13, 2005.

† Selection was made to offer a pertinent overview at the IFLA Satellite meeting no 17 organized by the Science and Technology and by the Health and Biosciences Libraries Sections, Oslo, August 13, 2005.
REFERENCES

1. Budapest Open Access Initiative. [Homepage on the Internet; cited 5 Aug 2005].
   <http://www.soros.org/openaccess/index.shtml>


   <http://www.zim.mpq.de/openaccess-berlin/berlindeclaration.html>

   <http://www.arl.org/arl/proceedings/138/guedon.html>

   <http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39902.htm>

6. NIH public access. [Page on the Internet; cited 4 Nov 2005].
   <http://publicaccess.nih.gov/>


<http://www.wellcome.ac.uk/doc_WTX026830.html>

<http://bmj.bmjournals.com/cgi/content/full/bmj;330/7494/756>

<http://bmj.bmjournals.com/cgi/content/full/314/7079/497>


<http://www.ministerosalute.it/>

27. CIVR [homepage on the Internet; cited 4 Nov 2005].

<http://www.civr.it/>


<http://bmj.bmjournals.com/cgi/eletters/329/7471/0-h#80657>

29. Guedon JC. Open access citation index. [Web presentation; cited 4 Nov 2005].

<eprints.rclis.org/archive/00001003/02/Group_6.ppt>


<http://www.dlib.org/dlib/june04/harnad/06harnad.html>

32. JISC receives an extra £15m to support education and research. [Press release on the Internet; cited 4 Nov 2005].

<http://www.jisc.ac.uk/index.cfm?name=news_spendingreview>

33. 9 ICML [Homepage on the Internet] Salvador Bahia, 2005 Sept 20-23. [cited 9 Nov 2005].

<http://www.icml.org>


< http://www.dlib.org/dlib/october05/ozek/10ozek.html>