Maximizing Research Impact
Through Institutional and National
Open-Access Self-Archiving Mandates

Stevan Harnad
Canada Research Chair in Cognitive Sciences Université du Québec à Montréal
http://www.crsc.uqam.ca/

and

Department of Electronics and Computer Science University of Southampton
http://www.ecs.soton.ac.uk/~harnad/

Summary
No research institution can afford all the journals its researchers may need, so all articles are losing research impact (usage and citations). Articles made “Open Access,” (OA) by self- archiving them on the web are cited twice as much, but only 15% of articles are being spontaneously self-archived. The only institutions approaching 100% self-archiving are those that mandate it. Surveys show that 95% of authors will comply with a self-archiving mandate; the actual experience of institutions with mandates has confirmed this. What institutions and funders need to mandate is that (1) immediately upon acceptance for publication, (2) the author’s final draft must be (3) deposited into the Institutional Repository. Only the depositing needs to be mandated; setting access privileges to the full-text as either OA or Restricted Access (RA) can be left up to the author. For articles published in the 93% of journals that have already endorsed self-archiving, access can be set as OA immediately; for the remaining 7%, authors can email the eprint in response to individual email requests automatically forwarded by the Repository.

1 Preamble

When Harold Varmus's very timely and influential 1999 Ebiomed Proposal (a pot-pourri of ideas about publishing, journals, archiving, peer-review, and what would eventually come to be called "Open Access” or “OA”) (Bailey 2006) managed to elicit staunch opposition from its foes and constructive criticism from its friends -- but very little in the way of actual OA -- it led to the creation of the Public Library of Science (PLoS), whose first action was to launch an Open Letter, signed by 34,000 biologists worldwide, threatening to boycott their journals – i.e., to cease pub-
lishing in or refereeing for them -- unless by September 2001 they began to make their contents OA (within 6 months of publication).

Now suppose that -- in addition to performing the keystrokes required to sign the 2001 PLoS Open Letter (pledging to boycott journals unless they become OA journals), each of the 34,000 PLoS signatories had also performed (or deputized a librarian, secretary or student to perform for them) the few further keystrokes it would have taken to make just one of their own year-2001 articles OA by self-archiving it, free for all, on the web (Harnad 1978, 1990, 1991, 1995, 2003; 2006). The number of OA articles (34,000) resulting from just that minimal act would already represent 60% of the approximately 55,000 Biology articles indexed by ISI in 2001; it would also have exceeded twice the total number of articles published by both BioMed Central and PLoS journals from 2001 to the present day (c. 16,000). And all at the cost of only a few keystrokes more per article than what it cost to sign the PLoS petition.

Yet the only thing researchers did in 2001 was to sign the PLoS Open Letter demanding that their journals should give them OA. They then waited, passively, for the journals to comply with their demand for OA. Most journals did not comply directly; of the 24,000 peer-reviewed journals that exist in 2006, only about 2000 of them (less than 10%) have converted to (or already were) OA ("gold") journal (Harnad et al. 2004). However, since 2001, in response to researchers' expressed wish for OA, over 90% of journals have given their authors their "green light" to self-archive their own articles online to make them OA if they wish (Figure 1). Yet today most researchers still seem ready to keep on waiting, passively, for more OA journals to be created or converted, one by one. Meanwhile spontaneous self-archiving continues to hover at about 5-25%, depending on the field and year.
There seems to be a note of inconsistency in this. Researchers feel they need and want OA badly enough to demand it from their journals, even threatening (rather idly, as it turns out to have been a bluff) to stop submitting to and peer-reviewing for the journals that decline to give them the OA they need and want so much. The needing and wanting have an unassailable objective basis because the benefits of OA are clearly demonstrated by the objective evidence of the dramatic citation impact advantage provided by OA, so: But is there an equally unassailable subjective basis, if the needing and wanting are not sufficient to induce researchers to do (or delegate) for themselves the few keystrokes that are the only thing standing between them and 100% OA?

Researchers themselves have hinted at the answer: Yes, they need and want OA. But there are many other demands on their time too, and they will only perform the requisite keystrokes if their employers and/or funders require it, just as it is already their employers and funders who require them to do the keystrokes to publish (or perish) in the first place. It is employers and funders who set researchers' priorities, because it is employers and funders who reward researchers' performance (Diamond 1986; Garfield 1988). Today, although only about 15% of research is being self-archived spontaneously, 95% of researchers sampled report that they would self-archive if required to do so by their employers and/or funders: 81% of them willingly, 14% reluctantly; only 5% would not comply with the requirement (Swan & Brown 2005; Figure 2). And in the four
objective tests of this self-reported prediction so far, all four institutions that have mandated self-archiving have fully confirmed it, with their self-archiving rates well above the spontaneous 15% baseline rate and firmly on the road toward 100% (Southampton-ECS, Queensland University of Technology, U. Minho and CERN).

So an employer/funder self-archiving mandate is obviously what is missing. But what exactly needs to be mandated? Only the keystrokes for depositing the final draft of the article (plus its bibliographic OAI metadata) in the author's Institutional Repository (IR) (Swan et al. 2005) immediately upon acceptance for publication are required. Going on to set access-privileges to the article as "OA" (full-text access open webwide) is merely encouraged. Access to over 90% of these articles can already be set to OA with the blessing of their publishers. The rest can be restricted to IR-internal access (for institutional employees, employers and funders) for the time being, but their bibliographic metadata (author, title, journal, date, abstract, keywords) will still be as visible to all searchers and surfers webwide as those of the 90% that are already OA, allowing would-be users to email the author to request an eprint. Emailing eprints can bridge the gap until either the remaining non-green journals give self-archiving their blessing or the author tires of doing the superfluous keystrokes to email the eprints and simply does the last keystroke to set access at OA. Either way, mediated OA will already be providing effective 100% OA as of the implementation of the keystroke-policy.

Such an immediate-deposit ("keystroke") policy -- leaving no loopholes for any exceptions or delays -- is what Research Councils UK (RCUK) has been recommended by the UK Selective
Committee on Science and Technology to mandate. The rest of the planet will follow suit. And Nature will take care of the rest.

2 Research Access and Impact

The 24,000 peer-reviewed journals (and conference proceedings) that exist today publish about 2.5 million articles per year, across all disciplines, languages and nations. No university or research institution anywhere, not even the richest, can afford to subscribe to all or most of the journals that its researchers may need to use (Odlyzko 2006). Hence no article is accessible online to all of its potential users webwide; and hence all articles are currently losing some of their potential research impact (usage and citations). This means that in the online era both the rate and the scale of research progress are less than what they could be.

This is confirmed by recent findings, independently replicated by many investigators, showing that articles for which their authors have supplemented subscription-based access to the publisher’s version by self-archiving their own final drafts free for all on the web (“Open Access,” OA) are downloaded and cited twice as much across all 12 scientific, biological, social science and humanities disciplines analysed so far (Lawrence 2001; Brody & Harnad 2004; Hajjem et al. 2005; Moed 2005b; Kurtz & Brody 2006). (Note: no discipline fails to benefit from self-archiving (Figure 3), they differ only awareness of OA and its possibilities.)
In all disciplines, articles (within the same journal issue) that are self-archived have more citations than those that are not. (Hajjem et al. 2005)

The total citation counts for articles submitted to the UK Research Assessment Exercise (RAE) -- which ranks all UK universities every four years according to their research performance, and funds them proportionately -- are also very closely correlated with the RAE ranking outcomes despite the fact that citations are not directly counted by the RAE. A higher number of citations is correlated with a higher RAE ranking (Smith & Eysenck 2002; Harnad et al. 2003). Hence citation counts are (i) robust indicators of research performance (Garfield 1973, Moed 2005a), (ii) they are not currently being maximised for those articles that are not self-archived and (iii) those articles that are being self-archived have a substantial competitive advantage over those that are not.

3 Institutional Self-Archiving Mandates Maximize Research Impact

Only 15% of the 2.5 million articles published annually are being spontaneously self-archived worldwide today. Creating an Institutional Repository (IR) and encouraging staff to self-archive
their articles therein is a good first step, but it is not sufficient to raise the self-archiving rate appreciably above the 15% baseline for spontaneous self-archiving. Adding library help to encourage and assist staff to self-archive raises the self-archiving rate somewhat, but insufficiently. (Sale, 2005)

The correct measure of institutional success in self-archiving is the ratio of annual self-archived articles in an institution’s IR relative to that institution’s total annual article output. The only institutions that are reliably approaching a 100% annual self-archiving rate today are those that not only create an IR and provide library help for depositing, but also adopt a self-archiving policy requirement or mandate (2006a, b, c).

A self-archiving mandate is a simple and natural extension of institutions’ already existing mandate to publish research findings (“publish or perish”); it is already linked to incentives (Waaijers 2006) by the fact that staff are promoted and funded on the basis of research performance indicators, of which citation impact is a prominent correlate, as in the RAE.

As noted above, two international, cross-disciplinary JISC surveys have found that 95% of authors will comply with a self-archiving mandate (81% willingly, 14% reluctantly). The four institutions worldwide that have adopted a self-archiving mandate to date (CERN in Switzerland, Queensland University of Technology in Australia, Minho University in Portugal, and the ECS Department at University of Southampton) have each confirmed the outcome of the JISC author surveys, with their institutional self-archiving rates reliably climbing toward 100%, whereas institutions without mandates remain at the 15% spontaneous self-archiving baseline rate.

4 Action: All research-active institutions and research funders should now mandate self-archiving

All research institutions should now maximise their own research impact and set an example for the rest of the world by adopting a self-archiving mandate (Sale 2006c). Research funders -- both governmental (Suber 2006) and private (Terry & Kiley 2006) -- should reinforce this by mandating that the research they fund must be self-archived in the fundee’s IR as a condition of the grant.

As indicated by the JISC survey and the empirical experience of the other 3 mandating institutions: there is no need for any penalties for non-compliance with the mandate; the mandate (and its own rewards: enhanced research access and impact) will take care of itself.

What needs to be mandated: The author/fundee,

1. immediately upon acceptance for publication,
2. must deposit into the university’s Institutional Repository
(3) both the full-text and the bibliographic metadata (author, date, title, journal, etc.)

(4) of the final accepted draft (not the publisher’s proprietary PDF)

(Note that only the depositing itself needs to be mandated. Setting the access privileges to the full-text can be left up to the author, with Open Access strongly encouraged, but not mandated. This makes the university’s self-archiving mandate completely independent of publishers’ self-archiving policies.)

The IR software then allows authors to choose to set access as Open Access (OA) or Restricted Access (RA):

OA: both metadata and full-text are made visible and accessible to all would-be users web-wide

RA: metadata are visible and accessible web-wide but the full-text is not

The decision as to whether to set full-text access as OA or RA can be left up to the author; 93% of authors will immediately set full-text access as OA; for the remaining 7%, the Eprints software still makes it possible for any would-be user web-wide to request an eprint of the full-text automatically by email -- by just cut-pasting their own email address into a box and clicking; the author immediately receives the request and can instantly email the eprint with one click. The result will be 100% access to all university research output, 93% immediately and directly, with one keystroke, 7% indirectly after a short delay, with a few extra keystrokes by user and author.

5 The Importance of Prompt Action

Research institutions and funders should not delay in adopting self-archiving mandates: Self-archiving is effortless, taking only a few minutes and a few keystrokes (Carr & Harnad 2005); library help is available too (but hardly necessary). 100% OA is both optimal and inevitable -- for research, researchers, their universities, their funders, and the tax-paying public that supports both the research and the universities. It will also give early adopters a strong competitive impact-advantage over later adopters.

With their self-archiving policy, early adopters are not only providing a model for emulation be the rest of the research world but at the same stroke they are maximizing their own research impact and research impact ranking. Institutional mandates need have no penalties or sanctions in order to be successful; they need only be formally adopted, with the support of departments, the
library, and computing services. The rest will take care of itself naturally of its own accord, as the experience of Southampton ECS, Minho, QUT and CERN has already demonstrated.

The OA Impact Advantage (currently 50-250%) will of course shrink as OA approaches 100%. Right now we are at about 15% OA self-archiving and the advantage is in part (no one can say how large a part) a competitive advantage of the minority 15% OA self-archivers (the head-start vanguard) over the laggard 85% non-OA majority. That makes it partly a race; and clearly, the race is to the swift and the battle to the strong. The competitive advantage is more reason for an individual, institution or nation (like the UK) to self-archive right now (as the RCUK will, we hope, soon be doing).

6 The Anatomy of the OA Impact Advantage: EA + (AA) + (QB) + QA + (CA) + UA

The OA impact advantage arises from at least the following 6 component factors, three of them temporary (2,3,5), three of them permanent (1,4,6):

1. **EA: EARLY ADVANTAGE**, beginning already at the pre-refereeing preprint stage. Research that is reported earlier can begin being used and built upon earlier. The result turns out to be not just that it gets its quota of citations sooner, but that quota actually goes up, permanently. This is probably because earlier uptake has a greater cumulative effect on the research cycle.

2. **(AA): ARXIV ADVANTAGE**, the special advantage of self-archiving specifically in ArXiv for physicists, because it is a central point of call: OAI-interoperable Institutional Repositories are likely -- for many reasons -- to supersede this, so it will eventually make zero difference which OAI-compliant IR one deposits in, as access will be through OAI cross-archive harvesters, not directly through individual OAI Archives.

3. **(QB): QUALITY BIAS**, arising from article/author self-selection; this does not play a causal role in increasing impact: The higher-quality (hence also higher-impact) articles/authors are somewhat more likely to be self-archived/self-archivers in these early (15%) days of self-archiving: this bias will of course vanish as self-archiving approaches 100%.

---

1 Actually, 8% more is OA too, via OA journals, but as the impact advantage is harder to calculate for OA journals -- because we are not comparing within the same journal and year -- we leave it out of these calculations. The same reasoning applies, however.
4. **QA: QUALITY ADVANTAGE**, allowing the high-quality articles to compete on a level playing field, freed of current handicaps and biases arising from access affordability differences. A permanent effect.

5. **(CA): COMPETITIVE ADVANTAGE**, for self-archived papers over non-self-archived ones, in early (15%) days; this too will of course disappear once self-archiving nears 100%, but at this moment it is in fact a powerful extra incentive, for the low % self-archiving fields, institutions and individuals.

6. **UA: USAGE ADVANTAGE**: OA articles are downloaded and read at least twice as much. This too is a permanent effect. (There is also a sizeable correlation between early download counts and later citation counts (Brody, Harnad & Carr 2005).)

Of these six component factors contributing to the OA impact advantage, only EA, QA, and UA remain operative in the few fields that are already close to 100% OA, such as Astrophysics and High Energy Physics. Everywhere else, however, the current 15% self-archiving rates still need to do a lot of climbing to reach 100%; so for those individuals, institutions, fields and nations the CA still matters a great deal today.

The UK, being the only country currently contemplating a nation-wide self-archiving mandate thereby stands to gain the biggest competitive advantage by being the first to do so. I have estimated that the UK’s gain in research impact would be the equivalent of having invested £1.5bn more into funding research (Harnad 2005). Have I overestimated this advantage in the longer-term, given the likelihood that other countries will follow suit, thereby cutting down on the CA component? It was partly to minimise this that I based my estimate on the lower end of the 50-250% OA impact advantage, underestimating it by using 50%. (It could also be 5 times as great.)

And whereas the Competitive Advantage will indeed shrink and disappear, the Early Advantage, Quality Advantage and Usage Advantage will be going strong. Kurtz et al (2004a, b; Kurtz & Brody 2006) have shown that although articles in a 100% OA field (Astrophysics) do not have longer reference lists, hence do not cite more articles overall, they do have three times higher usage rates (UA). So authors can at last find, access, and decide which articles to cite purely on the basis of their relative merit and quality (QA), no longer biased by the affordability (hence the accessibility) of the journal in which they happen to be published. So whereas the competitive horse-race (for who self-archives to gain the CA first) will be over at 100% OA, the cognitive horse-race (for which researcher finds what earlier: EA) will continue to favour the swift and the strong.

It is hence fair to say that although the annual £1.5 billion pounds-worth of potential impact that the UK is currently losing because it only self-archives 15% of its research output will shrink (as other nations' self-archiving policies catch up), how much it shrinks will then depend only on the true merit of British research rather than either the UK’s head-start in self-archiving or the current differential affordability/accessibility of journals.
APPENDIX: Southampton University Resources for Supporting Open Access Worldwide

1. U. Southampton ECS department was the first department and institution in the world to adopt a self-archiving mandate (2001).


3. ECS designed the first and most widely used software for creating institutional archives (Eprints, 2000), now already used by about 200 institutions worldwide; ECS also created Citebase (2002), the citation-based OA search engine (well before Google Scholar).

4. ECS conducted many of the seminal studies empirically demonstrating the citation impact advantage of self-archiving across all disciplines; ECS also maintains the growing and widely used bibliography of the accumulating findings on the OA Impact Advantage.

5. ECS/Eprints maintains ROAR, the Registry of Open Access Repositories, tracking the number, size and growth of IRs and their contents worldwide.

6. ECS/Eprints maintains ROARMAP, the Registry of Open Access Repository Material Archiving Policies, tracking the institutions worldwide that have adopted self-archiving policies, from recommendations to full mandates.

7. ECS/Eprints maintains the ROMEO Directory of Journal Policies on Author Self-Archiving: 93% of the nearly 9000 journals registered to date (including all the principal publishers and the core ISI journals) have already formally endorsed author self-archiving; only 7% of journals have not.

8. ECS/Southampton successfully lobbied the UK Parliamentary Select Committee in 2004 to mandate self-archiving; this led directly to the RCUK self-archiving mandate proposal, the Berlin 3 Policy Recommendation (formulated at Southampton) and the development of RAE submission mechanisms for the world’s two principal IR softwares (Eprints, and MIT’s Dspace, both written by Southampton’s Rob Tansley).

REFERENCES


Moed, H. F. (2005a) Citation Analysis in Research Evaluation. NY Springer.


