The self-archiving principle: a momentous trek

Nishith K Singh

Postgrad Med J 2007;**83**:564–567. doi: 10.1136/pgmj.2006.056887

In the existing scholarly publishing empire, authors give away their valued research work to various commercial journals, thereby restricting free accessibility to the published useful work. Triggered by the gargantuan promise of the internet, the self-archiving principle is a new and revolutionary concept which potentially lets all research work become freely available online. It involves deposition of research documents at a publicly accessible website, and its proponents see the initiative as a means to set entire author works free of all access and impact barriers. This review briefly discusses the allied concepts, the course and implications of the initiative.

In the present publishing world, authors and researchers give away their valuable work copyright in return for “value” gains. Their work is traded for claims of priority, precedence and presence and the earthly rewards of recognition by peers, citation, chance of career advancement, and procurement of grant funds.¹ Such a royalty-free model has been surprisingly sustained, as opposed to the royalty-based work in books. The publisher, on the other hand, seeks to exploit the sale value of an article by toll-gating the access.^{1,2} The impact and access of scientific work is, therefore, believed to be conceptually restricted. It is a widely held view that such power resting with the publishers is detrimental to the scholarly goals of scientific exchange. The open access (OA) movement aims to make all of the scholarly work freely available on the internet through the self-archiving tool and through submission in OA compliant journals. The central view of the OA movement is that the retention of copyright by authors would benefit scientific progress for the public good by permitting rapid distribution and exchange of information.

Self-archiving is a protocol, based on the internet, which has been developed with the aim of empowering authors with the ability to make their work free of all access barriers. The present review traces the goals and journey of the self-archiving principle since its inception, and comments on the changes induced by the gradual acceptance of the self-archiving concept in the existing byzantine scholarly publishing empire.

BIRTH OF THE INSURRECTION

What provoked the idea of self-archiving? The present scholarly publishing order can be conceptually divided into the authors and researchers, the librarians, the commercial publishers, the research funding agencies, and the taxpayer. For a long time, librarians have been disconcerted over the

ever rising journal subscription costs,³ and the funded scholarly work has been widely believed to have been impeded in a major way. One might also argue that taxpayer's money goes into funding research, and thus they should get a free ride into the published literature. On the other hand, the publishers' policies have long been haunted by the “Ingelfinger rule”, named after Franz J Ingelfinger, the editor of *New England Journal of Medicine*.^{4,5} The rule was originally invoked to boycott any submission of a manuscript for publication if the unrefereed pre-print has been self-archived (deposited) or published earlier. The idea behind such a proposition was to protect public health from the influence of unrefereed research work which could be potentially hazardous. This rule (which is also referred to as “prepublication embargo”), being only a journal policy and apparently conferring immunity to journal inflowing revenues, has faced major criticism, as author's valuable work is theoretically reduced in impact and accessibility.⁶ Proponents of self-archiving have long since dreamt of easily accessible peer-reviewed published work. The influence of the internet and the world wide web in shaping the communication interface and information dissemination provided the much needed platform for developing open access to information before publication.

The first online pre-print service, arXiv.org, was developed in 1991 by physicist Paul Ginsparg. It allowed thousands of scientists to share ideas over the internet before publishing, by posting their work pre-prints online. Three years later, in a paradigm shifting view posted on the internet by cognitive scientist Steven Harnad, a suggestion of immediate self-archiving of research papers by all authors was encouraged with the aim to forestall the publishers' control over manuscript access and improve impact of the research work. The proposal, known as the “subversive proposal”, urged institutions and researchers to post their works (pre-prints, peer-reviewed published post-prints, updates, etc) on the internet, making it accessible world wide, thus eliminating the access barrier and also boosting the impact.^{2,7} The intention, apparently utopian, is to free all peer-reviewed research articles from a barrier which would otherwise cease to diminish in a world driven by commercial prospects. This by no means connotes self-publishing; neither does it preclude a peer-review process. Quality control and certification of a published research work is a public health imperative and the unrefereed work would be

Correspondence to:
Dr Nishith K Singh,
Southern Illinois University
School of Medicine,
Department of Medicine,
PO Box 19636, Springfield,
IL 62794-9636, USA;
nishith_singh2007@yahoo.
com

Received 29 December
2006
Accepted 10 April 2007

tagged so as to differentiate it from a post-print refereed work. In the proposed model, the “obligatory” cost of peer-reviewing research work could be borne by the archiving entity (researchers, institutions, etc),² and the traditional publishers would provide the peer review expertise and let the author keep the copyright of his or her work. This is still profitable when taking into consideration the cost savings from cancelled journal subscriptions over time as the archive reaches a critical mass.

Harnad has further proposed a self-archiving kit for implementation, metatagging of information so that it is retrievable in a decentralised manner, free distribution of OA archive softwares, and an initial sponsor wave through universities and research libraries.⁸ Eprints (digital archives of files) are now being developed as a repository for tagged files. All repositories the world over will function as a single functional pool because of common metatagging standards and thus are interoperable. Self-archiving has thus been proposed as the implementing tool for OA which could have a quick and widespread effect, at least in theory.

THE EARLY EVENTS

In 1999, a life sciences version of arXiv.org, E-biomed, was suggested by Nobel Prize winner Harold Varmus, leveraging on the logical acceptance of the self-archiving initiative. The version was finally launched as PubMed Central, having overcome tough resistance from those publishers with commercial interests and the learned societies. As a result of the resistance, PubMed Central has limited the journals archived, even after 6 years following its conception. But the initial encouraging author response to free indexing at PubMed Central urged Varmus to create the Public Library of Science (PLOS) in November 2000.⁹ Authors and researchers had also raised concerns regarding the “authenticity” of self-archived material, the possibility of disruption of the present scholarly publishing infrastructure, and the copyright contravention by archiving of the post-prints on-line.¹⁰ Though such misgivings were not expected to last, the initial popularity of PubMed Central faded gradually and even some of the signatories (thousands of authors who pledged to shun publishers who refused free archiving of the published material) demonstrated

a lukewarm response.¹¹ It was slowly learned that the closely linked commercial links of the powerful biomedical publishing companies and research institutes posed a fierce resistance to any such movement. In yet another blow to the initial efforts to implement self-archiving, it was realised that disorganised self-archiving promotes scattering of information over several anonymous sites which are poorly accessible, thus amounting to loss of information.¹¹

But gradually over time, under the growing demand of OA advocates, the OA publishers started to come into being. Notable among them were Vitek Tracz’s BioMed Central and the restructured PLoS journals, the PLoS Biology and PLoS Medicine.⁹ These publishers publish articles for a charge from the author (the author-pay model) and in return the work is made freely accessible on the internet. The burden of expense thus switched hands, for want of a greater reach. But this was seen as opposing the idea of self-archiving by Harnad, as it gave rise to a new “breed” of publishers. It was a deviation from Harnad’s original concept, which had assumed that authors would continue to use traditional journals, but then self-archive their papers.¹¹ Also the commercial publishers and the librarians themselves, who had been at the forefront of the OA movement, expressed concerns regarding the economic viability of the new author-pay model.^{9 10} With the subsequent Budapest Open Access Initiative (BOAI) in 2001, the OA movement garnered a new look. Self-archiving gained ground as it was adopted as a “green road” to OA along with voluntary submitting to OA journals as the “golden road” to OA.^{9 12 13} In the following years the Bethesda and Berlin declarations, advocating OA, were announced. OA publishing gained more popularity as a result of the declarations, but self-archiving was only slowly catching up. Harnad believes OA journals have limited size and spread, whereas self-archiving can conceptually provide access to all scholarly works overnight.¹¹

TASTING SUCCESS

With the mounting pressure of congress budget policies and the OA advocates like PLoS, the National Institutes of Health (NIH) enacted its OA policy on 2 May 2005.¹⁴ The policy requests that authors of peer reviewed publications resulting from NIH funded research voluntarily submit a copy of their final manuscript to the National Library of Medicine (NLM) for archiving in PubMed Central, the NLM’s open access electronic repository for life sciences journals. Furthermore, the Wellcome Trust in the UK, which is a major research funding organisation, has also shown concordance with such moves by setting requirements, such as: all grantees awarded funds after 1 October 2005 must make their published results freely available in PubMed Central no later than 6 months after publication.^{15 16} The Research Councils UK has shown an inclination in this direction too and might be a major player in the near future.¹⁷ These shifts have been welcome among the self-archiving advocates as these accede to the archiving concept, but the response of the voluntary author and reader (these are overlapping tribes) over time has remained small to date.¹⁸ In June 2006, the US House of Representatives voiced strong support for the NIH view on the proposal of mandatory submission of all NIH funded research to PubMed Central by framing language included in the FY07 Labor HHS appropriations bill, directing NIH to make mandatory submissions from its grantees. The bill marks a new direction to the OA movement through self-archiving.

THE FUTURE: AS TIME KEEPS A WATCH

This story raises obvious questions. How serious is the threat to publishers? arXiv.org has been around for over a decade, eprinting scholarly articles in the disciplines of physics,

The triggered revolution

- Skyrocketing subscription costs of leading journals took its toll on librarians, limited access to research work, and forced scientists to look for alternate publishing options
- Computer and physics researchers have long been self-archiving pre-prints (pre-peer reviewed versions) of their scientific work on-line: <http://www.arxiv.org/>
- Self-archiving of pre-prints and published post-prints on publicly accessible websites was proposed by Steven Harnad in 1994 as an immediate solution to the access/impact
- Budapest Open Access Initiative (BOAI, 2001) adopted self-archiving as the “green road” to open access of all research work
- The National Institutes of Health (2005) and recently the UK Wellcome Trust’s self-archiving policy towards the goal of open access are standing testimony to the rising acceptance of this approach
- Prevailing low awareness and conflicting interests of the stakeholders might preclude a swift reform

Key references

- Harnad S. The self-archiving initiative [Erratum: *Nature* 2001;**411**:522]. *Nature* 2001;**410**:1024–5.
- Harnad S. Open access to peer-reviewed research through author/institution self-archiving: maximizing research impact by maximizing online access. *J Postgrad Med* 2003;**49**:337–42.
- Albert KM. Open access: implications for scholarly publishing and medical libraries. *J Med Libr Assoc* 2006;**94**:253–62.
- Steinbrook R. Public access to NIH-funded research. *N Engl J Med* 2005;**352**:1739–41.
- Terry R. Funding the way to open access. *PLoS Biol* 2005;**3**:364–6.

mathematics, computer science and quantitative biology. The publishing industry in those disciplines has not been adversely affected.¹⁰ How far is the OA movement going to convince and meet the needs of the scientific audience? In a reported telephone survey of authors submitting to the *BMJ*, it was found that among the respondents, there is a lack of an awareness regarding OA, and most consider perceived journal quality as a more important factor than open access when deciding where to submit papers.¹⁹ In another study, involving an electronic survey of authors of research papers submitted to the *BMJ*, *Archives of Disease in Childhood*, and *Journal of Medical Genetics* in 2004, it was found that authors had reservations regarding making a payment for publications and choosing OA over the quality of a journal.²⁰ The present day publishing world has a hierarchy of standards evolved within the system which subjects the articles to an intuitively conceptualised “quality skimming” process, and the reader has high acceptability of information from the high-standard journals. Proponents of copyright argue that such a system has certain blessings in terms of conferring protection against plagiarism, tagging accuracy with the cited data and regulating dissemination (for example, through bibliographical databases such as Medline). The copyedited, formatted article soaked in the journal’s proprietary content is very much valued by the readers. The mix-up of denotations in self-archiving, publishing and OA is bound to be a yet another roadblock to any major reform.

Is the self-archiving exercise finally able to answer these concerns and provide OA as it promises? How will it affect the future of the scholarly publishing world? The answers to all these questions are largely unknown but the present publishing model and the self-archiving movement is bound to coexist as opposing trends for some time. It is yet to be seen whether self-archiving is a friendly adjunct to, or if it will eventually mean Armageddon for, the existing publishing establishment.

Source(s) of support: None

Conflicting interest: None to declare.

Past affiliation: Department of Internal Medicine, All India Institute of Medical Sciences, New Delhi India

REFERENCES

- 1 Crawford BD. Open-access publishing: where is the value? *Lancet* 2003;**362**:1578–80.
- 2 Harnad S. The self-archiving initiative. The self-archiving initiative [Erratum, *Nature* 2001;**411**, 522]. *Nature*, 2001;**410**:1024–5.
- 3 Kyriolidou M. Journal costs: current trends & future scenarios for 2020. *ARL Bimonthly Rep* 2000;**210**.

- 4 Altman LK. The Ingelfinger rule, embargoes, and journal peer review—Part 1. *Lancet* 1996;**347**:1382–6.
- 5 Altman LK. The Ingelfinger rule, embargoes, and journal peer review—part 2. *Lancet* 1996;**347**:1459–63.
- 6 Germeis AE. Beyond the Ingelfinger Rule: the intellectual property ethics after the end of biomedical journals’ monopoly. *Med Inform Internet Med* 1999;**24**:165–70.
- 7 Harnad S. A subversive proposal. In: Okerson A, O’Donnell J, 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> (originally posted June 27 1994 <http://www.arl.org/scomm/subversive/sub01.html>).
- 8 Harnad S. Open access to peer-reviewed research through author/institution Self-Archiving: maximizing research impact by maximizing online access. *J Postgrad Med* 2003;**49**:337–42.
- 9 Albert KM. Open access: implications for scholarly publishing and medical libraries. *J Med Libr Assoc* 2006;**94**:253–62.
- 10 Alma Swan. Open access self-archiving: an introduction. Key Perspectives Ltd. (May 2005) www.ecs.soton.ac.uk/~harnad/Temp/jiscsum.pdf.
- 11 Poynder R. Ten years after. *Info Today*. 2004;**21**: 1, 23–4, 44, (9).
- 12 Budapest Open Access Initiative, authors. Budapest Open Access Initiative. Feb 16 2003). <http://www.soros.org/openaccess/read.shtml>.
- 13 Harnad S, Brody T, Vallieres F, et al. The access/impact problem and the green and gold roads to open access. *Ser Rev* 2004;**30**:310–4.
- 14 Steinbrook R. Public access to NIH-funded research. *N Engl J Med* 2005;**352**:1739–41.
- 15 Terry R. Funding the way to open access. *PLoS Biol* 2005;**3**:364–6.
- 16 Wellcome Trust. Wellcome Trust and National Library of Medicine in talks for worldwide open access archive (press release). http://www.wellcome.ac.uk/doc_WTX022826.html.
- 17 Vogel G, Enserink M. Information sharing. Europe steps into the open with plans for electronic archives. *Science* 2005;**308**:623–4.
- 18 Suber P. First fruits of the NIH public-access policy. *SPARC Open Access News* [serial online]. 2005 (88)..
- 19 Schroter S, Tite L, Smith R. Perceptions of open access publishing: interviews with journal authors. *BMJ* 2005;**330**:756.
- 20 Schroter S, Tite L. Open access publishing and author-pays business models: a survey of authors’ knowledge and perceptions. *J R Soc Med* 2006;**99**:141–8.

APPENDIX

CRUCIAL WEB LINKS

1. Self-Archiving FAQ. <http://www.eprints.org/openaccess/self-faq/#publisher-forbids>
2. Budapest Open Access Initiative. <http://www.soros.org/openaccess/>
3. arXiv. <http://www.arxiv.org/>
4. American Scientist Open Access Forum. <http://www.cogsci.soton.ac.uk/~harnad/Hypermail/Amsci/index.html>
5. Eprints.org: Journal Self-Archiving policies. <http://romeo.eprints.org/stats.php>
6. Ten years after <http://www.infotoday.com/it/oct04/poynder.shtml>
7. For Whom the Gate Tolls? <http://eprints.ecs.soton.ac.uk/8705/01/resolution.htm#1.4>
8. Open Archives Initiative. <http://www.openarchives.org/>

GLOSSARY

- **Author-pay models:** By “Author-pay” model, some publishers, such as BioMed Central and Public Library of Science (PLOS), rely on author fees to support the minimum cost online publication and then making it freely accessible on their websites (open access). Some believe that it has the potential to introduce bias by supporting the wealthier nations.
- **Budapest Open Access Initiative:** The Budapest Open Access Initiative arises from a small but lively meeting convened in Budapest by the Open Society Institute (OSI) on 1–2 December 2001. The purpose of the meeting was to accelerate progress in the international effort to make research articles in all academic fields freely available on the internet.

- **Eprints:** Eprints are the digital texts of peer-reviewed research articles, before (the pre-refereed “pre-prints”) and after refereeing (the refereed “post-prints”).
- **Internet:** The internet is the large network of electronically connected computers, interconnected in order to “decentralise” and thus make secure the sensitive information from being destroyed in face of an attack on one computer. The computers communicate using various protocols like TCP/IP etc. These protocols are rules of digital data transfer.
- **Open access:** As per the Budapest Open Access Initiative, “open access” for an article stands for free availability on the public internet, permitting any user to read, download, copy, distribute, print, search, or link to the full texts of article, trawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.
- **Open access journal (OA journal):** These are journals which use a funding model wherein an author or an institution is not charged for access to an on-line published work. As per the Directory of Open Access Journals (DOAJ), there are more than 2600 OA journals (<http://www.doaj.org/>) (accessed April 6, 2007).
- **PubMed Central (PMC):** PubMed Central is the US National Institutes of Health (NIH) *free* digital archive of biomedical and life sciences *full-text* journal literature. Participation in PMC is voluntary and is also accessible through the database PubMed (a service of US National Library of Medicine) which has millions of articles (citations from the Medline database) from thousands of journals worldwide, though not all of them are free.
- **Self-archiving:** Self-archiving stands for deposition of a digital document in a public accessed website. It is not always peer-reviewed and if peer-reviewed (refereed) the peer-reviewed material deposited is the *published* and refereed document and thus is not equivalent to a scholarly publication, nor means self-publishing (vanity press). Depositing involves a simple web interface where the depositor copy and pastes in the “metadata” (date, author-name, title, journal-name, etc) and then attaches the full-text document.
- **World wide web:** The world wide web is a huge collection of files which are interlinked with each other through the use of click-able “hyper-links”. It was conceived by Tim Berners-Lee at the CERN laboratory in Geneva in 1990. It comprises more than 60% of all information on the hard-wired internet.

Access a vast information database with *Toll-Free linking*

“Toll-free” linking gives you immediate access to the full text of many of the cited articles in a paper’s reference list—FOR FREE. With the support of HighWire’s vast journal catalogue, a huge reference library is now open to you. If HighWire hosts the journal, you can view the full text of the referenced article, completely free of charge by following the Free Full Text links in the references.