

Advancing Science By Self-Archiving Refereed Research

Stevan Harnad

Southampton University, UK

The Editor of SCIENCE, Dr. Floyd Bloom has written an editorial about NIH's E-biomed initiative (<http://www.nih.gov/welcome/director/ebiomed/ebiomed.htm>)

Floyd E. Bloom [Editorial] "Just a Minute, Please" SCIENCE 285 (5425) p. 197, 9 Jul 1999 (<http://www.sciencemag.org/cgi/content/summary/285/5425/197>)

This is a reply to his editorial. To summarize, Dr. Bloom is writing ex officio as the Editor of Science, published by the American Association for the Advancement of Science. Science is a hybrid journal. It contains articles by salaried staff writers and commissioned articles written for a fee. It is important to note that these articles are in no way at issue here.

But Science also contains refereed research reports, submitted by their authors for free, with the sole objective of making the research findings available as broadly as possible once they have met Science's rigorous standards of peer review. It is these refereed articles only that are under discussion here, and the issue is a simple one: Should NIH/E-biomed provide a free public Archive, modelled on the NSF-supported Los Alamos Eprint Archive in Physics (LANL) (<http://xxx.lanl.gov/>), in which the authors of these refereed research reports can self-archive them online publicly, free for everyone, everywhere, forever?

Dr. Bloom is arguing that they should not be, and we will shortly examine his reasons. But we can be confident that Dr. Bloom will revise his views when more fully informed of the objectives of E-biomed and the scientific potential of free public archiving of refereed research on the World Wide Web, for Dr. Bloom represents the American Association for the Advancement (not the secondary sale or suppression!) of Science.

At the moment, Dr. Bloom's reservations are motivated by two factors: Concern about the quality of the scientific research literature (and this concern is commendable, his journal being the representative of research standards of the highest quality) and concern about the revenue stream of his journal, which is the financial resource that is currently supporting those high standards of quality. It is here that I am afraid that Dr. Bloom is being somewhat short-sighted and perhaps even a little partisan too, unconsciously placing the interests of the maintenance of that revenue stream above the interests of the science that AAAS is dedicated to advancing.

It is undeniable that in the present PostGutenberg Era a conflict of interest has arisen between researchers and the current means of production of their published refereed research reports. There is a way to resolve this conflict, however, although it at first appears counterintuitive; and as the resolution is clearly to the benefit of science, and at the same time provides the revenue stream to sustain the essential service provided by the publishers of science -- quality control and certification in the form of peer review and editing -- there is every reason to believe that AAAS will find it fully compatible with its mission.

The resolution is a two-stage one.

First, it is necessary to identify and acknowledge the conflict of interest:

For scientific researchers, the reports of their (usually publicly funded) research findings are GIVE-AWAYS: They seek neither royalties nor fees; they seek only the eyes and minds of their fellow-researchers worldwide, present and future, so as to maximize the impact of their findings on the future course of research (and thereby also on the course of their careers and their livelihoods).

Researchers are accordingly interested in having their findings first quality-controlled and certified (QC/C) through peer review, and then made freely accessible to everyone. In the Gutenberg Era, the only way they could come anywhere near that goal was to treat their work exactly the same way trade authors (who wrote for fees or royalty) treated theirs, namely, to assign copyright to a publisher, who would then charge for access to the work in order to cover the substantial expenses of paper publication and distribution and to make a fair profit, where possible, for both himself and his author (in the form of royalties or fees).

But the scientists reporting their research findings in refereed journals were never interested in fees or royalties, for those would represent access barriers, restricting their findings to only those individuals and institutions that could afford to pay for them [via subscription, site license, or pay-per-view (S/L/P)].

Nevertheless, scientists had to live with these S/L/P barriers, for all the world as if they were trade authors seeking royalties or fees for their work, because in the Gutenberg Era there simply was no alternative way to reach even that privileged subset of the potential readership of their article (not a large populace even in the best of times).

In the PostGutenberg era of global digital networks, however, there is at last an alternative, and not only researchers, but research itself, and hence all of society, would be the losers if we failed to take full advantage of it. For now we no longer have to rely on the expensive, inefficient and access-limiting technology of print on paper to disseminate refereed research findings. They can be SELF-ARCHIVED by their authors in public online archives like E-biomed (and its spectacularly

successful model, LANL) and thereby accessible to one and all without any financial firewalls.

<http://www.arl.org/sc/subversive/i-overture-the-subversive-proposal.shtml>

<http://www.cogsci.soton.ac.uk/~harnad/nature.html>

Free public self-archiving, however, is only the first of the two stages of resolving the conflict of interest between research and its current means of publication. As long as there continues to be a demand for the paper version, it (and its proprietary online counterpart) can continue to be sold via S/L/P, which can continue to fund (among other things) QC/C (peer review). But meanwhile the worldwide research community will also have the self-archived online version on its desktops for free. And there is every reason to believe that they will grow increasingly reliant on it.

http://xxx.lanl.gov/cgi-bin/show_weekly_graph

Eventually, this is likely to shrink S/L/P revenues, and here it may look as if we are approaching a catastrophe point, for part of that revenue is paying for the maintenance of the quality standards of that literature. But a very simple solution is available, once we recognize that the S/L/P revenues are largely being paid for by their researchers' institutions. Let us call this "reader-institution end" funding. All that is needed to continue covering QC/C costs is to switch from reader-institution end funding to author-institution end funding, covered fully by the S/L/P savings. The big difference is that reader-institution-end S/L/P is access-blocking, holding the literature hostage to access tolls, whereas author-institution end funding makes access completely free.

This is the second stage of the resolution of the conflict of interest, and it has the further advantage (although this is more controversial, because no one has the exact figures yet) that it will save institutions a great deal of money. For the cost of QC/C alone -- once publishers have scaled down to providing this essential service alone, leaving the access providing and preservation entirely to public online archives like LANL and E-biomed -- is likely to be much lower than current S/L/P expenditure. Indeed it is likely to be less than 1/3 of it, by current estimates.

See the American Scientist Discussion Forum threads on this:

<http://amsci-forum.amsci.org/archives/American-Scientist-Open-Access-Forum.html>

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/economics.journals.txt>

This means that researchers benefit (access to their findings is expanded, potentially infinitely), their institutions benefit (both from S/L/P savings and from their researchers' enhanced impact), and research itself benefits (in both productivity and pace). Refereed journal publishers will unfortunately need to downsize, but in exchange they will have a stable, permanent niche that is compatible with the new medium rather than at odds with it.

Now I proceed to reply to Dr. Bloom's editorial on a quote/comment basis:

Proponents [of the E-biomed Archive] acknowledge that cooperating journals could lose subscription income and suggest that journals recover their costs through submission and acceptance fees charged to authors. E-biomed may be free to users, but it will not be free to taxpayers or authors submitting through peer review.

We can now understand that this passage is based on a misunderstanding. Tax payers are already sustaining our educational and research institutions, including their S/L/P budgets, which will be REDUCED rather than increased by the switch to up-front payment in the online-only era.

And the costs of providing public research archive facilities such as LANL and E-biomed will be minuscule compared to the size of the literature and the benefits conferred; moreover, most of the infrastructure is in place already, in this increasingly networked world, and pooling resources with the rest of the disciplines (after Physics and the Biomedical Sciences) will make the marginal costs even more minimal.

So there is nothing whatsoever in this passage to deter us from resolving this conflict of interest in the way just described.

[E-biomed has] much support from quarters long known to advocate a more open scientific literature that would banish the alleged cabals of editors, biased reviewers, and expensive commercial presses with generally irrelevant content.

There are as always extremists around who want to banish QC/C, but leveller heads are bent on preserving it, and indeed the entire scenario just described is predicated on just that.

<http://helix.nature.com/webmatters/invisible/invisible.html>

So this objection too is invalid.

Lurking behind the public discussions are some potentially troubling elements:

What if the major journals choose not to cooperate out of concern that their ability to survive and maintain quality control and timeliness are threatened by the diversion of authors and competent reviewers into the NIH system?

There was a little confusion in the initial draft of the E-biomed proposal. The eventual goal is cooperation with the refereed journals, in the form of official "overlays" on the archive, authenticated by them. But in the first stage, author self-archiving of their refereed drafts will suffice to free the literature.

<http://www.nih.gov/welcome/director/ebiomed/com0509.htm#harn45>

Nor is there any "diversion of authors and competent reviewers into the NIH system." There is no "NIH system," merely a public archive in which authors can deposit their papers (both refereed reprints and, if they wish, unrefereed preprints).

There is only one respect in which the major journals need to "cooperate," and one certainly hopes they will do so, otherwise this will escalate the conflict of interest instead of resolving it to the benefit of science: Publishers must not attempt to use copyright restrictions as a weapon to continue to hold the literature hostage to access tolls by forbidding public self-archiving.

This is THE central issue, and at the heart of all of this. SCIENCE itself has published a collective call for the retention of such author rights

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

along with a dissenting editorial by Dr. Bloom.

<http://www.sciencemag.org/cgi/content/summary/281/5382/1451>

Some prior comments on that exchange in SCIENCE are appended at the end of this reply. Let it only be noted here that progressive publishers are already resolving this conflict in a fair and rational way, in the interests of the scientific community they serve, rather than their own S/L/P revenue streams. A model in this regard (and they will be duly recognized by historians for this) is the American Physical Society (APS), publisher of the journals with the highest QC/C standards and impact factors in Physics. Dr. Bloom's homonymous APS counterpart, Dr. Blume, is one of the cosignators of the above copyright reform proposal in SCIENCE. For APS copyright policy, see:

<http://www.cogsci.soton.ac.uk/~harnad/Hypermail/Author.Eprint.Archives/>

Will societies whose members' future careers rely on NIH funding be willing to resist the cooptation of their journals' editorial and peer review systems?

Nothing is being co-opted. The NSF-funded LANL Physics Archive stands as a model for the kind of cooperative solution that will prevail. Journals, editorial boards and peer review will continue to exist, independent and intact. The only issue is whether they should be allowed to continue to try to hold this give-away literature hostage to S/L/P access tolls, against the interests of research and researchers.

What will the real costs be to authors, peer-reviewed journals, and scientific societies?

Yes, what will they be indeed, once the obsolete Gutenberg "add-ons" are phased out and only the essential QC/C costs remain?

Does a monopolistic archive under government control by the major research funder enhance scientific progress better than the existing journal hierarchy, which provides multiple alternatives to authors and readers?

Multiple journals -- indeed the entire hierarchy that currently exists -- will continue to exist for authors and readers. Nor will it be government controlled. (As always, quality will be controlled by peer reviewers, who, like the authors, do their work for free! QC/C costs are for IMPLEMENTING peer review, not for actually performing it.)

NIH will fund E-biomed, just as NSF funds LANL. The cost will be minuscule, and still smaller as more disciplines join in the self-archiving initiative. And once S/L/P expenditures shrink, savings will prevail, including savings on government-supported institutional serial budgets.

Pluralism will be, if anything, enhanced by a firewall-free global research literature. The objective is to free the literature from market restrictions that are no longer justified or necessary, not to take over a market!
(The word "monopoly," so clearly out of place here, will recur later in this reply in the context of certain collaborative firewall practises on the part S/L/P providers...)

What about research in disciplines outside what the National Library of Medicine considers biomedical?
There are plans for vetting the unrefereed clinical preprint sector to safeguard public health, but no planned restrictions of any sort on the refereed sector, any more than there are any such restrictions on the LANL Archive. (One wonders what Dr. Bloom has in mind here?)

What about research not sponsored by NIH or even U.S. federal funds?

The answer to this question is so obvious, one can only wonder why it was raised: What about research not sponsored by NSF in LANL? What about LANL's 14 mirror sites around the world? Why on earth would an archive dedicated to freeing access to the refereed research literature for the world scientific community through self-archiving have any interest in blocking access to any of it? (The only interests vested in blocking access to this corpus at the moment are certainly not governmental ones...)

Without answers to these and other questions, it is hard to determine the feasibility of the proposal.

(The answers are in each case so trivially obvious that one can only wonder what the real source of the reservations about feasibility might be!)

Science and other journals are eager to identify the advantages of the E-biomed proposal and are actively looking for changes that could benefit scientific publishing.

The advantage of the E-biomed proposal is that it will free the refereed journal literature, to the benefit of science, scientists, and humankind. The only change required at the moment is a copyright policy that clearly recognizes the no-royalty/no-fee author's right to self-archive along the lines of the APS policy.

<http://www.cogsci.soton.ac.uk/~harnad/Hypermail/Author.Eprint.Archives/0006.html>

For example, the E-biomed server would provide a venue for online publication of negative results and thus allow others to avoid experimental repetition.

Among the much more profound benefits of public online self-archiving of refereed reprints and unrefereed preprints there is also the more modest one of being able to self-archive negative results, both those that have been accepted by refereed journals, and those that were not.

On the other hand, if NIH really wants to improve access to the literature, they could digitize the peer-reviewed literature published before 1995.

The retroactive peer-reviewed literature is most certainly welcome in the free public archives, and will most certainly be deposited there, both by individual authors and by digitization initiatives (neither LANL nor E-biomed is a digitization initiative: they are public self-archiving initiatives).

But exactly what is the benefit to science of restricting availability to the pre-1995 literature?

In addition, all would benefit if NIH developed software for online journal submittals and provided access to a common search engine that could survey all peer-reviewed sciences across all journal lines.

The first benefit, though undeniable, is likewise not E-biomed's mandate. (Why should NIH develop submission software tools?) On the other hand, the practise of self-archiving will certainly help accelerate the development of such tools, and it will hasten and expand authors' using them. Moreover, once the second stage is reached, official journal overlays on E-biomed will allow automatic online submission to the journals via the archive, as is already being implemented on LANL in collaboration with the APS.

As to providing the capacity to "survey all peer-reviewed sciences across all journal lines," this will be trivially provided by E-biomed and any number of generic search engines as soon as the self-archiving initiative is well under way, and E-biomed is well stocked with papers searchably tagged as "U" (unrefereed preprint) or "R" (refereed reprint, together with journal name "Jx").

But the principal advantage of this free public archive will be that it will indeed be "across all journal lines" without any of

the financial firewalls that criss-cross the proprietary online corpus as it now stands -- a state of affairs that some would like to see turned into a "click-through" monopoly governed by interpublisher S/P/L fee agreements!

<http://www.cogsci.soton.ac.uk/~harnad/citation.html>

It may be instructive to recall an earlier congressional reaction, as Albert Henderson, editor of Publishing Research Quarterly did in his response to E-biomed on 6 May. In the Sputnik aftermath, an E-biomed-like proposal was made that Congress accelerate U.S. scientific research by establishing a unified information system similar to what had been created in the Soviet Union. The Senate's advisory panel responded: "The case for a Government-operated, highly centralized type of center can be no better defended for scientific information services than it could be for automobile agencies, delicatessens, or barber shops." Surely other creative solutions can be found to what NIH considers problems. Are they prepared to listen, or is this a done deal?

Both Dr. Henderson and Dr. Bloom might benefit from being reminded (if they will only listen!) that unlike the producers of cars, bagels and haircuts, the producers of refereed journal articles wish to give them away for free. And there is no earthly reason why any government should not wish to help them do so, to the eternal benefit of science and society worldwide.

This would have been as welcome in the Sputnik era as it is today, but we had not yet reached the PostGutenberg Galaxy at that time.

The only costs that remain to be paid are those for the SERVICE of implementing QC/C, costs that it will make incomparably more sense for the author-institution to pay up-front, out of S/L/P savings, thereby freeing the literature for one and all, along with a considerable institutional saving, rather than at the access-denying reader-institution end, for the reasons described above.