Open Access:
Why should we have it?

Alma Swan
Key Perspectives Ltd, 48 Old Coach Road, Truro, TR3 6ET, United Kingdom
www.keyperspectives.co.uk

*This article derives from a presentation made at the meeting **"Zichtbaar** onderzoek. Kan Open Archives daarbij helpen?" / **Visible** research. Can OAI help? This was a "good practice" conference organised by AWI (Flemish Ministry for Economy, Enterprise, Science, Innovation and Foreign Trade) and VOWB (Flemish Organisation of Scientific Research Libraries) May 2006*

One frequently reads statements to the effect that Open Access is difficult to define or that it has many meanings. Whilst it is true that the term has a wide variety of applications in other settings, from the right to roam across the British countryside through systems for seeing your doctor to a kind of bone density test[a], in the scholarly communications sense it is actually rather easy to define what Open Access is. It is the free (gratis) online availability of the research results that scholars give away themselves (peer-reviewed journal articles and conference papers, mostly), provided by authors upon acceptance for publication and made permanently available without restrictions on use.

Open Access is not about the literature and research output from which scholars normally expect to derive some financial benefit, such as books and monographs that commonly pay authors a royalty on sales.: no-one is suggesting that the authors of these types of output should give them away, now or in times to come.

So having defined Open Access as free, immediate, permanent and unrestricted, let's move on to why we should have it. Certainly its introduction is causing all manner of upheaval, anxiety and argument, things we could all do without unless there are very persuasive reasons for backing the cause. What are these reasons? What is Open Access going to offer that is of sufficient benefit to make the struggle worthwhile?

I propose four main reasons as to why Open Access is beneficial for the way scholarly research is carried out and how its findings are used, and is thus incontrovertibly beneficial for human society as a result. I mention the latter because the stakeholders are, after all, not just the immediate players in the game: we all have stakes in there, too – researchers, research institutions, nations and global society as a whole. We all have an interest in the efficient and

[a] Examples courtesy of Peter Suber, whose daily trawl of the web for the term 'open access' returns him articles on over 40 topics where the term is in common use, and which he then has to filter. If you ever thought that putting together his daily blog on Open Access (http://www.earlham.edu/~peters/fos/fosblog.html) is straightforward and quick, think again.

effective progress of scholarly endeavour. The reasons I offer, then, for why Open Access is the way to go are these:

i) Open Access means there is greater visibility and accessibility, and thus impact from scholarly endeavour
ii) Open access means there is more rapid and more efficient progress of scholarly research
iii) Open Access means there can be better assessment, better monitoring and better management of science
iv) Open Access means that novel information can be created using new computational technologies

These are not just personal hunches. There is evidence for each, as I shall now go on to explain.

Open Access brings greater visibility and impact for research

Evidence is now accumulating that **open access increases the impact** of scientific work[1, 2 ,3, 4, 5]. Stevan Harnad's teams in Montreal and Southampton have carried out the most wide-ranging and extensive studies on this issue. Their robot crawls the Web, searching for scholarly articles that are openly accessible in full-text. Once articles are located, the number of citations to these articles are measured and compared to the number of citations to articles *in the same issue of the same journal* thus ensuring that like is not being compared to unlike. Comparing articles in different research fields, or between different journals, would be a very badly controlled experiment, but the methodology used here avoids this potential pitfall.

The data that have so far come out of this series of studies, which is ongoing, have demonstrated conclusively that open access doubles downloads and increases citations by an average of around 50% (this rate varies with discipline, from around 40% for biology to 250% for physics, so 50% is a conservative average figure)[6,7].

Given that, and since only 15% [b]of research around the world is currently open access, we can translate these findings about the loss of potential usage and impact (downloads and citations respectively) into figures that are meaningful in terms of the way research is funded. An example from my own country serves to show what I mean here. The current budget for the eight UK Research Councils is 3.5 billion GBP per annum. There is much *more* money pouring into research and development in the UK, of course, but for the purpose of my argument this particular example of public funding through the central funding bodies suffices.

[b] This is the average amount of Open Access material available, calculated by the robot used by Stevan Harnad's research groups in Montreal and Southampton to crawl the Web and locate full-text Open Access articles

If open access increases impact (citations) by an average of 50%, as Harnad's work shows, then **potential impact worth 1.49 billion GBP** is being lost every year if the output from the research funded by the UK Research Councils remains closed. A recent paper by economists Houghton and Sheehan has drawn similar conclusions[8].

Open access brings more rapid and more efficient progress for scholarly research

The high energy physics repository, arXiv, which has been in operation since 1991, provides the perfect experimental system for studying the deposition behaviour, usage patterns and impact of open access material. The repository contains around 400,000 documents, of which just over half are postprints, that is, they have been peer-reviewed[c].

Brody has looked at the pattern of citations to articles deposited in arXiv, specifically at the length of the delay between when an article is deposited and when it is cited, and has published the aggregated data for each year from 1991 to the present[9]. For simplicity, in Figure 1 below I have shown only the data for alternate years. These show that as more papers are deposited and more scientists use the repository, the time between an article being deposited and being cited has been shrinking dramatically, year upon year. This is important for research uptake and progress, because it means that in this area of research, where articles are made available at – or frequently before – publication, the research cycle is accelerating. The height of the curves in Figure 1 is not particularly significant because they simply show that the number of articles being deposited is growing each year. What *is* important is the *shape* of the curves. Those for earlier years show that it used to take a much longer time for new findings to be used and cited in further research, whereas for later years articles are being cited much earlier. Put simply, **the research cycle in high energy physics is approaching maximum efficiency** as a result of the early and free availability of articles that scientists in the field can use and build upon rapidly. Note that some articles are deposited in arXiv before publication (either at the time of acceptance by a journal or even before this, as a preprint, and this explains the negative number at the left end of the X-axis.

[c] Data obtained with the help of Dr Tim Brody

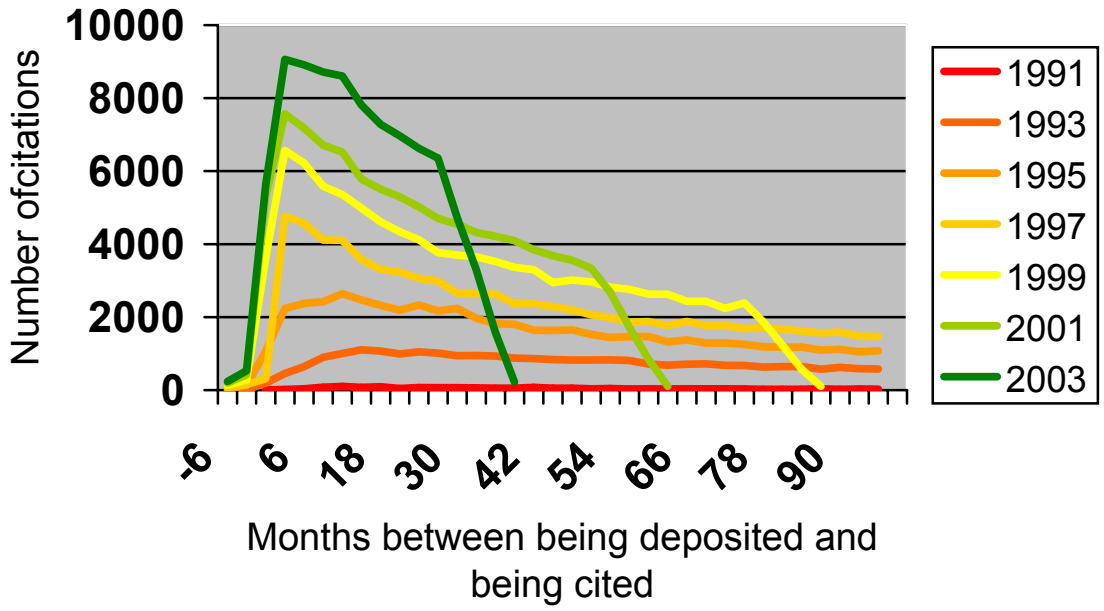

Figure 1: Time taken for articles in the arXiv database to be cited
(constructed from original data provided by Dr Tim Brody)

Open Access will enable better assessment, better monitoring and better management of science

Work that is now going on in the field of scientometrics (bibliometrics specifically applied to the scientific research literature) is pointing the way to what will be possible in future in terms of tracking the way the literature is used, how scholarly research effort is built upon, and how to identify effective science and scientists[10].

The citation-tracking software, Citebase[d], developed at Southampton University by Tim Brody, currently works on the UK mirror site of the arXiv repository (high energy physics) and some other open access article sources. It records the references each article cites and links these to the citing articles, thus mapping the complex web of citations within the bodies of literature in these collections. Using Citebase, it is possible to track how a field of research has developed, grown, split into sub-fields or declined. It is possible to work backwards to see where an idea first arose and who was responsible for it; it is possible to analyse who are the (highly cited) leading researchers in the field (considered to be 'authorities') and who cites them frequently; it is possible to see which articles are frequently – or always – cited alongside certain other articles; and it is possible to trace the development of ideas and theories, their growth rate, their maturation,

[d] Citebase: a citation-tracking tool for the scholarly literature www.citebase.org/help/

their directionality, the diversification of a field into daughter fields of research, and so forth.

Until the development of this type of analytical tool bibliometrics was something of an infant field with severe limitations on the methodologies that could be utilised, because the full-text of articles was simply not available for such tools to work upon. Now there are enormous possibilities and these will provide the means not only for researchers to better understand how their own work is being used and how their field is developing, but this also opens up a wealth of avenues of investigation for bibliometricians and for research funders, research managers and research planners to do their jobs much more effectively. ***These tools will enable us to measure, assess and manage scientific productivity and progress much better than is currently possible***, but they depend on having a critical mass of open access material on which to work.

Open Access will enable novel information to be created using new computational technologies

Alongside the bibliometrics opportunities described above, exciting new developments in text-mining and data-mining are beginning to show what can be done to create new, meaningful scientific information from existing, dispersed information using computer technologies[11,12, 13]. Research articles and accompanying data files can be searched, indexed and mined using semantic technologies to put together pieces of hitherto unrelated information that will further science and scholarship in ways that we have yet to begin imagining. These technologies are just in their infancy at the moment.

Real scientific advances will be made using them but to work effectively the technologies need to be applied to the full-text articles of the open access corpus: literature and data hidden behind journal or databank access restrictions are invisible to the computer tools that can do this work and so it is crucial that we free up the results of current research in order to generate the benefits that lie in wait. The longer we wait for open access to happen, the longer we delay the advantages to science and society that these technologies will bring.

Conclusion

These, then, are the reasons for which open access is worth the struggle. I have briefly described the tangible benefits for scholarship and society. There is much to be looked forward to and benefits for stakeholders of all kinds. But most important of all is that the payoff from our investment in science, technology and scholarship will be maximised. Society pays for research to be done, partly in the spirit of human curiosity about the world we live in, but also in the hope that tangible payoffs will be forthcoming. We pay up, and we do so expecting that the

results will be achieved as efficiently as possible. Every so often in the development of human societies a phase-shift occurs, after which things are quite changed and developments proceed at a new pace. The World Wide Web has brought such a phase-shift upon us, and it is now incumbent upon the research community to take advantage of this for the benefit of us all.

References

1. Kurtz, Michael. Restrictive access policies cut readership of electronic research journal articles by a factor of two. http://opcit.eprints.org/feb19oa/kurtz.pdf (accessed 30 October 2006)

2. Harnad, Stevan ; Brody, Tim Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine*, 2004, Vol 10 n°. 6, www.dlib.org/dlib/june04/harnad/06harnad.html (accessed 30 October 2006)

3. Antelman, Kristin Do open-access articles have a greater research impact? *College & Research Libraries,* 2005, Vol. 65 n°. 1, pp. 372-282.

4. Wren, Jonathan Open access and openly accessible: a study of scientific publications shared via the internet. *British Medical Journal* 2005 Vol. 330, p. 1128 (14 May), doi:10.1136/bmj.38422.611736.E0 (published 12 April 2005) http://bmj.bmjjournals.com/cgi/content/full/330/7500/1128?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&author1=wren&andorexactfulltext=and&searchid=1&FIRSTINDEX=0&sortspec=relevance&resourcetype=HWCIT (accessed 30 October 2006)

5. Eysenbach, Gunther Citation advantage of open access articles. *PLoS Biology* 2006, Vol. 4, n°. 5 http://biology.plosjournals.org/perlserv/?request=get-document&doi=10%2E1371%2Fjournal%2Epbio%2E0040157 (accessed 30 October 2006)

6. Hajjem, Chawki; Gingras, Yves; Brody, Tim; Carr, Leslie; and Harnad, Stevan. Open Access to Research Increases Citation Impact. 2005. Technical Report, Institut des sciences cognitives, Université du Québec à Montréal. http://eprints.ecs.soton.ac.uk/11687/ (accessed 30 October 2006)

7. Hajjem, Chawki; Harnad, Stevan; Gingras, Yves .Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. *IEEE Data Engineering Bulletin*, 2005, Vol. 28, n°**.** 4 pp. 39-47. http://eprints.ecs.soton.ac.uk/12906/ (accessed 30 October 2006)

8. Houghton, John & Sheehan, Peter. The Economic Impact of Enhanced Access to Research Findings, 2006 Centre for Strategic Economic Studies Victoria University. http://www.cfses.com/documents/wp23.pdf (accessed 30 October 2006)

9. Brody, Tim; Harnad, Stevan; Carr, Leslie. Earlier web usage statistics as predictors of later citation impact. *Journal of the American Association for Information Science and Technology* (JASIST), 2005, Vol. 57 **n°.** 8 pp. 1060-1072. http://eprints.ecs.soton.ac.uk/10713/01/timcorr.htm (accessed 30 October 2006)

10. Harnad, Stevan; Carr, Leslie; Brody, Tim; Oppenheim, Charles. Mandated online RAE CVs Linked to University Eprint Archives: Increasing the predictive power of the UK Research Assessment Exercise while making it cheaper and easier. *Ariadne*, 2003, Vol 35. http://www.ariadne.ac.uk/issue35/harnad/ (accessed 30 October 2006)

11. The NeuroCommons http://sciencecommons.org/projects/data/index.html (accessed 30 October 2006)

12. Murray-Rust Peter. Open source, open data and the science commons. From Peter Murray-Rust's blog 'A Scientist and the Web', 7th September 2006. http://wwmm.ch.cam.ac.uk/blogs/murrayrust/?p=15 (accessed 30 October 2006)

13. Lynch, Clifford. Open Computation: Beyond Human-Reader-Centric Views of Scholarly Literatures. In *Open Access: Key Strategic, Technical and Economic Aspects*, edited by Neil Jacobs. Chandos, Oxford, 2006. http://www.cni.org/staff/cliffpubs/OpenComputation.htm (accessed 30 October 2006)