

# The Access/Impact Problem and the Green and Gold Roads to Open Access

[Stevan Harnad](#) (U. Québec/Montréal), [Tim Brody](#) (Southampton U.), [François Vallières](#) (OST/U. Québec/Montréal)  
[Les Carr](#) (Southampton U.), [Steve Hitchcock](#) (Southampton U.) [Yves Gingras](#) (U. Québec/Montréal),  
[Charles Oppenheim](#) (Loughborough U.), [Heinrich Stamerjohanns](#) (Oldenburg U.), [Eberhard R. Hilf](#) (Oldenburg U.)

It was the research journal affordability problem and the resulting university libraries' journal budget crisis that first brought the research access/impact problem to light, but the journal affordability problem and the research access/impact problem are not the same problem.

There are about 24,000 peer-reviewed research journals -- <http://www.ulrichsweb.com/ulrichsweb/analysis/> -- publishing about 2.5 million articles per year. Because of spiralling price rises, libraries have been able to afford to subscribe to fewer and fewer of those journals (despite bundled "Big Deal" licenses), and are hence providing their users with access to a smaller and smaller fraction of those yearly 2.5 million articles, even though, in the online age, we would have expected the opposite. That is the journal affordability problem.

What the journal affordability problem unmasked was another problem: Most would-be users at most universities cannot access most of the 2.5 million articles published yearly (because their university cannot afford the access-tolls). As a consequence, much of the potential impact of those inaccessible articles is lost. An article's research impact is the degree to which its findings are read, used, applied, built-upon and cited by users in their own further research articles.

Research impact is a measure of the progress and productivity of research. This is why researchers' careers (their salaries, promotions, tenure, funding, prestige, prizes) depend on impact; this is 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 way they spend tax-payers' money) reward impact:

It is not enough to merely do the research and then put your findings in a desk-drawer; that is no better than not doing the research at all. Researchers must submit their research to peer review and then "publish or perish" it so others can use and apply their findings. But getting findings peer-reviewed and published is not enough either: Users must find the findings useful, as proved by their using and citing them. (The three-fold repetition of the "u-word" here was intentional!) And to be able to use and cite them, users must first be able to access them. That is the research access/impact problem.

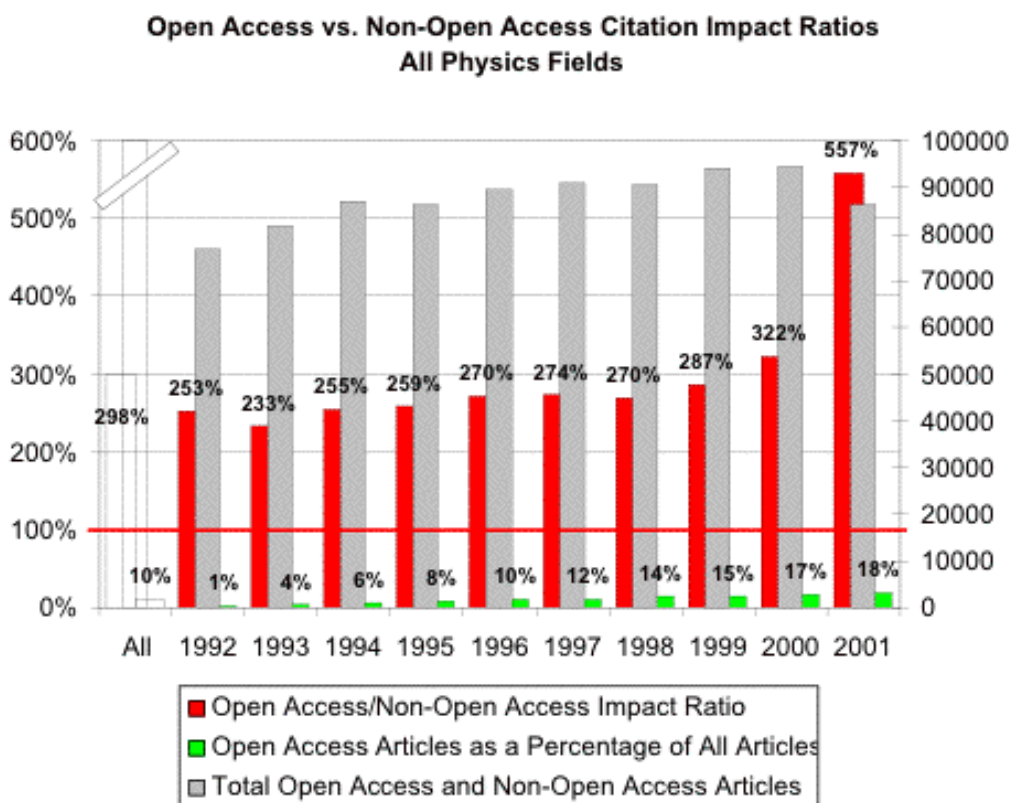
To see that the journal affordability problem and the research access/impact problem are not the same problem one need only note that even if all 24,000 peer-reviewed research journals were sold to universities at cost -- i.e. with not a penny of profit -- it would still be true that almost no university could afford all or even most of the 24,000 journals even at those lower access-tolls. <http://fisher.lib.virginia.edu/cgi-local/arlbin/arl.cgi?task=setuprank> Hence it would remain true even then that most would-be users could not access most of the yearly 2.5 million articles, and that all 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.

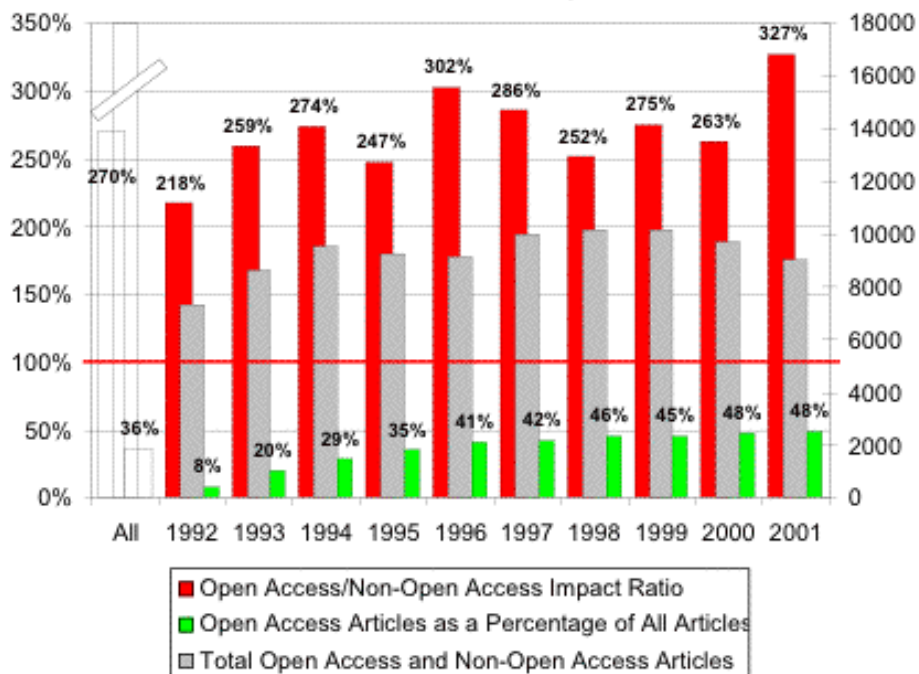
How big is the access/impact problem? Estimates are emerging, and their size is quite astounding: [Lawrence \(2001\)](#) reported that in computer science the citation impact of articles that are accessible online toll-free -- let us call that "Open Access" (OA), in line with the definition provided in 2001 by the Budapest Open Access Initiative: <http://www.soros.org/openaccess/read.shtml> -- is 336% higher. Kurtz et al. ([2003](#), [2004](#)) have reported similar effects in

astrophysics.

We are charting this OA-impact effect across all disciplines as well as across time in a study using a 10-year sample of 14 million articles from the Institute for Scientific Information (ISI) database. We are comparing the matched citation counts of OA versus TA (Toll Access) articles by trawling the web to find which of the 14 million articles within the same journal and year are and are not OA. Results are already available for physics, and the effects there are at least as dramatic as [Lawrence](#) reported, and seem to peak especially within 3 years of the paper's publication date ([Brody et al. 2004](#)):



**Open Access vs. Non-Open Access Citation Impact Ratios  
Nuclear and Particle Physics**



How did some of the articles in those TA journals become OA? Because their authors "self-archived" them on the web (i.e., made them accessible toll-free for all would-be users): <http://www.eprints.org/self-faq/>

We know that physicists have been self-archiving in growing numbers since 1991, in a central archive called Arxiv -- [http://arxiv.org/show\\_monthly\\_submissions](http://arxiv.org/show_monthly_submissions) -- and that computer scientists have meanwhile been doing the same on their own websites, which are then harvested by Citeseer: <http://citeseer.ist.psu.edu/cis>.

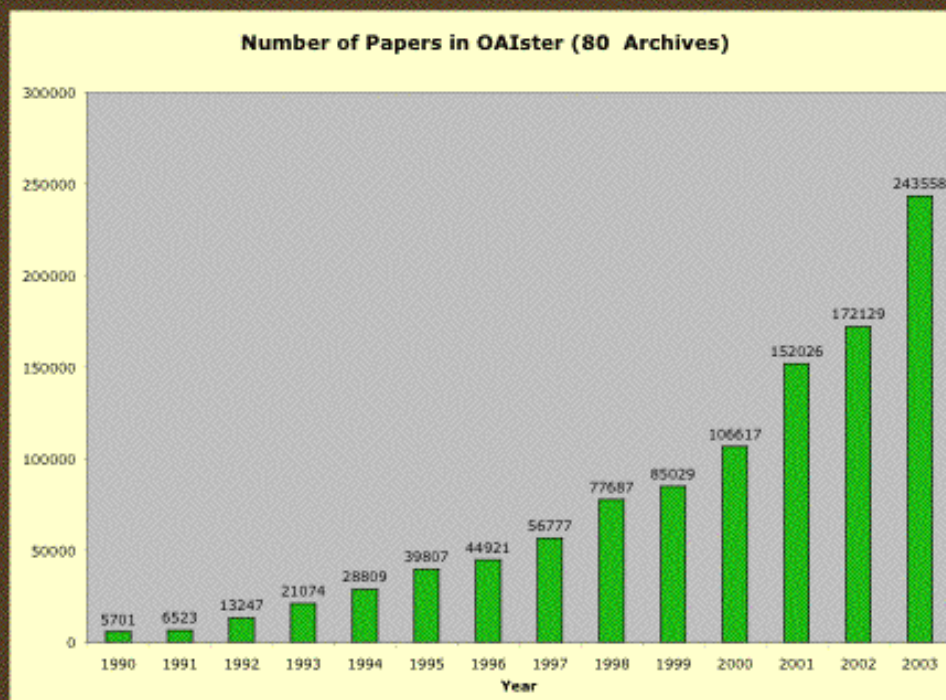
**GNU**

But the self-archiving method with the biggest potential to provide OA is self-archiving in one's own university's OAI-compliant Eprint Archives: <http://software.eprints.org/handbook/>. OAI-compliance means using the Open Archive Initiative's metadata-tagging protocol: <http://www.openarchives.org/OAI/openarchivesprotocol.html>. OAI-compliance makes those many distributed archives "interoperable" with one another, so that they can all be harvested by cross-archive harvesters such as OAIster -- <http://oaister.umdl.umich.edu/o/oaister/> -- into a single, global seamlessly-searchable virtual OA archive:



OAIster, a cross-archive search engine, now covers over 250 OAI Archives (about half of them Eprints.org Archives) indexing over 3 million items (but not all research papers, and not all full-texts). Below are data for just the full-text research papers with 1990-2003 creation dates.

<http://oaister.umdl.umich.edu/o/oaister/>



This sort of global OA archive can then be enhanced with a "google" for the research literature such as Citebase -- <http://citebase.eprints.org/> -- which counts citations instead of links, and can rank articles by either the citation impact or the "download impact" of the article or the author ([Hitchcock et al. 2003](#)). 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.



# Time-Course of Citations (red) and Usage (hits, green)

Witten, Edward (1998) String Theory and Noncommutative Geometry *Adv. Theor. Math. Phys.* 2 : 253

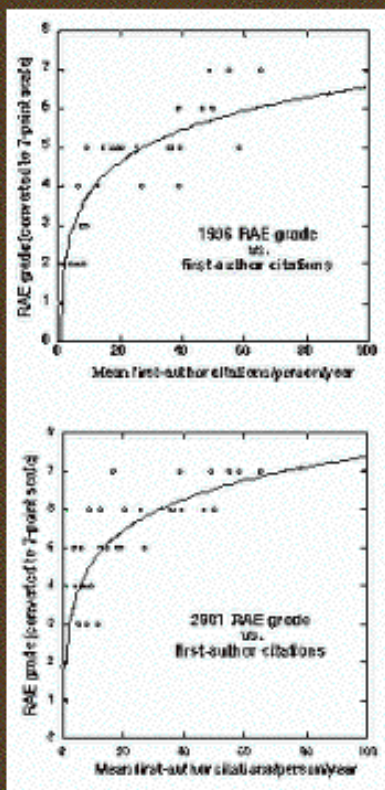


1. Preprint or Postprint appears.
2. It is downloaded (and sometimes read).
3. Eventually citations may follow (for more important papers).
4. This generates more downloads, etc.

Such performance indicators and predictors can be included in standardized university OAI CVs -- [http://paracite.eprints.org/cgi-bin/rae\\_front.cgi](http://paracite.eprints.org/cgi-bin/rae_front.cgi) -- and then harvested by research assessors and evaluators to chart the progress and direction of research as well as to help make decisions on promotion and funding ([Harnad et al. 2003](#)):



# Research Assessment, Research Funding, and Citation Impact



“Correlation between RAE ratings and mean departmental citations +0.91 (1996) +0.86 (2001) (Psychology)”

“RAE and citation counting measure broadly the same thing”

“Citation counting is both more cost-effective and more transparent”

(Eysenck & Smith 2002)

<http://psyserver.pc.rhbnc.ac.uk/citations.pdf>

There is now evidence that as many as 40% of authors are already providing OA for their articles by one or the other of these three means of self-archiving (arbitrary websites, central disciplinary archives, distributed university archives) ([Swan & Brown 2004](#)):

This 40% now needs to be systematically increased to 100%, and the institutional self-archiving route is the most promising way to achieve that, because universities and their researchers share in the benefits of maximising research impact, and share in the costs of lost impact. Hence universities are in the best position to implement their own OA provision policies:

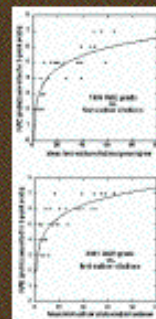
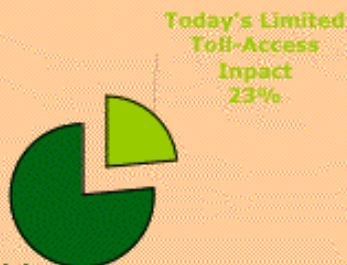
<http://www.eprints.org/signup/sign.php>

More than 100 universities worldwide -- <http://archives.eprints.org/eprints.php?page=all> -- already have Eprint archives. The adoption of official university OA provision policies will help to maximise the number of Eprint archives as well as the number of articles in them, in part by encouraging the 40% of their researchers who already self-archive to deposit their articles in their own university's Eprint Archive, in part by encouraging those of them who do not yet self-archive to do start doing so, for the sake of the dramatically enhanced impact that the citation studies are demonstrating that OA will generate.



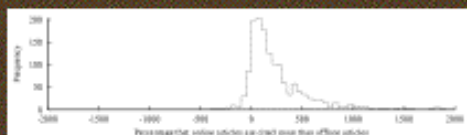
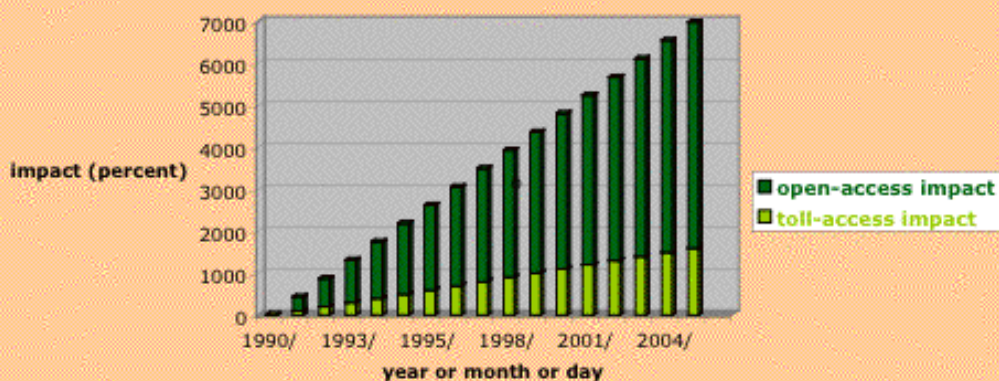
**Quo usque tandem patientia nostra...?**  
*How long will we go on letting our cumulative daily/monthly/yearly research-impact losses grow, now that the online medium has made it all preventable?*

**What we stand to gain:**



.91 correlation with UK research ranking and funding

**Yearly/Monthly/Daily Impact Loss**



336% higher impact

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

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

**WSIS Declaration:** [http://www.itu.int/ws/s/documents/doc\\_multi-en-1161|1160.asp](http://www.itu.int/ws/s/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 fervently expressed desire for OA, the JISC/Romeo survey of over 7000 journals indicates that 55% (60% according to Cox & Cox's [2003] ALPSP survey) have already declared themselves "green" -- that is, they officially endorse author self-archiving:

<http://www.lboro.ac.uk/departments/ls/disresearch/romeo/Romeo%20Publisher%20Policies.htm>

Almost 1000 journals (i.e., approaching 5%) are even "gold" -- that is, they are OA journals, making all their own contents OA: <http://www.doaj.org/> To cover their costs, however, many of these gold journals have had to adopt the OA journal cost-recovery

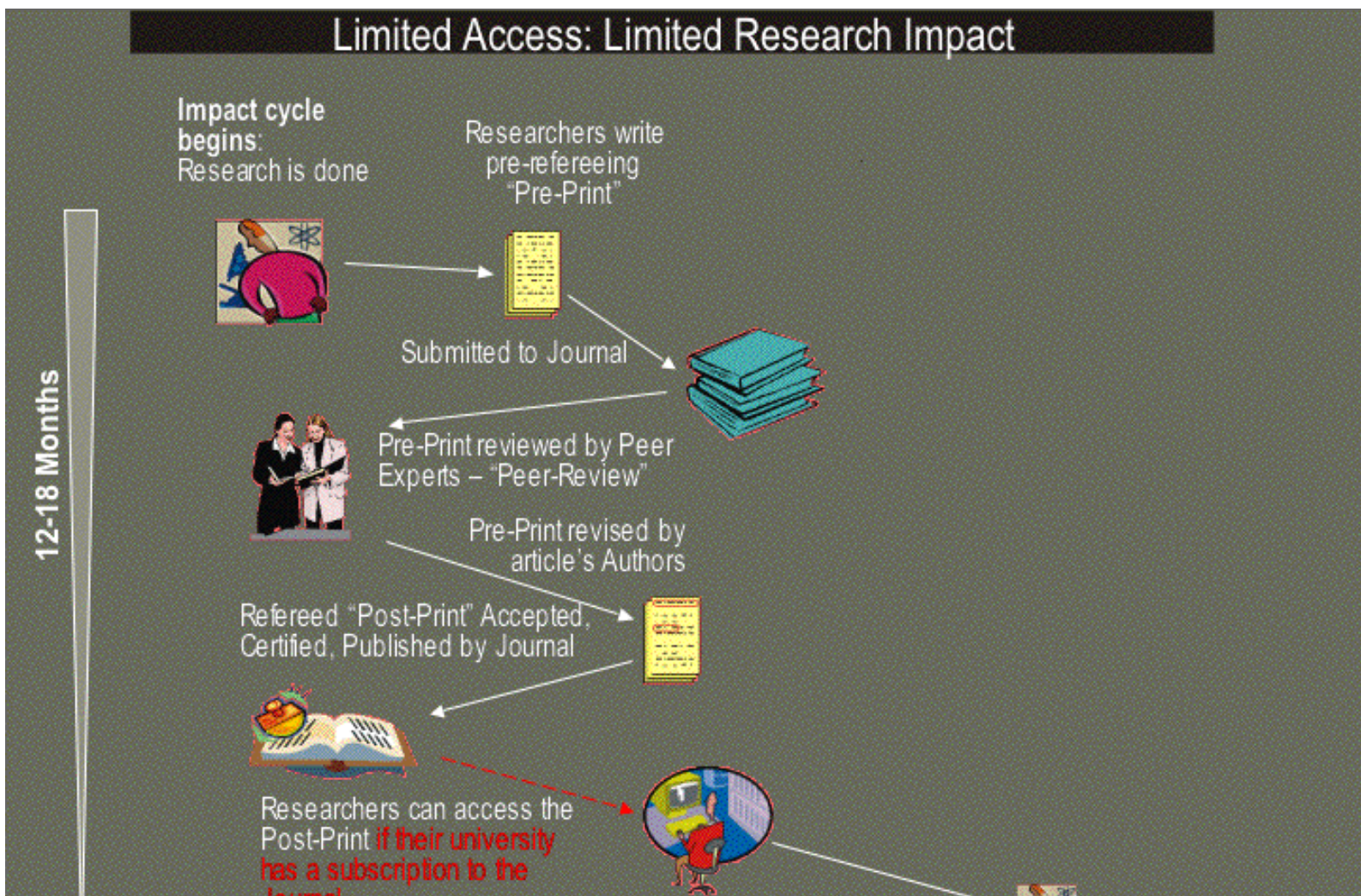


model ([Harnad 1995](#)): Instead of the user-institution paying the journal access-tolls for incoming articles, the author-institution pays the journal peer-review and publication costs per outgoing article. (Not all OA Journals have as yet registered themselves in DOAJ: e.g., in physics, cf. <http://de.physnet.net/PhysNet/journals.html> .)

It is the riskiness and untestedness of this gold journal cost-recovery model that makes publishers more willing to go green rather than gold in response to the research community's demand for OA at this time. Publishers note that physics journals have been green since 1991 and yet there still has not been any cancellation pressure: Universities that can afford to pay for the TA version do so. Users at universities that cannot afford the TA version use the authors' self-archived OA versions. One prominent born-gold journal -- Journal of High Energy Physics <http://jhep.sissa.it/> -- 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. But its contents remain 100% OA, because 100% of its authors self-archive them.

Publishers have done their part in response to the research community's demand for OA, by giving self-archiving the green light. 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 24,000 journals to convert to gold. And it certainly isn't fair for the research community 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, for their own articles, that OA that they purport to want and need so much -- by simply self-archiving them!

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., with only the golden road to OA, overlooking the faster, surer and already more heavily travelled green road. We think this oversight is a spin-off of conflating the journal-affordability problem with the access/impact problem. Let us hope that the mounting evidence of the powerful impact-generating effects of OA will at last persuade the 60% of authors (and their institutions) who have not yet done so to take to the green road so we can all enjoy the benefits of 100% OA at last.

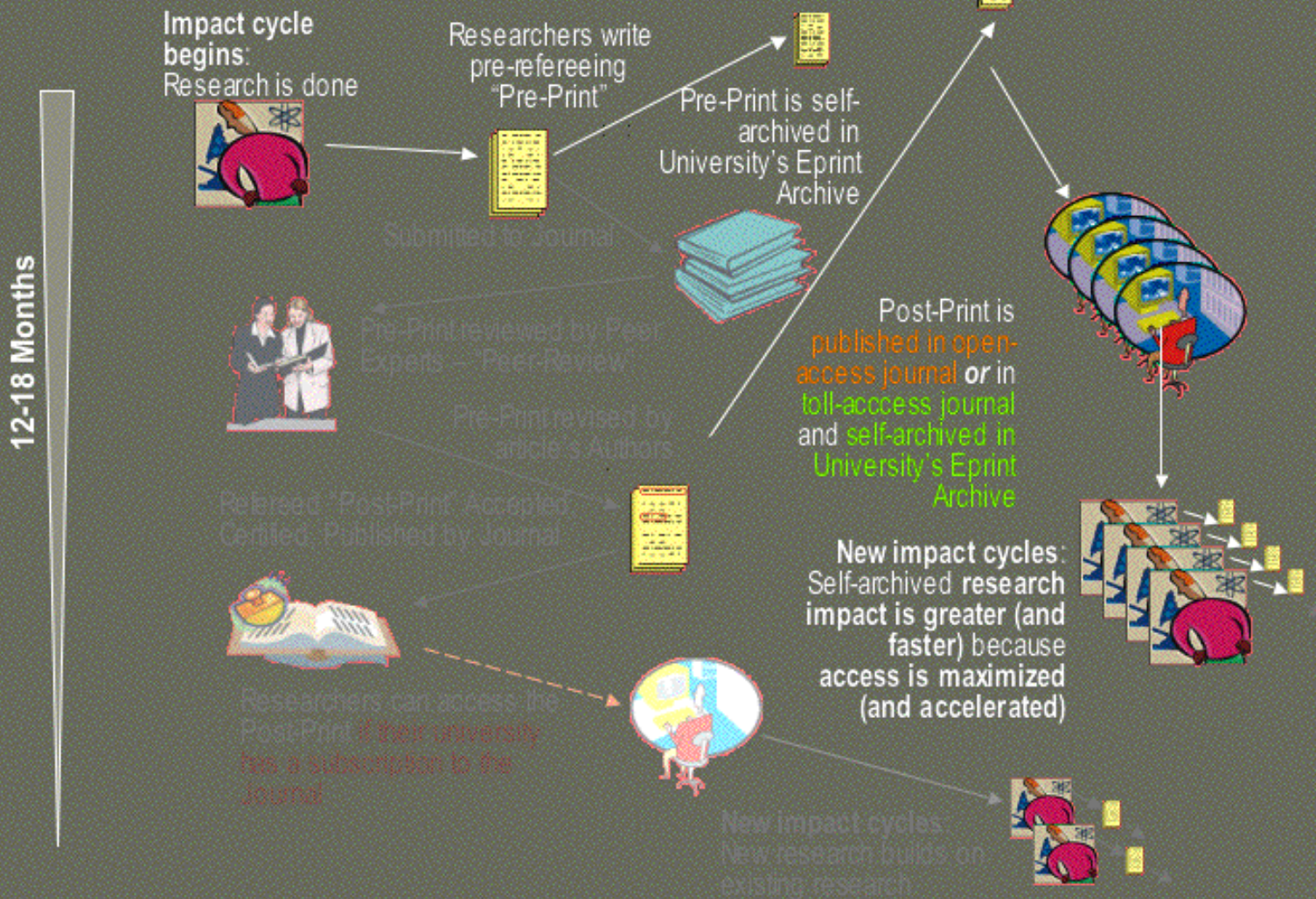




has a subscription to the Journal

New impact cycles:  
New research builds on existing research

## Maximized Research Access and Impact Through Self-Archiving



### References

Brody, T., Stamerjohanns, H., Harnad, S. Gingras, Y. & Oppenheim, C. (2004) The effect of Open Access on Citation Impact. Presented at: National Policies on Open Access (OA) Provision for University Research Output: an International meeting, Southampton, 19 February 2004. <http://opcit.eprints.org/feb19prog.html> <http://www.ecs.soton.ac.uk/~harnad/Temp/OA-TAadvantage.pdf>

Cox, J. & Cox, L. (2003) Scholarly Publishing Practice: The ALPSP report on academic publishers' policies and practices in online publishing. Association of Learned and Professional Society Publishers. <http://www.alpsp.org/2004pdfs/SFpub210104.pdf>

Harnad, S. (1995) Electronic Scholarly Publication: Quo Vadis? *Serials Review* 21(1) 70-72 (Reprinted in *Managing*

Information 2(3) 1995) <http://cogprints.ecs.soton.ac.uk/archive/00001691/00/harnad95.quo.vadis.html>

Harnad, S., Carr, L., Brody, T. & Oppenheim, C. (2003) Mandated online RAE CVs Linked to University Eprint Archives: Improving the UK Research Assessment Exercise whilst making it cheaper and easier. *Ariadne* 35 (April 2003).  
<http://www.ariadne.ac.uk/issue35/harnad/>

Hitchcock, S., Woukeu, A., Brody, T., Carr, L., Hall, W., and Harnad, S. (2003) Evaluating Citebase, an open access Web-based citation-ranked search and impact discovery service <http://opcit.eprints.org/evaluation/Citebase-evaluation/evaluation-report.html>

Kurtz, Michael J.; Eichhorn, Guenther; Accomazzi, Alberto; Grant, Carolyn S.; Demleitner, Markus; Murray, Stephen S.; Martimbeau, Nathalie; Elwell, Barbara. (2003) The NASA Astrophysics Data System: Sociology, Bibliometrics, and Impact. *Journal of the American Society for Information Science and Technology* <http://cfa-www.harvard.edu/~kurtz/jasis-abstract.html>

Kurtz, M.J. (2004) Restrictive access policies cut readership of electronic research journal articles by a factor of two, Michael J. Kurtz, Harvard-Smithsonian Centre for Astrophysics, Cambridge, MA <http://opcit.eprints.org/feb19oa/kurtz.pdf>

Lawrence, S. (2001) Online or Invisible? *Nature* 411 (6837): 521. <http://www.neci.nec.com/~lawrence/papers/online-nature01/>

Swan, A. & Brown, S.N. (2004) JISC/OSI Journal Authors Survey Report.  
[http://www.jisc.ac.uk/uploaded\\_documents/JISCOAreport1.pdf](http://www.jisc.ac.uk/uploaded_documents/JISCOAreport1.pdf)  
<http://www.ecs.soton.ac.uk/~harnad/Hypermail/Amsci/3628.html>