Self-Archiving, Metrics, and Mandates

Stevan Harnad

Open access (OA) means free online access to published research articles. Some 2.5 million research articles are published every year in 25,000 peer-reviewed journals across all fields and all languages. The authors of those articles are employed and their research is funded so that it can be used, applied, and built on. The degree to which the research is used, applied, and built on is called its impact. The number of times an article is cited is one measure of impact.

Because researchers' salaries and funding depend on the *impact* of their research and because impact in turn depends on access, researchers have always wanted to maximize access to their work. Now the Web makes it possible for all researchers to "self-archive" their articles in institutional repositories (see *roar.eprints.org*).

Self-archived articles are preprints or postprints that the author deposits in an online repository and that are freely accessible. Thus, they do not *substitute* for the peer-reviewed journal articles, [1] but rather *supplement* the limited or expensive access that publishers provide (in much the way reprints were sent to requesters in the paper era). The OA versions come in a variety of forms. They can be the publisher's pdf; the author's revised, refereed, and accepted final draft; or an unrefereed preprint. Some have full references to the publisher's URLs and DOIs.

If maximal impact is the goal and if citations are one measure of impact, an important way to estimate the value of OA is to measure the increase in citations of articles that are made OA. A series of studies of citation counts across more than a dozen fields—beginning with computer science [2], then physics [3], then the biologic and social sciences and the humanities [4]—have consistently found that OA articles are cited 25% to over 250% more than non-OA articles. That is called the OA impact advantage. The accompanying figure shows this effect for a variety of fields. More detailed data by field can be found at *opcit.eprints.org/oacitation-biblio.html*.

The method is simple: The metadata on all the articles indexed by the ISI science and social-science indexes (on a licensed CD-ROM) are fed to a software robot that trawls the Web to try to find an OA version of each article. On the average, about 15% of articles are being self-archived today. Once the free versions are found, the logarithms of their citation counts are compared with those for non-OA articles in the same journal and year. The OA:non-OA citation ratio is the OA advantage. Some fields, such as chemistry, have low rates of self-archiving (the American Chemical Society is particularly opposed to OA), so OA and non-OA in such fields cannot now be compared. However, in other fields, such as physics, self-archiving has been extensive. For astrophysics papers posted as preprints in arXiv.org, Kurtz et al. [5] found that "the effect of e-printing on citation rates in astronomy and physics is significant."

One question is whether article-quality differences are a factor in OA–non-OA differences. Self-selection for quality is indeed one of five potential factors that contribute to the OA advantage: (1) *early advantage* (earlier OA, more citations), (2) *quality advantage* (the top 10% of articles benefit more from OA than the bottom 90%), (3) *use advantage* (more downloads of OA articles), (4) *competitive advantage*, and (5) *quality bias* (selectively making better articles OA). However, the last two effects vanish when all articles are self-archived, for instance, if mandates are put into place. Kurtz et al. [6] (in astrophysics) and Moed [7] (in condensed-matter physics) concluded that authors' selective archiving of their higher-quality papers is indeed one of the factors that influence whether they deposit them in the arXiv repository before publication in a journal. We have compared [8] the usual, spontaneous self-selected self-archiving with self-archiving mandated by authors' institutions. If the OA advantage were due all or mostly to self-selection (quality bias), the advantage should be smaller or nonexistent for mandated self-archiving, which reduces or eliminates self-selection bias, particu-

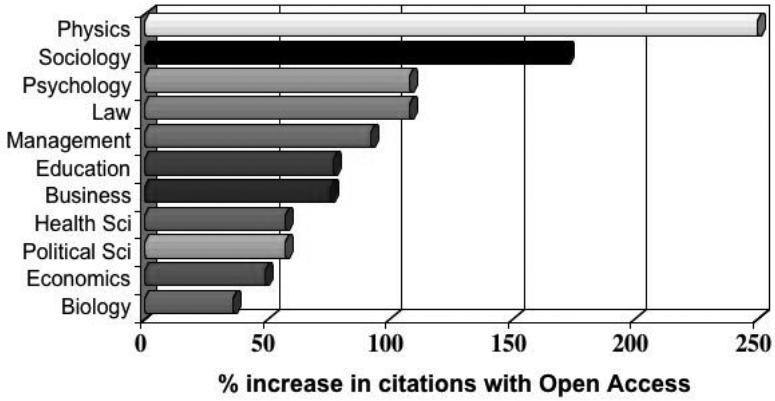
larly in institutions that have already approached 100% compliance, such as CERN. But there is no detectable difference in the OA advantage (for CERN or the other three mandating institutions: Queensland University of Technology, the University of Minho, and the University of Southampton), so the overall contribution of the quality bias is very small.

The effects of embargoing access for 6 to 12 months have not yet been estimated. It is hard to measure the amount of loss in use and citations when OA is delayed. It would no doubt vary among fields (some of which develop faster than others), but research [6] suggests that delay results in not just a temporary but a permanent loss in the research cycle: fewer accesses, fewer uses, fewer citations. Because the items just listed propagate in parallel, this means less productivity and progress.

Although examining surges at the end of an embargo is not the most effective or direct way of testing the OA advantage, Brody and others [9,10] have found indications of download increases when one item in arXiv is cited in another, newly deposited item in arXiv, and citation increases when an item is newly deposited. They have also found a correlation between early downloads and later citations [9] and shown that in physics, the interval between an item's first being deposited and its beginning to be cited keeps shrinking as self-archiving grows.

Despite the OA advantage and despite the link between impact and researchers' salaries and funding, only 15% of researchers are self-archiving spontaneously today. I have dubbed that paradox "Zeno's paralysis" (it has at least 34 easily remedied causes: *eprints.org/openaccess/self-faq/#32-worries*) [11]. Institutions and funders already mandate that their researchers must publish (or perish); they are now also beginning to mandate that they self-archive to maximize their research impact. Fourteen universities and research institutions and 17 research funders worldwide already mandate OA self-archiving, and several even bigger multi-institutional and national funding agency mandates have been proposed and are under consideration

Association of open access and citation rate



(Data: Brody and Harnad, 2004 [3]; Hajjem, Harnad, and Gingras, 2005 [4])

Figure. Open-access citation advantage: in all fields tested, open-access articles have a higher citation impact than non–open-access articles.

(*eprints.org/openaccess/policysignup/*).

The UK has the strongest momentum toward OA. The first and one of the most widely used (free) softwares for creating OA institutional repositories was developed in the UK (*eprints.org*, University of Southampton). The UK Parliamentary Select Committee was the first to recommend mandating OA self-archiving, and five of the seven UK research councils have already mandated it. In addition, the UK has a Research Assessment Exercise (RAE) in which the research impact of every department of every UK university is ranked by an assessment panel and each department is awarded substantial top-sliced research funding in proportion to its rank. The UK RAE panel rankings turned out to correlate highly with citation counts in most fields (for example, 0.91 in 1996 and 0.86 in 2001 in psychology) [12]. Panel rankings are now being converted to metrics.

Besides citation counts, OA will provide a rich spectrum of potential metrics, including download counts, download and citation growth and decay rates, book-citation counts (from Google Books and Google Scholar), and co-citations. OA versus non-OA download counts, however, are much harder to compare than citations. Studies are just beginning, and downloads need to be tested jointly with other potential metrics. In 2008, the RAE will conduct a parallel exercise—both metrics and panel rankings—in which the metrics can be validated and calibrated against the panel rankings, discipline by discipline. The outcome of the validation exercise can then be used to create research-impact metrics. A prototype scientometric engine, citebase (*citebase.eprints.org*), has already been developed that can apply the metrics not only to navigation and evaluation but as an incentive to motivate and accelerate OA self-archiving and OA self-archiving mandates worldwide. [13-15]

Further analyses will be needed to test and validate the data from the 2008 UK RAE. Once the metrics are validated field by field against the panel rankings, each with its own (beta) weights for each metric, then OA versus non-OA impact can be compared with the full metric equation and each of its validated components. Metric displays can then also be built into

the repository and harvesting software so that anyone can use OA metrics for evaluation and navigation, and authors can see directly the benefits conferred by OA. OA through self-archiving is optimal and inevitable for research, researchers, their institutions and funders, the vast research and development industry, and the taxpaying public that funds research. OA scientometrics is poised to usher in the OA era at long last.

References

1. Swan A. Open access self-archiving: an introduction. Technical Report, Joint Information Systems Committee, Higher Education Funding Council for England[Internet]. 2005 [cited 2007 Sep 5]. Available from: eprints.ecs.soton.ac.uk/11006/.

2. Lawrence S. Free online availability substantially increases a paper's impact. Nature [Internet]. 2001 [cited 2007 Sep 5];411:521. Available from: www.nature.com/nature/debates/e-access/Articles/lawrence.html.

3. Brody T, Harnad S. Comparing the impact of open access (OA) vs non-OA articles in the same journals. D-Lib Magazine [Internet]. 2004 [cited 2007 Sep 5];10(6). Available from: eprints.ecs.soton.ac.uk/10207/.

4. Hajjem C, Harnad S, Gingras Y. Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. IEEE Data Engineering Bulletin [Internet]. 2005 [cited 2007 Sep 5];28(4):39-47. Available from: eprints.ecs.soton.ac.uk/12906/.

5. Kurtz MJ, Eichhorn G, Accomazzi A, Grant C, Demleitner M, Murray SS, Martimbeau N, Elwell B. The bibliometric properties of article readership information. J Am So Inf Science and Technol [Internet]. 2005 [cited 2007 Sep 5];56(2):111-128. Available from: adsabs.harvard.edu/cgi-bin/bib_query?2005JASIS..56..111K%20target=%22_BLANK%22.

6. Kurtz MJ, Eichhorn G, Accomazzi A, Grant C, Demleitner M, Henneken E, Murray SS. The effect of use and access on citations. Inf Process Manage [Internet]. 2005 [cited 2007 Sep 5];41(6):1395-1402. Available from: adsabs.harvard.edu/cgi-bin/bib_query?2005IPM....41.1395K%20target=%22_BLANK%22.

7. Moed HF. The effect of open access on citation impact: an analysis of ArXiv's condensed matter section. J Am Soc Inf Sci Technol [Internet]. 2007 [cited 2007 Sep 5];58(13):2047-2054. doi: 10.1002/asi.20663.

8. Hajjem C, Harnad S. The open access citation advantage: quality advantage or quality bias? Technical Report, Electronics and Computer Science, University of Southampton; 2007. Available from: eprints.ecs.soton.ac.uk/13328/.

9. Brody T, Harnad S, Carr L. Earlier Web usage statistics as predictors of later citation impact. J Am Assoc In Sci Technol [Internet]. 2006 [cited 2007 Sep 5];57(8):1060-1072. Available from: eprints.ecs.soton.ac.uk/10713/.

10. Brody T. Evaluating research impact through open access to scholarly communication [dissertation]. [Southampton (UK)]: University of Southampton [Internet]; 2006 [cited 2007 Sep 5]. Available from: eprints.ecs.soton.ac.uk/13313/.

11. Harnad S. Opening access by overcoming Zeno's paralysis. In: Jacobs N, editor. Open access: key strategic, technical and economic aspects. Oxford, UK: Chandos [Internet]; 2006 [cited 2007 Sep 5]. Available from: eprints.ecs.soton.ac.uk/12094/.

12. Harnad S, Carr L, Brody T, Oppenheim C. Mandated online RAE CVs linked to university eprint archives: improving the UK research assessment exercise whilst making it cheaper and easier. Ariadne [Internet]. 2003 [cited 2007 Sep 5];(35). Available from: www.ariadne.ac.uk/issue35/harnad/.

13. Brody T, Carr L, Gingras Y, Hajjem C, Harnad S, Swan A. Incentivizing the open access research Web: publication-archiving, data-archiving and scientometrics. CTWatch Quarterly [Internet]. 2007 [cited 2007 Sep 5];3(3). Available from: eprints.ecs.soton.ac.uk/14418/01/ctwatch.html.

14. Harnad S. Open access scientometrics and the UK research assessment exercise. In: Torres-Salinas D, Moed HF, editors. Proceedings of 11th annual meeting of the International Society for Scientometrics and Informetrics, Madrid, Spain, 2007 [cited 2007 Sep 5];11(1):27-33, Madrid, Spain. Available from: eprints.ecs.soton.ac.uk/13804/.

15. Shadbolt N, Brody T, Carr L, Harnad S. The open research Web: a preview of the optimal and the inevitable. In: Jacobs N, editor. Open access: key strategic, technical and economic aspects. Oxford, UK: Chandos [Internet]; 2006 [cited 2007 Sep 5]. Available from: eprints.ecs.soton.ac.uk/12453/.

STEVAN HARNAD *is Canada Research Chair in Cognitive Science at the University of Quebec at Montreal and professor in electronics and computer science at the University of Southampton, UK. For more information, visit www.ecs.soton.ac.uk/~harnad.*