

A reflection on open-access, citation counts, and the future of scientific publishing

Xavier Bosch

Department of Internal Medicine, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Spain

Received: 2008.12.30, **Accepted:** 2009.01.28, **Published online:** 2009.03.31

© L. Hirschfeld Institute of Immunology and Experimental Therapy, Wrocław, Poland 2009

Corresponding author: Xavier Bosch, M.D. Ph.D., Department of Internal Medicine, Hospital Clínic, University of Barcelona, Villarroel 170, 08036-Barcelona, Spain, e-mail: xavbosch@clinic.ub.es

In the new era of wide dissemination of information, the opportunities to communicate and share the results of scientific research are immense. The consequences of this change are, however, somewhat unpredictable. For one thing, the more people have access to information and ideas, the more ideas we will have and, in theory, scientific knowledge may expand to levels never seen before. Nevertheless, one may wonder whether this expansion of knowledge will have beneficial or harmful effects. The widespread diffusion of erroneous ideas, whether from scientists or laymen, has the potential to do substantial damage due to the pervasive influence of the Internet. However, this same ubiquity will, surely, ensure that mistaken ideas or theories, at least those propounded by reputable scientists, will also be discussed and refuted where necessary.

In the field of scientific publishing, the scrutiny of ideas and results, i.e. peer review, appears to have worked quite well until the arrival of the World Wide Web. Moreover, science is largely a self-correcting process. It is also true that the mechanisms used by scientific journals to communicate new scientific data have contributed greatly to the present state of things. But it is also easy to imagine that scientific knowledge could spread much faster in the new information age than hitherto. It is thus understandable that the open-access movement has received such overwhelming support from many sectors. For scientists, open access implies not only immediate access to the full written article, but also to data and metadata that can be used for re-analysis. It also means major visibility; in other words, if someone is looking for a paper and does not have access through the Internet, they may simply disregard it unless they really need it.

However, it is also understandable that traditional publishers, who no doubt have been those mainly

responsible for and the guarantors of the transmission of most scientific discoveries for the betterment of society so far, are reluctant to accept the new system. Although the true cost of scientific publishing has not been determined, many journals are profitable to their owners. Journals owned by for-profit companies may well contend that they are conducting sound business, while revenue from publications is used by not-for-profit organizations to further advance the cause of the organization (DeAngelis and Musacchio 2004). In addition, publishers can always argue that they took the initial financial risk in starting journals and are entitled to a return on their investment.

Much has been written on the topic of open access over the last few years, especially since the launch of the Public Library of Science (PLOS) in October 2003. Richard Smith, former editor of the BMJ, has perhaps too radically compared the business of publishing to theft, where publishers make money by preventing ideas from breeding more ideas and stopping the expansion of knowledge (Smith 2007). Smith, who was BMJ's chief editor for 25 years, suggested that journals should bid for the best papers, rather than researchers transferring their copyright to journals for free. The traditional system of scholarly publishing certainly reveals some striking ironies. A reviewer who has worked to shape an article and has refereed the article for free may have no access to the final version of the published article if they or their institution has no subscription to the journal.

At the other end of the spectrum are those who argue that we are inundated by information but lack the time to evaluate it and that open access worsens this situation by opening the gates to increasing amounts of unfiltered information (Esposito 2007), which is, in my opinion, sometimes really junk information. There are also concerns about whether reputational incentives will

be sufficient to ensure quality maintenance in open-access journals (European Commission 2006). Yet if a journal does not have a reputation for high quality, fewer people will read it and fewer would-be authors will wish to publish in it. Open-access for-profit publisher BioMed Central (BMC) has allowed groups of researchers to start new open-access journals or convert existing journals to open access; editorial control remains with the researchers and the company provides them with a free publishing system (BioMed Central). Is it right to take such a risk or is it better to continue relying upon the traditional system with skilled editors? It is undeniable that the open-access movement has helped widen the spectrum of publications, with many new journals having been created. However, whether the quality of what is published in these new journals is sufficient is another matter.

We are all facing a profound crisis in the world of scientific publishing. With over 24,000 scholarly journals being published, and library budgets stagnating, if not decreasing, individual research institutions may no longer afford to subscribe to more than a fraction of journals (European Commission 2006). In parallel, open access is increasingly gaining ground. Although scientists have, in general, reacted with a certain conservatism and passivity to the movement (Gallagher 2007), submissions to open-access journals are increasing. However, the future financial viability of different journals is unclear. Thus PLoS has not yet reached self-sustainability and BMC may have broken even. In fact, Springer, the second largest publisher of science, technology, and medicine, recently acquired the latter. The open-access publisher Hindawi, in contrast, seems to be financially healthy, with article processing charges covering the company's costs (The Scholarly Publishing and Academic Resources Coalition 2007).

But what is the effect of free access to scientific literature on article citations? Results from recent studies are not very encouraging in this regard. Initially, the observed positive correlation between the open-access status of a given article and higher citation counts was interpreted as causal, since there was both an intuitive and sociological appeal to this mechanism. However, closer analyses have not confirmed any causal relationship, and have actually shown a more complex set of contributors to the effect.

The majority of studies that have tried to analyze this matter have failed to determine accurately the date of the earliest dissemination of each article and then to impose a defined citation window, which must be used if citation analysis of open-access status is to produce definitive results. Even the most robust methods developed to date are unable to show causality unequivocally nor can they generalize the effects observed at the author level or across a large number of varied disciplines (Craig et al. 2007).

In a forthcoming (to be published in 2009) study analyzing the citations to articles published in eleven biological and medical journals from 2003 to 2007 that

employed author-choice open-access models, only two journals showed positive and significant open-access effects after controlling for known explanatory predictors of citations. Analyzing all journals together, the authors found a small but significant increase in article citations of 17%. However, they concluded that the open-access citation advantage, widely promoted in the literature, is considerably overstated for biological and biomedical literature. They also found evidence suggesting that the open-access advantage is declining by about 7% per year, from 32% in 2004 to 11% in 2007 (Davis 2009).

A report published in November 2008 by the College of Management of Technology (Ecole Polytechnique Federale de Lausanne) analyzed a sample of 4388 biology papers published between May 2004 and March 2006 by the Proceedings of the National Academy of Sciences. This prestigious journal started to offer an open-access option to authors in May 2004 for a fee of US \$1000. This paper showed that at least part of the larger number of citations received by open-access papers was due to a self-selection effect rather than a diffusion effect and that the diffusion effect of open access was smaller than previously thought (Gaule and Maystre 2008).

In another study published recently by the BMJ (Davis et al. 2008), a randomized trial of open access, Davis and colleagues sampled papers due for publication in a group of physiology journals and randomly allocated them to either open-access or subscription-access publication. They found significantly higher online usage of open-access articles, but no significant difference in citation rates between the two groups in the first year after publication. Again, the study suggested that previous findings of a citation advantage from open access may have been the result of self-selection, with more highly citable articles being more likely to be published in open-access journals. Davis and colleagues' finding that open access provided no citation advantage despite increased readership may also be explained by the fact that journal readers who generate citations already have subscription access to journals (Davis et al. 2008; Godlee 2008).

Coming back to the current crisis of the scientific publishing system, a change in it is inevitable. Despite the legitimate interests of every traditional publisher, the model will continue to change so that information reaches everyone. The future model may be one that, while preserving the content of prestigious journals, would also allow immediate free access to all original research data. Some journals, such as Science, are so outstanding that they may safely offer a full open-access model already. Yet the financial uncertainties of such a sudden change mean that implementation of a mixed system, rather than a full transformation, would be more reasonable in the short or medium term. Under this transitory mixed or hybrid model, an author-pays mode would ensure the free dissemination of original research while a pay-to-view/subscription-based mode would

cover the rest of the content with added value, such as Reviews, Editorials, Perspectives, News, and Book Reviews. However, subscription fees should also decrease proportionally. Traditional journals could still produce, if they wish, a print version of the original papers and/or the full issue, for which interested parties would pay the required fee.

However, this is not enough. The new U.S. National Institutes of Health (NIH) mandatory policy might be transitional. Open access to articles between six (in the case of Wellcome Trust funded research) and twelve (NIH) months after publication is, in theory, completely useless, at least in some fields in which science advances incredibly rapidly. In practice, no scientist will wish to wait so long to read an article that may help advance his or her research. These policies may thus be regarded as too bland and aimed mainly at satisfying publishers. If the philosophy of open access is immediate access, why wait six or twelve months? If the overall model eventually proves successful, the natural evolution of these mandatory policies should be towards simultaneous free availability of original research in both the journal and the public repository. This natural evolution should also include the authors' keeping the copyright and being able to reuse their work as many times as they wish.

Finally, as far as the relationship between open-access papers and citation counts is concerned, there is a need for methodologically sound and well-interpreted studies of such relationship across diverse disciplines. Such studies need to ensure that the signature distorted distribution of citation patterns and artifacts such as selection bias and early view are accounted for. Truly randomized studies (with articles randomly selected from the same journal for open-access treatment) such as that by Davies et al. (Davis et al. 2008) may offer one approach, provided these can be managed practically and other factors can be controlled for (Craig et al. 2007).

REFERENCES

- BioMed Central. Independent, open access journals. <http://www.biomedcentral.com/independent/>
- Craig ID, Plume AM, McVeigh ME et al (2007) Do open access articles have greater citation impact? A critical review of the literature. *J Informetrics* 1:239–248
- Davis PM (2009) Author-choice open access publishing in the biological and medical literature: a citation analysis. *J Am Soc Inf Sci Technol* 60:3–8
- Davis PM, Lewenstein BV, Simon DH et al (2008) Open access publishing, article downloads, and citations: randomised controlled trial. *BMJ* 337:a568
- DeAngelis DC, Musacchio RA (2004) Access to JAMA. *JAMA* 291:370–371
- Esposito JJ (2007) Open access 2.0. *The Scientist*, November 1, 2007, <http://www.the-scientist.com/toc/2007/11/1/>
- European Commission (2006) Study on the economic and technical evolution of the scientific publication markets in Europe. January 2006. http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf
- Gallagher R (2007) OA, OK?. *The Scientist*, November 1, 2007, <http://www.the-scientist.com/toc/2007/11/1/>
- Gaule P, Maystre N (2008) Getting cited: does open access help? College of Management of Technology (Ecole Poly-technique Federale de Lausanne). November 2008. <http://ilemt.epfl.ch/repec/pdf/cemi-workingpaper-2008-007.pdf>
- Godlee F (2008) Open access to research. *BMJ*, 337:a1051
- Smith R (2007) The highly profitable but perhaps unethical business of publishing medical research. In: Smith R (ed) *The Trouble with Medical Journals*. The Royal Society of Medicine Press, pp 211–222
- The Scholarly Publishing and Academic Resources Coalition (2007) Matrix Analysis. November 2, 2007. <http://www.arl.org/sparc/meetings/ala07/matrixhtml.html>