Serials Review

The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update[1]

Stevan Harnad, Tim Brody, François Vallières, Les Carr, Steve Hitchcock, Yves Gingras, Charles Oppenheim, Chawki Hajjem, and Eberhard R. Hilf

Available online 6 March 2008

The research access/impact problem arises because journal articles are not accessible to all of their would-be users; hence, they are losing potential research impact. The solution is to make all articles open access (OA, i.e., accessible online, free for all). OA articles have significantly higher citation impact than non-OA articles. There are two roads to OA: the "golden" road (publish your article in an OA journal) and the "green" road (publish your article in a non-OA journal but also self-archive it in an OA archive). About 10% of journals are gold, but over 90% are already green (i.e., they have given their authors the green light to self-archive); yet only about 10–20% of articles have been self-archived. To reach 100% OA, self-archiving needs to be mandated by researchers' employers and funders, as they are now increasingly beginning to do. Serials Review 2008; 34:36–40.

The research *journal-affordability problem* and the resulting university libraries' journal budget crisis were what first brought the research *article-access/impact problem* to light, but the journal-affordability problem and the article-access/impact problem are not the same.

Harnad is Canada Research Chair in Cognitive Science, Université du Québec à Montréal, Quebec, Canada, and Professor of Electronics and Computer Science, University of Southampton, UK; e-mail: harnad@uqam.ca.
Brody is Post doctoral Fellow, Electronics and Computer Science, University of Southampton, UK; e-mail: tdb01r@ecs.soton.ac.uk.
Vallières is Computer Analyst, Observatoire des Sciences et Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal, Quebec, Canada; e-mail: vallieres.francois@uqam.ca.
Carr is Senior Lecturer, University of Southampton, UK; e-mail: lac@ecs.soton.ac.uk.
Hitchcock is Postdoctoral Fellow, University of Southampton, UK; e-mail: S.Hitchcock@ecs.soton.ac.uk.
Gingras is Professor, Université du Québec à Montréal, Quebec, Canada; e-mail: gingras.yves@uqam.ca.
Oppenheim is Professor, University of Loughborough, Leicestershire, UK; e-mail: C.Oppenheim@lboro.ac.uk.
Hajjem is Doctoral Candidate, Université du Québec à Montréal, Quebec, Canada.
Hilf is Professor, University of Oldenburg, Oldenburg, Germany; e-mail: hilf@physnet.physik.uni-oldenburg.de.

According to *Ulrichsweb* [http://www.ulrichsweb.com/ulrichsweb/analysis/], about 25,000 peer-reviewed research journals exist worldwide, across all disciplines and languages, publishing about 2.5 million articles per year. But because journal prices keep rising and library budgets are limited, each university can afford only a small portion of that total. That means their users have access to only a fraction of those articles, even though, in the online age, we would have expected otherwise. This is the research journal-affordability problem.

What the journal-affordability problem unmasked was a further problem: As a consequence of the fact that most of their would-be users at most universities cannot access most of the 2.5 million articles published yearly (because their universities cannot afford the journal access-tolls), a significant portion of the potential research impact of those inaccessible articles is being lost. An article's research impact is the degree to which its findings are read, used, applied, built upon, and cited by researchers in their own further research and applications. Research impact is a measure of the progress and the productivity of research. That is why researchers' careers (their salaries, promotions, tenure, funding, prestige, prizes) depend on their impact; it is also why their universities (which co-benefit from the research funding, progress, and prestige), as well as their research funding agencies (which are answerable for the

0098-7913/$–see front matter.
doi:10.1016/j.serrev.2007.12.005

way they spend tax-payers' money), reward research impact.

Merely to do the research and then put your findings in a desk drawer is no better than not doing the research at all. Researchers must submit their research to peer review[7] and then "publish or perish," so others can use and apply their findings. But getting findings peer reviewed and published is not enough either: Other researchers must find the findings useful, as proved by their actually using and citing them. And to be able to use and cite them, they must first be able to access them. That is the research article access/impact problem.

To see that the journal-affordability problem and the article access/impact problem are not the same one need only note that even if all 25,000 peer-reviewed research journals were sold to universities *at cost*—i.e., with not a penny of profit—it would still be true that almost no university has anywhere near enough money to afford all or even most of the 25,000 journals, even at minimal access-tolls: http://fisher.lib.virginia.edu/cgi-local/arlbin/arl.cgi-task=setuprank. Hence, it would remain true even then that not all would-be users could access all of the yearly 2.5 million articles, and hence that that potential research impact would continue to be lost.

So although the two problems are connected (lower journal prices would indeed generate somewhat more access), solving the journal-affordability problem does not solve the research access/impact problem.

How big is the access/impact problem? Estimates are emerging, and their consistency and size are quite striking. Lawrence[15] reported that in computer science the citation impact of conference articles whose full texts are accessible online toll-free—let us call that "Open Access" (OA), in line with the definition provided in 2001 by the Budapest Open Access Initiative: http://www.soros.org/openaccess/read.shtml—is 336% higher than the impact of non-OA articles. Kurtz et al.[13,14] have reported similar effects in astrophysics, and Odlyzko[16] in mathematics.

We have charted this OA-impact advantage across all disciplines, as well as across time in a study using a twelve-year sample of fourteen million articles from the Institute for Scientific Information (ISI) database. We compared the matched citation counts of OA versus non-OA articles by trawling the Web to find which of the fourteen million articles within the same journal and year are and are not OA. In field after field, the results confirm what *Lawrence* reported (Fig. 1).[5,10]

How did some of the articles in those non-OA journals become OA? Because their authors "self-archived" them on the Web (i.e., made them accessible online toll-free for all would-be users): http://www.eprints.org/self-faq/. Physicists have been self-archiving in growing numbers since 1991, in a central archive called Arxiv (http://arxiv.org/show_monthly_submissions), as have computer scientists on their own Web sites, which are then harvested by Citeseer: http://citeseer.ist.psu.edu/cis.

However, the self-archiving method with the greatest potential to provide OA is self-archiving in one's own university's OAI-compliant Institutional Repositories

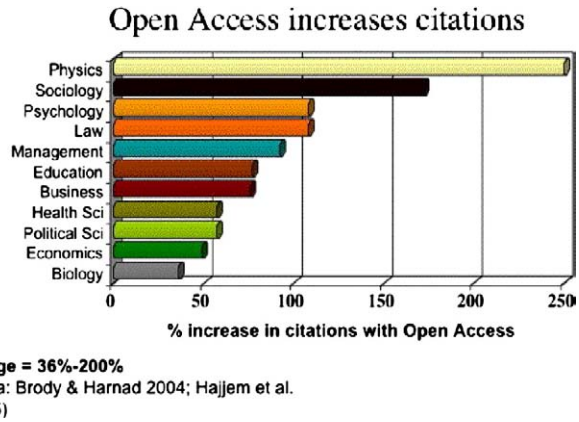

Fig. 1. Average citation ratios for articles in the same journal and year that were and were not made OA by author self-archiving. Date span: 1992–2003. Source: Hajjem et al.[5] and Harnad and Brody.[10]

(IRs). There are already over one hundred such institutional archives worldwide (http://roar.eprints.org/), and they are growing rapidly, but not yet rapidly enough (see Fig. 2).

OAI compliance means using the open archive initiative's metadata-tagging protocol to tag the critical information (author, title, date, etc.) in a uniform way (http://www.openarchives.org/OAI/openarchivesprotocol.html). OAI compliance makes those many distributed archives "interoperable," so that they can all be harvested by cross-archive harvesters such as OAIster (http://oaister.umdl.umich.edu/o/oaister/) or BASE (http://base.ub.uni-bielefeld.de/) into a single, global seamlessly searchable virtual OA archive.

This global OA archive can then be enhanced with a harvester for the research literature such as Google Scholar or Citebase (http://citebase.eprints.org/), which counts citations instead of links and can rank articles by the citation impact or even the "usage impact" (downloads) for the article or the author.[10,12] Early-days measures like the Citebase download/citation correlator (http://citebase.eprints.org/analysis/correlation.php) can even predict eventual citations two years later from the number of downloads today (see Fig. 3 for an area of physics in which the correlation between downloads and citations is about 0.4).

Such performance indicators and predictors can be included in standardized university OAI CVs (http://paracite.eprints.org/cgi-bin/rae_front.cgi) and then harvested by research assessors and evaluators to chart the progress and the direction of research as well as to help make decisions on promotion and funding.[9,11,18] There is evidence that perhaps as many as 39% of authors are already providing OA for *at least one* of their articles by one or the other of the three means of self-archiving (arbitrary Web sites, central disciplinary archives, distributed university archives).[20,21] This 39% now needs to be systematically increased to 100%, *for all articles*, and the institutional self-archiving route is the most promising way to achieve that because universities and their researchers share in the benefits

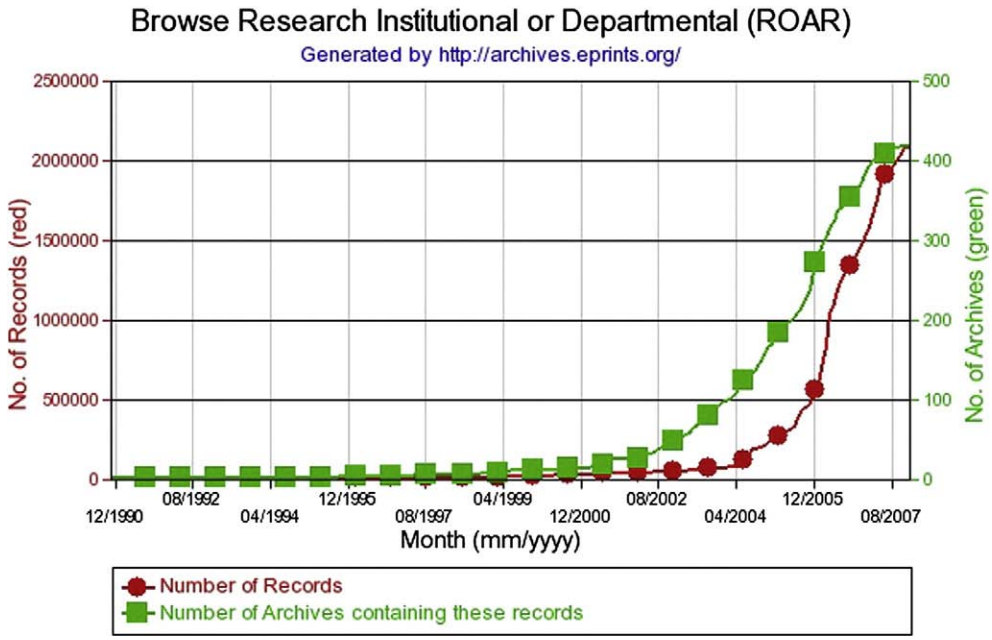

Fig. 2. Growth of institutional archives and contents. Displays a graph of all archives that have been flagged as "Research Institutional." The date stamps of records as exported by the archive's OAI-PMH interface are used to plot a cumulative graph of records over time. The date of the earliest OAI-PMH record is used to show the number of cumulative archives over time (green, scale right). The number of metadata records exported by an archive may not reflect the number of full-text, publicly accessible documents (red, scale left).

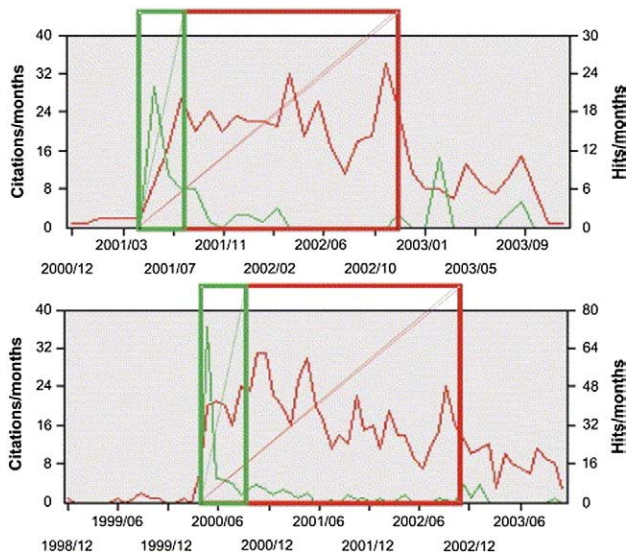

Fig. 3. The download/citation cycle across time. In most areas of physics, the correlation between downloads and citations is between 0.3 and 0.4.[3] These graphs show the time course of downloads (smaller left box) and citations (larger right box) that would be included in calculating the correlation for two papers, where downloads were included up to four months after deposit and citations up to two years. The effect is cyclic, downloads generating citations and citations generating further downloads.

of maximizing research impact and share in the costs of lost impact.

All signs are favorable: There has been a great increase in OA consciousness in the past year, with many Declarations and Statements in support of OA worldwide such as the following:

Berlin Declaration: http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html

WSIS Declaration: http://www.itu.int/wsis/documents/doc_multi-en-1161|1160.asp

Bethesda Statement: http://www.earlham.edu/~peters/fos/bethesda.htm

Budapest Open Access Initiative: http://www.soros.org/openaccess/view.cfm

Public Library of Science: http://www.plos.org/about/history.html

Wellcome Trust Statement: http://www.wellcome.ac.uk/en/1/awtvispolpub.html

IFLA Statement: http://www.ifla.org/V/cdoc/open-access04.html

In response to the research community's expressed desire for OA, the latest JISC/Romeo survey of more than 10,000 journals indicates that over 90% are already "green"; that is, they have given their official green light to author self-archiving (http://romeo.eprints.org/stats.php).[4] Almost 3,000 journals (i.e., about 10% of all

journals) are even "gold"; that is, they are OA journals, making all their own contents OA: http://www.doaj.org/. To cover their costs, however, some of these gold journals have had to adopt the OA journal cost-recovery model:[6] Instead of the user–institution paying the journal access-tolls for incoming articles, the author–institution pays the journal peer review and publication costs per outgoing article.

Currently, the riskiness and the untestedness of this gold journal cost-recovery model make most publishers more willing to go green rather than gold in response to the research community's demand for OA. Physics publishers note that their journals have been green since 1991, and yet there still has not been any cancellation pressure.[19] Universities that can afford to pay for the official non-OA version do so. Users at universities that cannot afford the non-OA version use the authors' self-archived OA versions. One prominent "born-gold" journal—*Journal of High Energy Physics* (http://www.iop.org/EJ/journal/1126-6708)—has even successfully made the transition backwards from gold to green in order to make ends meet after a few years of being toll-free. Yet its contents remain 100% OA because 100% of its authors self-archive them. (Nevertheless, the Sponsoring Consortium for Open Access Publishing in Particle Physics (SCOAP3) has also been converting some physical journals to Gold OA recently: http://www.scoap3.org.)

Most publishers have done their part in response to the research community's demand for OA by giving their green light to author–institution self-archiving. It is now time for more of the research community to take them up on it. It is not enough to sit and wait for all 25,000 journals to convert to gold (http://www.eprints.org/self-faq/#31.Waiting). In addition, it certainly is not fair for researchers to demand that publishers make all the sacrifices and take all the risk upon themselves while the research community does not bother to take the risk-free step of providing OA (which they purport to want and need so much) for their own articles—by simply self-archiving them.[8]

The research community is ready at last to update its existing "publish or perish" mandate to require also providing Open Access to the articles it publishes in the online era. The UK Parliament Science and Technology Committee (http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39903.htm) has recommended (and the US Congress http://www.taxpayer-access.org/ has already voted in favor of) legislation to the effect that as one of the conditions for receiving research funding it should be mandatory for the fundee not merely to publish but also to self-archive all the articles resulting from the funded research. Six of the seven UK Research Councils went on to implement a Green OA self-archiving mandate, and numerous other mandates are at the proposal stage (see the Registry of Open Access Repository Materials Archiving Policies [ROAOMAP]: http://www.eprints.org/openaccess/policysignup/).

In an author survey, Swan and Brown[20,21] report that the vast majority of their author sample indicated that *they would self-archive willingly if their employer (or funding body) required them to do so*! Hence, universities and research funders are in the best position to usher in the OA era by adopting and implementing their own institutional Green OA self-archiving mandates (http://www.eprints.org/signup/sign.php).

More than five hundred universities worldwide already have Institutional Repositories. The adoption of official university OA self-archiving mandates will help to maximize the number of such archives, as well as the number of articles in them—by incentivising self-archiving not only for the sake of the enhanced impact it has been shown OA will generate[5,10] and the rewards of enhanced impact in assessing, crediting, and rewarding research productivity and progress,[9] but for the rich potential of e-science and e-scholarship on the Open Research Web.[2,17]

Along with the substantial recent rise in OA consciousness worldwide, there has also been an unfortunate tendency to equate OA exclusively with OA journal publishing (i.e., the golden road to OA) and to overlook the faster, surer, and already more heavily traveled green road of OA self-archiving. This oversight is probably a spin-off of conflating the journal-affordability problem with the access/impact problem. Perhaps the mounting evidence of the powerful impact-generating effects of OA, plus incentives from their employers and funders, will at last induce the 61% of authors who have not yet done so to take to the green road so that we can all enjoy the benefits of 100% OA.

To date thirty-seven Green OA self-archiving mandates have been adopted worldwide, and nine more have been proposed. Some of those mandates (such as that of NIH in the US, RCUK in the UK and ERC in Europe) have been very big ones, but the majority have so far been research funder mandates (22) rather than university mandates (12), even though virtually all research originates from universities, not all of it is funded, and universities share with their own researchers and students the benefits of showcasing maximizing the uptake of their joint research output. Among the proposed mandates, two are very big multi-university proposals (one for all 791 universities in the forty six countries of the European University Association and one for all the universities and research institutions of Brazils). The world's universities are OA's sleeping giant, and they are where the last big burst of Green OA mandates will come from, the one that takes us across the finish line to 100% OA.

Notes

1. A preliminary version of part of this article appeared as: S. Harnad, T. Brody, F. Vallieres, L. Carr, S. Hitchcock, Y. Gingras, C. Oppenheim, H. Stamerjohanns, E. Hilf, (2004) "The green and the gold roads to Open Access." *Nature Web Focus*. http://www.nature.com/nature/focus/accessdebate/21.html; additionally, a prior version of this article appeared in *Serials Review,* v. 30, no. 4 (2004): 310–314, which the authors were invited to update here.
2. T. Brody, L. Carr, Y. Gingras, C. Hajjem, S. Harnad and A. Swan, "Incentivizing the Open Access Research Web: Publication-Archiving, Data-Archiving and Scientometrics," *CTWatch Quarterly* 3 (3) (2007). http://eprints.ecs.soton.ac.uk/14418/.

3. T. Brody, S. Harnad and L. Carr, "Earlier Web Usage Statistics as Predictors of Later Citation Impact," *Journal of the American Association for Information Science and Technology (JASIST)* 57 (8) (2006): 1060–1072. http://eprints.ecs.soton.ac.uk/10713/.
4. J. Cox and L. Cox, "Scholarly Publishing Practice: The ALPSP Report on Academic Publishers' Policies and Practices in Online Publishing," Association of Learned and Professional Society Publishers, 2003. http://www.alpsp.org/2004pdfs/Sfpub210104.pdf.
5. C. Hajjem, S. Harnad and Y. Gingras, "Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How It Increases Research Citation Impact," *IEEE Data Engineering Bulletin* 28 (4) (2005): 39–47. http://eprints.ecs.soton.ac.uk/11688/.
6. S. Harnad, "Electronic Scholarly Publication: Quo Vadis?" *Serials Review* 21 (1) (1995): 70–72 (Reprinted in *Managing Information* 2(3) 1995). http://cogprints.ecs.soton.ac.uk/archive/00001691/00/harnad95.quo.vadis.html.
7. Stevan Harnad, "The Invisible Hand of Peer Review," *Nature* (1998) [online] (5 Nov. 1998). http://helix.nature.com/webmatters/invisible/invisible.html. Longer version in Exploit Interactive 5 (2000). http://www.exploit-lib.org/issue5/peer-Ereview/.
8. S. Harnad, "Opening Access by Overcoming Zeno's Paralysis," in N. Jacobs (Ed.), *Open Access: Key Strategic, Technical and Economic Aspects*, 2006, Chandos. http://eprints.ecs.soton.ac.uk/12094/.
9. S. Harnad, "Open Access Scientometrics and the UK Research Assessment Exercise," in D. Torres-Salinas, H.F. Moed (Eds.), *Proceedings of 11th Annual Meeting of the International Society for Scientometrics and Informetrics*, vol. 11 (1), 2007, pp. 27–33, Madrid, Spain. http://eprints.ecs.soton.ac.uk/13804/.
10. S. Harnad and T. Brody, "Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals," *D-Lib Magazine* 10 (6) (2004) June. http://www.dlib.org/dlib/june04/harnad/06harnad.html.
11. S. Harnad, L. Carr, T. Brody and C. Oppenheim, "Mandated Online RAE CVs Linked to University Eprint Archives: Improving the UK Research Assessment Exercise Whilst Making It Cheaper and Easier," *Ariadne* 35 (2003, April). http://www.ariadne.ac.uk/issue35/harnad/.
12. S. Hitchcock, A. Woukeu, T. Brody, L. Carr, W. Hall and S. Harnad, "Evaluating Citebase: An Open Access Web-Based Citation-Ranked Search and Impact Discovery Service," 2003. http://opcit.eprints.org/evaluation/Citebase-evaluation/evaluation-Ereport.html.
13. Michael J. Kurtz, Guenther Eichhorn, Alberto Accomazzi, Carolyn S. Grant, Markus Demleitner, Stephen, S. Murray, Nathalie Martimbeau and Barbara Elwell, "Worldwide Use and Impact of the NASA Astrophysics Data System Digital Library," *Journal of the American Society for Information Science and Technology* 55 (2004). http://cfa-www.harvard.edu/~kurtz/jasist1.pdfhttp://cfa-www.harvard.edu/~kurtz/jasist1.pdf.
14. Michael J. Kurtz, Guenther Eichhorn, Alberto Accomazzi, Carolyn S. Grant, Markus Demleitner, Stephen S. Murray, Nathalie Martimbeau and Barbara Elwell, "The Bibliometric Properties of Article Readership Information," *Journal of the American Society for Information Science and Technology* 55 (2004). http://cfa-www.harvard.edu/~kurtz/jasist2.pdf.
15. S. Lawrence, "Online or Invisible?" *Nature* 411 (6837) (2001): 521. http://www.neci.nec.com/~lawrence/papers/online-nature01/.
16. A.M. Odlyzko, "The Rapid Evolution of Scholarly Communication," *Learned Publishing* 15 (2002): 7–19. http://www.catchword.com/alpsp/09531513/v15n1/contp1-1.htm.
17. N. Shadbolt, T. Brody, L. Carr and S. Harnad, "The Open Research Web," in N. Jacobs (Ed.), *Open Access*, 2006, Chandos. http://eprints.ecs.soton.ac.uk/12453/.
18. A. Smith and M. Eysenck (2002), "The Correlation Between RAE Ratings and Citation Counts in Psychology." *Technical Report, Psychology, University of London, Royal Holloway.* http://psyserver.pc.rhbnc.ac.uk/citations.pdf.
19. A. Swan, "Open Access and the Progress of Science," *American Scientist* 95 (3) (2007). http://eprints.ecs.soton.ac.uk/13860/.
20. A. Swan and S.N. Brown, *JISC/OSI Journal Authors Survey Report* (2004a), http://www.jisc.ac.uk/uploaded_documents/JISCOAreport1.pdf http://www.ecs.soton.ac.uk/~harnad/Hypermail/Amsci/3628.html.
21. A. Swan and S.N. Brown, "Authors and Open Access Publishing," *Learned Publishing* 17 (3) (2004b): 219–224. http://www.ingentaselect.com/rpsv/cw/alpsp/09531513/v17n3/s7/.