



# How and Why To Free All Refereed Research From Access- and Impact-Barriers Online, Now

Stevan Harnad, Les Carr, Tim Brody

## Abstract

Researchers publish their findings in order to make an impact on research, not in order to sell their words. Access-tolls are barriers to research impact. Authors can now free their refereed research papers from all access tolls immediately by self-archiving them on-line in their own institution's Eprint Archives. Free [eprints.org](http://eprints.org) software creates Archives compliant with the [Open Archives Initiative](#) metadata-tagging Protocol OAI 1.0. These distributed institutional Archives are interoperable and can hence be harvested into global "virtual" archives, citation-linked and freely navigable by all. Self-archiving should enhance research productivity and impact as well as providing powerful new ways of monitoring and measuring it.

=====

Why do scientists (and scholars) do research and report their findings? In a word, it is so that their findings will have an impact -- not just in the narrow sense of the "citation impact factor" (the number of subsequent research reports that cite their findings [[Garfield 1955](#)]), but impact in the broadest sense: Researchers want their work to make a difference, to build upon the work of others, and to be built upon in turn by others. They want to make a contribution to human knowledge; and it is no contribution if it is not noticed and has no consequences.

How do researchers maximize the impact of their research findings? By making them public through publishing them, so that any potentially interested fellow-researcher anywhere in the world, now and at any future time, can access and use them. The findings are published in peer-reviewed journals, which thereby perform a double service for research and researchers. They not only (i) make the findings accessible to the world (on-paper in the Gutenberg era, both on-paper and on-line in the PostGutenberg era), but they (ii) certify their quality-level too. They do this by implementing peer review [[Harnad 1998/2000](#)].

Peer review is the evaluation and validation of the work of experts by qualified fellow-experts (referees) as a precondition for acceptance and publication, so that the research community at large can know which work is likely to be worth the time and effort of reading and trying to build upon. Peer review is not a red-light/green-light, accept/reject system: It is a dynamic interaction between the author and referees, mediated by and answerable to a qualified expert (the Editor). It sometimes involves several rounds of revision and re-refereeing before a final draft can be certified as having

met the quality standards of a particular journal. There is a hierarchy of journal-quality in most fields, with the higher quality journals tending to have the higher rejection rates and higher impact factors [[Yamazaki 1995](#)], grading all the way down to a vanity press at the bottom. Peer review accordingly also performs the function of filtering and triage, sign-posting the resultant literature for navigation. It has been suggested that most papers are eventually accepted somewhere in their field's hierarchy [[Lock 1985](#)], but this may differ from field to field [[Hargens 1988](#)]).

In the on-paper era, journals provided the double service of quality-control and certification [QC/C] plus dissemination for refereed research reports. Providing that service cost money, which then had to be recovered (along with a fair profit) from subscription (S) fees (mostly from institutions), and lately, in the on-line era, also from institutional site-license (L) and/or pay-per-view (P) fees. Let us call these three fee-based access tolls, jointly, S/L/P.

It is important to note an immediate conflict of interest between access-tolls and research here: Researchers conduct and report research for impact, as we have noted. But S/L/P fee-based access-barriers are necessarily also impact-barriers. Institutions and individuals that cannot or do not pay the S/L/P tolls cannot access the research: All this blocked access adds up to lost potential impact for the researcher. Is there any way to resolve this conflict of interest? There is, but to find it, we first have to clearly understand where the conflict resides.

Authors of refereed research reports are not representative of authors in general (not even of themselves, when wearing other hats); in fact, they are highly anomalous. Unlike the authors of books, who write their texts for royalty income, or the authors of magazine articles, who write them for fee income, the authors of refereed journal articles write solely for impact: Their texts are, and always have been give-aways, whereas most of the rest of the published literature is non-give-away [[Harnad, Varian & Parks 2000](#)]. The rewards for researchers (research-funding, salary, promotion, tenure, prizes) come from the impact of their research, not from the S/L/P toll income (which does not accrue to them in any case, but to their publishers). This is why access-barriers create a conflict of interest for this nonstandard minority, the give-away authors, but not for the majority, the non-give-away authors, on whose much more representative interests publishing in general is (rightly) modeled.

How much impact-loss do S/L/P access-barriers cause research and researchers annually? We can only make crude guesses at this point, because the data are not available. A direct estimate would require comparing the citation impact for comparable and representative samples of literature under metered and unmetered access conditions. At best, we can treat differences between S/L/P-limited and S/L/P-free access levels as estimates of upper bounds for potential differences in impact-levels (although even these could be underestimates just as readily as overestimates, if the relation between access and impact is nonlinear). A hint of what unmetered usage levels would look like comes from that small portion of the refereed literature that has already been freed of S/L/P in Physics. It is a reasonable assumption that most of the free daily downloads of papers from the Los Alamos Physics Archive and its 14 worldwide mirror sites represent a net increment in access, hence potential impact for those papers [[http://arXiv.org/cgi-bin/show\\_weekly\\_graph](http://arXiv.org/cgi-bin/show_weekly_graph)] over S/L/P baselines. The only users who could otherwise have accessed all those papers that freely would have been the lucky ones who happened to be at institutions that could afford online S/L/P access to all the journals in which those papers appeared. But no institution anywhere near that lucky (or wealthy) exists.

To see why no such lucky institution exists, we need to consider the total number of refereed journals currently published annually. A conservative estimate would be the 20,000 active refereed journals indexed by Ulrich's Periodicals Directory [<http://www.ulrichsweb.com/ulrichsweb/>]). A

conservative estimate of the average number of papers appearing in each would be 100 (the numbers range from 12 to 1200 according to ISI's Web of Science [<http://wos.mimas.ac.uk/>]) for an annual total of 2 million refereed papers. What is the average proportion of the 2 million annual papers that is currently inaccessible per research institution because of the limits on annual institutional S/L/P budgets? Even the most conservative estimate of 0.5 would mean that the lost potential access is enormous. The lost potential impact will also be some function of that figure [[Odlyzko 1998](#), [1999a](#), [1999b](#)].

The proportion will of course be different for Harvard, with perhaps the largest S/L/P budget in the world (but still short of being able to afford the annual total of 20K refereed journals), compared to universities in the developing world, or even the less wealthy universities in the U.S. [<http://fisher.lib.virginia.edu/newarl/index.html>]. But the Los Alamos Physics un-metered usage levels suggest that even Harvard researchers may have a good deal to gain -- both in terms of their own access to the research of others and the impact of their own research on others -- if the entire research corpus could be freed of all impact- and access-barriers.

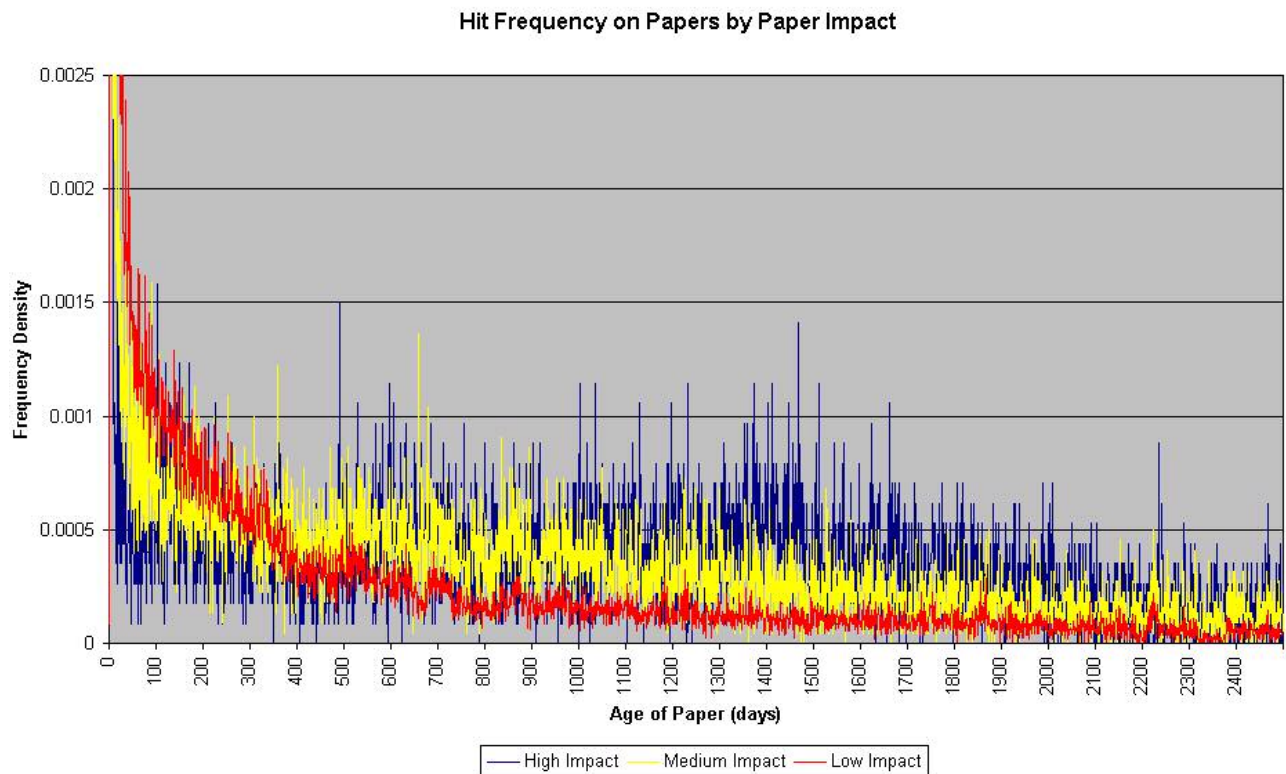
Well the good news is that it can be: Virtually all the papers in all the refereed journals in all fields can now be freed of all S/L/P barriers by author-institution self-archiving [[Harnad 1994](#)]. Physicists have been the first to recognize and exploit the feasibility of this, but even they are still doing it too slowly: At its present (linear) rate of growth (150K papers archived so far, 30K per year, annual growth 3.5K) [[http://www.openarchives.org/DC2001/warner\\_long.pdf](http://www.openarchives.org/DC2001/warner_long.pdf)], it will take the Physics Archive another decade to free the full annual refereed corpus of physics (at least 300K refereed papers published annually in physics, astrophysics, and mathematics according to ISI's Web of Science [<http://wos.mimas.ac.uk/>]). Other fields are even further behind [<http://cogprints.soton.ac.uk/>], and most have not even started. Why not? Future historians will need to answer this question, but as of January 2001 [<http://www.openarchives.org/DC2001/OpenMeeting.html>], it will certainly not be for lack of the universal means to do so, immediately.

Free "Eprint" archive-creating software (using only free resources) has just been designed [<http://www.eprints.org/>] to make it possible for all universities and research institutions worldwide to immediately create their own archives, in which all their researchers can then self-archive all their papers online (n.b., "eprints" include *both* pre-refereeing preprints and electronic refereed postprints, in electronic form). These Eprint archives are not only extremely easy and cheap to install and maintain, but they are all fully interoperable with one another, through compliance with the Open Archives Initiative's [<http://www.openarchives.org/> metadata tagging protocol OAI 1.0 released on January 23. This means that all the Eprint archives are like clones, and can be registered [<http://oaisrv.nsd.cornell.edu/Register/BrowseSites.pl>] and "harvested" into one (or many) global "virtual" archives [e.g. <http://arc.cs.odu.edu/>], so that researchers worldwide can search and retrieve the entire refereed corpus by discipline, topic, keyword, author, journal, etc., with no need to know which institutional Eprint archive a paper happens to be deposited in.

The distributed Eprint archives can be multiply mirrored at "twinned" sites for reliability and backup. Their contents can also be citation-linked, so users can surf from paper to paper via the "mother of all hyperlinks," reference citation. The OpCit Project [<http://opcit.eprints.org/>] has demonstrated this by citation-linking the centralized Physics Archive; that same feature can be applied to distributed Eprint Archives too [[Hitchcock et al. 2000](#)]. A barrier-free refereed corpus online also spawns new scientometric measures of impact, productivity, and the time-course and direction of evolving knowledge [[Harnad & Carr 2000](#)]. Citation-linking allows any user to retrieve the results of searches ranked by the papers', authors' or journals' citation impact [[Figure 1](#)]. Download impact [[Table 1](#)], prepublication immediacy factors, and still newer metrics can be

gathered and analysed from this digitized corpus to complement the classical citation impact measures, offering us a much deeper and richer analysis of the embryological stages in the development of knowledge, from the pre-refereeing preprint, through successive stages of revision, to the refereed, journal-certified postprint, to postpublication revisions, corrections, updates, commentaries, and responses -- all can be linked and threaded together in the Eprint Archives, charting a "scholarly skywriting" continuum [Harnad 1990] in the PostGutenberg Galaxy [Harnad 1991].

**Figure 1. How long and how often papers are downloaded from the Los Alamos Physics Archive:** Papers can be divided into those receiving high, medium and low numbers of citations (all papers are citation-linked). Note that the higher the citation impact, the greater the download longevity. For further data see: <http://opcit.eprints.org/tdb198/opcit/> and <http://opcit.eprints.org/ijh198/>



**Table 1. Citation Impact vs. Download Impact**

There is a significant positive correlation between how often a paper is cited (citation impact) and how often it is downloaded (download impact), but only in the case of highly cited papers. For further data see: <http://opcit.eprints.org/tdb198/opcit/> and <http://opcit.eprints.org/ijh198/>

Download Type	r	N
All Papers (8-month sample, total cites equally split)	+0.11155	63671
High Citation Papers (40+ cites) (2.0%)	+0.27293*	1981
Medium Citation Papers (13-39 cites) (7.7%)	+0.01288	5937
Low Citation Papers (1-12 cites) (46.5%)	-0.01412	30163

All that is needed in order to provide immediate, unlimited click-through, full-text access to the entire refereed research corpus online, for free, for all, forever, is for universities and research institutions to install Eprint Archives and for their researchers to fill them with all their papers, now. If (a) the enhanced access by their own researchers to the research of others and (b) the enhanced visibility and the resulting enhanced impact of their own research on the research of others are not incentive enough for universities to promote and support the self-archiving initiative energetically at this time, they should also consider that it will be an investment in (c) an eventual solution to their serials crisis and the potential recovery of 90% of their annual serials (S/L/P) budget [Harnad 1998, 1999]. (Note that the success of the self-archiving initiative is predicated on the same Golden Rule on which both refereeing and research themselves are predicated: If we all do our own part for one another, we all benefit from it: Give in order to receive...)

A more detailed account of what Researchers, Universities and Libraries can do right now to hasten the day, including answers to questions about copyright, preservation, embargo policies (such as Science's embargo policy [Harnad 2000a, 2000b]), educational implications [Light et al. 2000] and the future role of refereed journals, is archived online at: [http://library.cern.ch/HEPLW/4/papers/1/#Harnad, S. \(in prep\)](http://library.cern.ch/HEPLW/4/papers/1/#Harnad,S.(in%20prep))

We hope this Policy Forum will induce researchers (and historians and citizens) to reflect upon all the potential research impact being lost forever -- yearly, daily -- the longer we keep putting off doing the Optimal and Inevitable, now that it is entirely within our reach, and could become the Actual virtually overnight. (We are conducting a web survey to try to ascertain why people have and have not already begun to self-archive already. Readers of this paper are invited to complete the survey at: <http://www.ecs.soton.ac.uk/~chh398/Questionnaires>).

## REFERENCES

- [1] Garfield, E., (1955) Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science* 122: 108-111  
[http://www.garfield.library.upenn.edu/papers/science\\_v122\(3159\)p108y1955.html](http://www.garfield.library.upenn.edu/papers/science_v122(3159)p108y1955.html)  
<http://www.garfield.library.upenn.edu/papers/cseimpactfactor05092000.html>
- [2] Hargens, L. (1988) Scholarly Consensus and Journal Rejection Rates. *American Sociological Review* 53: 139-151.
- [3] Harnad, S. (1990) Scholarly Skywriting and the Prepublication Continuum of Scientific Inquiry. *Psychological Science* 1: 342 - 343 (reprinted in *Current Contents* 45: 9-13, November 11 1991).  
<http://cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.skywriting.html>
- [4] Harnad, S. (1991) Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge. *Public-Access Computer Systems Review* 2 (1): 39 - 53  
<http://cogsci.soton.ac.uk/harnad/Papers/Harnad/harnad91.postgutenberg.html>
- [5] Harnad, S. (1994) A Subversive Proposal. In: Ann Okerson & James O'Donnell (Eds.) *Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing*. Washington, DC., Association of Research Libraries, June 1995.  
<http://www.arl.org/scomm/subversive/toc.html>
- [6] Harnad, S. (1998) On-Line Journals and Financial Fire-Walls. *Nature* 395(6698): 127-128  
<http://www.cogsci.soton.ac.uk/~harnad/nature.html>

- [7]Harnad, S. (1998/2000) The invisible hand of peer review. *Nature* [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-review/>  
<http://www.cogsci.soton.ac.uk/~harnad/nature2.html>
- [8]Harnad, S. (1999) Free at Last: The Future of Peer-Reviewed Journals. *D-Lib Magazine* 5(12)  
December 1999 <http://www.dlib.org/dlib/december99/12harnad.html>
- [9]Harnad, S. (2000a) E-Knowledge: Freeing the Refereed Journal Corpus Online. *Computer Law & Security Report* 16(2) 78-87. [Rebuttal to Bloom Editorial in *Science* and Relman Editorial in *New England Journal of Medicine*]  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.scinejm.htm>
- [10]Harnad, S. (2000b) Ingelfinger Over-Ruled: The Role of the Web in the Future of Refereed Medical Journal Publishing. *Lancet* (in press)  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.lancet.htm>
- [11]Harnad, S. & Carr, L. (2000) Integrating, Navigating and Analyzing Eprint Archives Through Open Citation Linking (the OpCit Project). *Current Science* 79(5): 629-638.  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.citation.htm>  
<http://www.iisc.ernet.in/~currsci/sep102000/629.pdf>
- [12]Harnad, S. (in preparation) For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving, Now  
<http://www.cogsci.soton.ac.uk/~harnad/TP/resolution.htm>
- [13]Harnad, S., Varian, H. & Parks, R. (2000) Academic publishing in the online era: What Will Be For-Fee And What Will Be For-Free? *Culture Machine* 2 (Online Journal)  
<http://www.cogsci.soton.ac.uk/~harnad/Temp/Varian/new1.htm>  
[http://culturemachine.tees.ac.uk/frm\\_f1.htm](http://culturemachine.tees.ac.uk/frm_f1.htm)
- [14]Hitchcock, S. Carr, L., Jiao, Z., Bergmark, D., Hall, W., Lagoze, C. & Harnad, S. (2000) Developing services for open eprint archives: globalisation, integration and the impact of links. *Proceedings of the 5th ACM Conference on Digital Libraries*. San Antonio Texas June 2000.  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.acm.htm>
- [15]Light, P., Light, V., Nesbitt, E. & Harnad, S. (2000) Up for Debate: CMC as a support for course related discussion in a campus university setting. In R. Joiner (Ed) *Rethinking Collaborative Learning*. London: Routledge (in press).  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.skyteaching.html>
- [16]Lock, Stephen (1985) *A difficult balance : editorial peer review in medicine* London : Nuffield Provincial Hospitals Trust.
- [17]Odlyzko, A.M. (1998) The economics of electronic journals. In: Ekman R. and Quandt, R. (Eds) *Technology and Scholarly Communication*. Univ. Calif. Press, 1998.  
<http://www.research.att.com/~amo/doc/complete.html>
- [18]Odlyzko, A.M. (1999a) Competition and cooperation: Libraries and publishers in the transition to electronic scholarly journals, A. M. Odlyzko. *Journal of Electronic Publishing* 4(4) (June 1999)

and in J. Scholarly Publishing 30(4) (July 1999), pp. 163-185. The definitive version to appear in The Transition from Paper: A Vision of Scientific Communication in 2020, S. Berry and A. Moffat, eds., Springer, 2000.

<http://www.press.umich.edu/jep/04-04/odlyzko0404.html>

<http://www.research.att.com/~amo/doc/competition.cooperation.pdf>

[19]Odlyzko, A.M. (1999b) The rapid evolution of scholarly communication, to appear in the proceedings of the 1999 PEAK conference.

<http://www.research.att.com/~amo/doc/rapid.evolution.pdf>

<http://www.si.umich.edu/PEAK-2000/>

[20]Yamazaki, S. (1995) Refereeing System of 29 Life-Science Journals Preferred by Japanese Scientists. Scientometrics 33: 123-129.

## **BOX 1**

*What's wrong with this Picture?*

1. A brand-new PhD recipient proudly tells his mother he has just published his first article. She asks him how much he was paid for it. He makes a face and tells her "nothing," and then begins a long, complicated explanation...

2. A fellow-researcher at that same university sees a reference to that same article. He goes to their library to get it: "It's not subscribed to here. We can't afford that journal. (Our subscription/license/loan/copy budget is already overspent)"

3. An undergraduate at that same university sees the same article cited on the Web. He clicks on it. The publisher's website demands a password: "Access Denied:Only pre-paid subscribing/licensed institutions have access to this journal."

4. The undergraduate loses patience, gets bored, and clicks on Napster to grab an MP3 file of his favourite bootleg CD to console him in his sorrows.

5. Years later, the same PhD is being considered for tenure. His publications are good, but they're not cited enough; they have not made enough of a "research impact." Tenure denied.

6. Same thing happens when he tries to get a research grant: His research findings have not had enough of an impact: Not enough researchers have read, built upon and cited them. Funding denied.

7. He decides to write a book instead. Book publishers decline to publish it: "It wouldn't sell enough copies because not enough universities have enough money to pay for it. (Their purchasing budgets are tied up paying for their inflating annual journal subscription/license/loan costs...)"

8. He tries to put his articles up on the Web, free for all, to increase their impact. His publisher threatens to sue him and his server-provider for violation of copyright.

9. He asks his publisher: "Who is this copyright intended to protect?" His publisher replies: "You!"

*What's wrong with this picture?*

(And why is the mother of the PhD whose give-away work people cannot steal, even though he wants them to, in the same boat as the mother of the recording artist whose non-give-away work they can and do steal, even though he does not want them to?)

## **BOX 2**

### **Some Relevant Chronology and URLs**

*Psychology* (Refereed On-Line-Only Journal) (1989)

<http://www.cogsci.soton.ac.uk/psychology>

"Scholarly Skywriting" (1990)

<http://cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.skywriting.html>

Physics Archive (1991)

<http://arxiv.org/>

"PostGutenberg Galaxy" (1991)

<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad91.postgutenberg.html>

"Interactive Publication" (1992)

<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad92.interactivpub.html>

Self-Archiving ("Subversive") Proposal (1994)

<http://www.arl.org/scomm/subversive/toc.html>

"Tragic Loss" (Odlyzko) (1995)

<http://www.research.att.com/~amo/doc/tragic.loss.txt>

"Last Writes" (Hibbitts) (1996)

<http://www.law.pitt.edu/hibbitts/lastrev.htm>

NCSTRL: Networked Computer Science Technical Reference Library (1996)

<http://cs-tr.cs.cornell.edu/>

University Provosts' Initiative (1997)

<http://library.caltech.edu/publications/ScholarsForum/>

*CogPrints*: Cognitive Sciences Archive (1998)

<http://cogprints.soton.ac.uk/>

*Journal of High Energy Physics* (Refereed On-Line-Only Journal) (1998)



<http://jhep.cern.ch/>

Science Policy Forum (1998)

<http://www.sciencemag.org/cgi/content/full/281/5382/1459>

American Scientist Forum (1998)

<http://amsci-forum.amsci.org/archives/september98-forum.html>

OpCit: Open Citation Linking Project (1999)

<http://opcit.eprints.org/>

E-biomed: Varmus (NIH) Proposal (1999)

<http://www.nih.gov/about/director/pubmedcentral/pubmedcentral.htm>

Open Archives Initiative (1999)

<http://www.openarchives.org/>

Cross-Archive Searching Service (2000)

<http://arc.cs.odu.edu/>

Eprints: Free OAI-compliant Eprint-Archive-creating software (2001)

<http://www.eprints.org/>

Harnad Home Pages

<http://cogsci.soton.ac.uk/harnad/>

<http://www.princeton.edu/~harnad/>

## Author Details

**Stevan Harnad, Les Carr, Tim Brody**

[Intelligence/Agents/Multimedia Research Group](#)

[Electronics and Computer Science Department](#)

Southampton University

Highfield, Southampton

United Kingdom SO17 1BJ

<http://cogsci.soton.ac.uk/>

<http://cogsci.soton.ac.uk/>

For citation purposes:

Stevan Harnad, Les Carr, Tim Brody, "How and Why To Free All Refereed Research From Access- and Impact-Barriers Online, Now", High Energy Physics Libraries Webzine, issue 4, June 2001  
URL: <<http://library.cern.ch/HEPLW/4/papers/1>>