

Fixing the Broken Toaster: Scholarly Publishing Re-imagined

Julian H. Fisher, MD

Scholarly Exchange, Inc., Brookline, MA and Department of Neurology, Beth Israel
Deaconess Medical Center, Boston, MA

Contact: Julian H. Fisher, MD, Scholarly Exchange, Inc., 320 Dudley Street, Brookline, MA
02445, fisher@scholarlyexchange.org

Abstract

Ever-increasing journal subscription prices continue to stress the academic community in both economic and intellectual terms. Newly available tools and approaches to article production deployed in a collaborative manner enable two orders of magnitude of cost reduction. Open access – providing content globally at no cost – can be an economically viable and sustainable proposition.

Introduction

A basic lesson of mechanical engineering is that complex problems often have simple solutions if approached logically. The processes that make a toaster work are multiple, distinct and linear. The toaster that emits a ‘poof’ instead of light, heat, and nicely browned bread signals trouble, requires diagnostic analysis, and then a repair. The technician’s traditional approach breaks the process down into its logical steps, splits the sequence at any point, and tests each subset. The splitting-and-testing continues until the broken or defective link reveals itself, at which point the technician intervenes with the proper part to repair the problem efficiently. The speed with which the process occurs and normal functions return is nothing short of miraculous to those of us who lack the gene for engineering logic.

The time-honored tradition of scholarly publishing can be considered in the somewhat mechanistic light of the toaster – or the broken toaster. Many in the academic world consider the mechanism to be either broken or not functioning optimally – a device in need of repair.

The discussion boards and listservs of the scholarly publishing and library worlds abound with frequent, repetitive ping-pong-like debates about subscription costs, mandatory subscription bundling, alternative subscription pricing models, cancellation rates, open archiving, and definitions of copyright or fair use. The issues raised are the symptoms of an enterprise – scholarly publishing – that appears to have numerous maladies, costing more than it should, held hostage by commercial interests, dysfunctional in its efforts to disseminate information, and reluctant to adopt the new.

It is helpful to apply some engineering logic from the toaster analogy to the scholarly publishing process. Outlining, analyzing and testing the steps that lead to the production and dissemination of scholarly knowledge can be both startling and revealing. Rarely do the participants in this debate step back from the heat of the argument for an overview of the processes to:

- diagnose the problems
- propose rational solutions, and
- manage effective and successful implementation

Does the publishing process need to cost what it does? If not, why has the world not adopted newer, more cost-efficient approaches? This raises the greater social engineering issue: who or what is impeding progress? How can we overcome these hurdles? Only then can one determine whether the costs are fair, intellectual integrity (the peer review process) maintained, access enhanced, posterity (archiving) and the future (new knowledge) served. Despite the reported list of greater than 2500 scholarly journals flagged as open access, many are not truly free and open for current articles and the majority is still burdened by high production costs. [DOAJ]

The publishing process seems too rooted in history, too logical, perhaps even too self-evident to require a detailed description. Yet that is precisely what one needs to do, in light of new technologies, from manuscript preparation throughout the editorial review process and then to display, dissemination, and preservation. Studying the steps involved can reveal the relative efficiencies or inefficiencies of the system, to determine what is broken and how it might be fixed.

The cost of intellectual goods

Scholarly communication begins with the preparation of a manuscript, either research-oriented or didactic. It is understood within the community that the sharing of knowledge furthers both intellectual inquiry and careers (retaining a position or advancing in that community). In nearly all cases, faculty or industry salaries and grants of some kind have covered those information development costs – the scholarship or research itself, any associated data analysis, essential illustrations, and other supporting materials. While the work may have begun in handwritten form, a safe assumption is that it will have been converted into a word-processed document - an electronic format of some kind – in preparation for submission. [Sosteric] The cost at this point in the publishing process – the cost of goods - has always been considered and continues to be zero. In no other industry does one produce material of great value, give it to others, and allow them to sell it back to the community that created it and at a high price. [Bergstrom]

Objections sometimes arise at the point of manuscript submission that variable formatting of material confuses, delays or adds cost to the evaluative process. Practically every journal maintains submission guidelines, either detailed instructions to authors or a complex style sheet to be followed. It is the rare journal that can claim complete compliance with such guidelines. Some journals have promoted an electronic template into which their potential contributors can flow and format material. While this is workable, with variable cooperation a reality, it does not eliminate the final, more precise formatting and (potentially) tagging that occur after acceptance, in preparation for display.

Electronic preparation should, in theory, enable ready, appropriate layout of the material to the standards set by a given publication. Author individuality and serial submission (rejection by one publication leads to a submission to the next publication) make a journal's wishes for conformity to a style more desire than reality.

Yet this is not a concern, nor should it add any real cost. If the material submitted is not in a comprehensible and reviewable format, it becomes the author's responsibility and cost in time to reformat it, a part of the initial cost of intellectual goods that bears no financial burden for the publication. The initial submission should be a manuscript that can be converted easily to a viewable and comprehensible format for the editors and appropriate reviewers.

Only those few publications that consider themselves to be at the top of the intellectual food chain can insist upon adherence to guidelines and reject non-compliant submissions. One need not erect barriers at the outset, either technical or esthetic, that diminish the quantity and potentially the quality of incoming material, especially if a high percentage of submitted manuscripts face rejection.

Peer review costs

While there are rare exceptions, the rule has long been that peer review is a contributed and collaborative effort within a scholarly community. Each reviewer dedicates an amount of time (again covered by salary or grant or simply donated after hours) to evaluate submissions in that individual's field. It is a socialistic enterprise of sorts, benefiting the field and colleagues – from each according to his ability, to each according to his need. [Odlyzko1] The cost to the scholarly publication is customarily again zero.

There are, however, some minimal expenses involved in preparing the manuscript for distribution to section editors and reviewers. Recall that the originally submitted manuscript may have arrived in any of several electronic formats. Reviewers need a common cross-platform standard for viewing, and that commonly is the Portable Document Format (PDF), an open file format developed and maintained by Adobe.

Converting documents from a variety of formats into PDF is neither difficult nor costly, with software tools ranging in cost from free (the open source Open Office program as well as dedicated PDF converters, with some limitations) [Google1] [Wikipedia] to a several hundred dollars (a single copy of Adobe Acrobat) available. The conversion process to PDF takes a minute or two in minimally trained hands and should cost no more than a dollar or two in time and material. The resulting PDF can be sent to the reviewer as an e-mail attachment or made available for direct download (with e-mail notification) from an editorial management system. In the exceptional case that reviewers require paper documents, the cost of postage plus the secretarial function cost add another dollar or two to the publication cost.

Editorial management costs

Editorial oversight for incoming manuscripts involves both processes and people. Process control requires managing the flow of manuscripts among authors, editors, and reviewers. The people responsible for editorial progress are the editors, associate or section editors, managing editors, and editorial assistants who have traditionally controlled the flow of paper and now electronic files. At the end of the process stream are the copy editors, layout editors and the amorphous cloud of people whose role is, as needed, to format and tag content according to pre-set standards, for display, indexing, and archiving.

A working platform

Workflow software platforms now abound to manage the movement of electronic documents in an editorial system. They permit authors or editors to submit their electronic documents directly to the flow control platform and at the same time enter the necessary descriptive data (metadata) that will later uniquely identify the work. Submitting authors can provide guidance to the editorial office with keywords and other information that facilitate the review

process and perhaps eventually contribute to effective indexing and archiving. Such systems contain the submissions, the pathways that these potential articles will follow, the reviews, editorial decisions, and correspondence – in a word, everything of and about the manuscript, in a circumscribed place. Even for journals requiring some degree of paper, these software platforms prove helpful, if no more than a clever electronic tickler file and content hopper. Nothing need be overdue or lost.

These software programs far exceed in complexity the previous generation of systems that merely tracked the physical presence of paper-based manuscripts. Despite startling differences in acquisition cost that run from free (the open-source Open Journal Systems and DPubs, among others) [OJS] [DPubs] to tens or hundreds of thousands of dollars (and into the tens of millions) for the proprietary commercial and not-for-profit publishing systems [ScholarOne] [EditorialManager] [HighWire] [NYTimes], the differences in functionality are amazingly minimal.

One must reasonably factor in a cost for the implementation and support of such an all-in-one platform. Here again, the costs for essentially similar services range from free [SE] to a thousand dollars [OJS] to many orders of magnitude greater, with only slight differences in the services provided. An editorial platform already set up and made available to multiple journals as ready-to-use (the application service provider (ASP) model) is far more effective than an individually installed and configured system. Shared resources reduce cost to a minimum and enable almost immediate startup for new journal participants.

The editorial workforce

On the human side, one must acknowledge the time and effort that editorial staff contributes to the production of a finished article. That effort encompasses oversight of the editorial workflow process, editorial assignments for peer review, editorial decision-making, and copy editing for intellectual content, style and overall appearance.

For all but the largest-circulation (and often most lucrative) journals, senior editorial staff have long donated their time and effort to the cause as an integral part of their academic positions. In select cases, societies or publishers have contributed a stipend of varying amount to cover secretarial time, in effect paying a portion of salary for a manager or managing editor. Economic pressures and cost-cutting by for-profit publishers have caused a drying-up of these support funds. Fortunately, the amount of time required for workflow oversight has diminished with the newer and cleverer software platforms.

An editor of a low-volume journal can readily handle submissions in the 100 per year range with a 25% acceptance rate and spend no more than an hour a week on administrative functions in an automated manuscript management system. This has been the experience of numerous journals using the Open Journal Systems software implemented on the Scholarly Exchange site. [Fisher1] These are the functions formerly handled in a paper-based world by editorial assistants or managing editors. One can consider the collaborative production of a number of journals on such a platform, with a dozen or more publications sharing administrative support. [SE][Crow]

Editorial decision-making and article-editing functions require considerably more time. These, like the authors' and reviewers' time, are subsumed in the salaries or grants that faculty receive.

Putting shape to content

The only inescapable costs at the end of the editorial review process relate to copy editing and tagging. Journal editors may already have spent considerable (free) time working with the author on content presentation, and many editors consider their work the final stage for an accepted manuscript. Others insist on the more traditional refinement that professional copy editors bring to bear, adding semantic clarity to content. One open access publisher reduced its costs by making authors responsible for their own copy editing, advising that they retain the services of a professional copy editor for accepted articles as needed. [BioMedCentral] Other publications have begun to follow that model.

The cost of copy editing depends on the area of specialization, manuscript length, and regional pay rates. Nonetheless, allocating from \$25-\$75 for a 5-10 page manuscript seems a conservative yet realistic number.

While many journals are content to present the PDF format as the final and only viewable end-product, with a consequent cost of perhaps another dollar or two to convert the finally edited manuscript, those submitting material to such archives as the US National Library of Medicine's PubMedCentral must convert their unstructured material into highly structured and tagged ASCII text. Both metadata and data (content) must be fully identified at beginning and end of each element, to as finely granular a degree as the individual character level.

A Google search [Google2] yields a range of free-standing software programs in the multi-hundred dollar range to perform some degree of tagging, often to a user-configurable XML standard. Even the cleverest programs, attempting to digest variably entered word-processed documents and converting to the simplest XML document definitions, tend to make endless mistakes. Human oversight and intervention is necessary. Costs may mount to an extent that defeats the benefit of the automation.

That same Google search will also yield, as a viable alternative, commercial outsourcing options, companies that offer re-keying and tagging services. Balancing the automation followed by human intervention-and-cleanup against simple outsourced human re-keying often favors the latter. More complex document tagging definitions (such as the US National Library of Medicine DTD) can be accommodated through the outsourcing option, with re-keying and tagging costs well under a dollar per standard page of text. [Fisher2] A five-page article encoded for archival storage costs barely more than an elaborate Starbucks coffee preparation.

Displaying and storing the final product

The final product is no longer the monolithic printed page but rather a collection of tagged ASCII text, with a wide range of supporting files (sound, video and program files, to name but a few). They comprise the content and related metadata to be stored, converted, and displayed in a variety of ways at present and many more in the future.

Today's readers have needs that may be at variance with today's archivists, but the cost of preparing a final common pathway for a manuscript need no longer be prohibitive. The most

page-like appearance for many journal readers is the PDF, whether viewed on-screen or printed out – titles centered, citation at the top of each page, two-column display, all the niceties of the original printed product. Enhancing the value of PDF is its ability to be fully searchable electronically and its newly promulgated archival nature in an open file format environment – able to be understood electronically and up-converted to new and evolving storage standards. [PDF/A] Modifying or editing existing PDF files is a simple and feasible task. No longer need one consider the PDF an immutable concrete block of content.

For those readers eager to take advantage of the hyperlinked nature of the Web, the PDF document alone may not be sufficient. While one can embed higher-resolution images in a PDF file, file sizes and download times grown to levels unacceptable to many readers. The higher-quality images will also only be apparent in a printed copy.

The linking of thumbnail images to higher-resolution or enlarged versions and the inclusion of sound or animation may necessitate a HTML-based display file. Moving beyond PDF to this tagged-and-formatted state of the content entails some additional work. Fortunately, the preparation cost associated with this degree of structure within an article has dropped precipitously over recent years, through a combination of technology and outsourcing.

Indexing, archiving, and sharing

Submission to specialized indexed collections and archiving costs are, in many cost models, not included in traditional estimates of first-copy production costs. The very nature of the Web – information dispersed and distributed widely – and the plummeting costs of electronic storage open possibilities and eliminate costs that might have required consideration in the past.

Production costs formerly essential for indexing or archiving are either no longer important or can be accomplished at orders of magnitude less. There is a growing realization that the CrossRef digital object identifier (DOI) may be replaced by rapidly evolving academic literature search technologies from Google, Yahoo, and Microsoft, among others. The increasing trend toward self-archiving or university- or government-based digital archives leads to multiple copies of articles kept at different sites. This has been formalized with the open-source LOCKSS (Lots of Copies Keep Stuff Safe) software, enabling a journal in a digital archive at one institution to be replicated and kept fully synchronized with one, a hundred or a thousand other digital archives, all in real time, and essentially at no added cost. [LOCKSS]

The current state of digital archiving is another fluid entity with a variety of initiatives competing for attention and longevity. While for-profit publishers have under some duress agreed to deposit archival versions of their publications in selected university library archives, in the event of financial failure or other disappearance [Yale], the general publishing community has advanced the notion of fee-based secure storage, Portico. [Portico] Whether this project has the stigmata attached to CrossRef, with costs exceeding those acceptable to small independent publications, and is destined to CrossRef's ultimate fate in the face of Google/Yahoo/Microsoft is open to debate.

In the meanwhile, university librarians have been aggressively pursuing such independent projects as DSpace [DSpace] and Eprints [Eprints] based on commonly available and open-source standards as the Open Archive Initiative [OpenArchive]. These more democratic and

less commercial resources may in conjunction with the The Internet Archive [Archive] offer a viable solution.

Funding the enterprise

The ongoing debate about the cost of publishing a scholarly article inevitably leads to a variety of necessary payment schemes, from the all-too-familiar subscription model (reader pays) to newer author-pays schemes. The latter was a central tenet for BioMedCentral and Public Library of Science when they began, but both have begun to push fees considerably higher. Some commercial publishers have matched the concept but, as expected, at still higher rates.

While studies have pointed to first-copy paper or paper-plus-electronic production costs in the \$4000 - \$5000 range, [King] [Odlyzko2] [Bot] [ARL] the current experience with electronic-only as outlined here is considerably less. Recall that neither authors nor reviewers nor editors (in most cases) contribute to cost. There are minimal costs in the range of several dollars for the preparation and delivery of a reviewer-acceptable PDF document. There are similar costs for the final display formatting of a PDF. Per-article costs for optional professional copy editing and conversion with tagging can add in the range of \$50 and \$5 respectively to the cost. The platform upon which review, production, and display occur can add anywhere from nothing to \$50 to the cost, depending on the platform chosen and the article volume processed (for distribution of costs) [Table 1].

Hypothetically, if one reduces publishing costs substantially, by two orders of magnitude, the question of who pays – reader or author – becomes irrelevant. A journal article with a first-copy production cost of \$4000 and another with a cost in the \$40 range have vastly different funding requirements. The double-digit article is no less valid in terms of content, appearance or archivability – overall value – than the quadruple-digit one. For a year's worth of double-digit articles in a low-volume specialty journal (50 articles yearly, a not-uncommon figure), the out-of-pocket expense is \$2000 rather than \$200,000. The complexities of funding melt away, with a society, a research grant, a departmental fund or a library able to cover costs indefinitely and painlessly. Alternatively, a journal could levy a publication fee in the range of \$50-\$100 that would adequately cover expenses – well below the fees currently charged by many not-for-profits and a miniscule fraction of those charged by the profit-making entities.

In effect, the cost of one subscription for one library or one author publication fee practically covers the out-of-pocket costs for a journal with 50 articles distributed globally at no cost to readers.

The economic argument is controversial but supportable, and functioning examples are already in place [SE2]. The commercial world and many of the open access journals still wedded to old and established technology have difficulty accepting the cost model. Beyond that, there is resistance, perhaps better described as reluctance and doubt, from the very community that could benefit most, academics.

Facilitating change

Promoting change – fixing the toaster – requires an acknowledgement or an admission that the toaster is broken. Many in the scholarly community fail to see the process in those terms,

because the system appears to them to be functioning satisfactorily, and the sought-after change has not occurred. [Harnad] They choose, for reasons of career advancement, to submit to journals that have an established reputation, are affiliated with their professional or scholarly society or specialize in their field of expertise. They are unaware that increasing monopolistic practices and costs are reducing the exposure that an accepted article may have, apart from the ever-fewer privileged subscribers. Many feel that they are unaffected since their university continues as one of those subscribers. Therein lies the irony, that scholarship published even electronically in this Internet age should have decreasing visibility and availability.

From another vantage point, one could argue that the toaster is not broken, but that the price of toast has risen at a time that it should be falling. The existing system limps along at a cost that is not evident to academic users but is painfully apparent to administrators and librarians. The increasingly exorbitant prices are hidden in higher tuition fees, steeper institutional overheads charged for research grants, ever-longer lists of subscriptions cancelled, manipulative bundling arrangements, and restrictions on acquisitions of other reference resources - in effect the denial of access to intellectual resources.

The motivation to change has begun with the involvement of academic thought leaders, librarians, professors, deans, and presidents, from many disciplines. It will be encouraged by emerging research demonstrating the value of open access publishing. The principle of open access, giving away intellectual products of great value, requires that the means and cost of production be reduced to the barest minimum, as can now be done.

It behooves universities and research institutes to rethink their criteria for advancement and tenure, to consider new journals produced at lowest cost and maximal free distribution as recognized venues for scholarship of the highest grade along with the established brands.

Combining the motivation with the means requires that universities, philanthropic entities, and governments make a commitment to support a lowest-common-denominator approach to scholarly communication. The simple tools now exist, and there must be the support to begin using those tools. Analogous to the university presses of old (many of which are now cost-constrained) but with new models and better approaches, there need to be collaborative efforts that help journals make the transition from old technology to new – or re-invent themselves completely. [Crow][SE] [Okerson]

Scholars need to understand that low cost of production does not imply low quality of content, that low cost enables greater distribution of knowledge and enhanced academic prestige. Editors require help and encouragement in using these tools. Centers of learning, governments, and philanthropic organizations have a new role: to facilitate the adoption of these newer, simpler, less costly platforms and methods. Only then will journals be capable of producing the same toast as before while sharing it more widely and at dramatically lower cost to society. The dream of open access becomes a real, viable, and sustainable proposition.

Acknowledgment:

I thank my fellow directors at Scholarly Exchange, Yuwei Shi (Fisher School of International Business, Monterey Institute) and Mike Sosteric (Athabasca University) for the extended discussions that have led to this re-thinking of the entire publishing process.

References

[Archive] <http://www.archive.org/index.php>

[ARL] ARL Monograph and Serial Costs in ARL Libraries, 1986-1999, <http://www.arl.org>

[Bergstrom] Free Labor for Costly Journals? Bergstrom, T. (2000, Dec 1).
<http://www.econ.ucsb.edu/~tedb/jep.pdf>.

[BioMedCentral] <http://www.biomedcentral.com/independent/develop/copyediting>

[Bot] The Cost of Publishing an Electronic Journal: a general model and a case study
Marjolein Bot, Johan Burgemeester Hans Roes
<http://www.dlib.org/dlib/november98/11roes.html>

[Crow] Publishing cooperatives: an alternative for non-profit publishers. Raym Crow
http://www.firstmonday.org/issues/issue11_9/crow/index.html

[Eprints] <http://www.eprints.org/>

[Fisher1] Personal communications with Scholarly Exchange journal editors during 2006.

[Fisher2] Personal communication, Mala Morris, NewGen Imaging Systems (India),
<http://www.newgenimaging.com/>, November 10, 2005.

[Google1] “free web conversion to pdf adobe”

[Google2] “XML conversion”

[DOAJ] <http://www.doaj.org>

[DPubs] <http://www.dpubs.org/>

[DSpace] <http://www.dspace.org>

[EditorialManager] <http://www.editorialmanager.com/>

[Harnad] Free at Last: The Future of Peer-Reviewed Journals. Stevan Harnad,
<http://www.dlib.org/dlib/december99/12harnad.html>

[HighWire] <http://www.highwire.org/>

[King] Economic Cost Models of Scientific Scholarly Journals. Donald W. King, Carol Tenopir presented to the ICSU Press Workshop, Keble College, Oxford, UK, 31 March to 2 April 1998 <http://www.bodley.ox.ac.uk/icsu/kingppr.htm>

[LOCKSS] <http://www.lockss.org/>

[NYTimes] Amount Elsevier invested in 2000 to put journals online: \$60 million
David D. Kirkpatrick. (2000, Nov 2). As publishers perish, libraries feel the pain, New York Times.

[Odlyzko1] Peer and non-peer review, A. M. Odlyzko. Peer Review in Health Sciences, 2nd ed., F. Godlee and T. Jefferson, eds., BMJ Books, 2003, pp. 309-311.

[Odlyzko2] The economics of electronic journals, A. M. Odlyzko. First Monday 2(8) (August 1997), http://www.firstmonday.org/issues/issue2_8/odlyzko/index.html/, and Journal of Electronic Publishing 4(1) (September 1998), <http://www.press.umich.edu/jep/>. Definitive version on pp. 380-393 in Technology and Scholarly Communication, R. Ekman and R. E. Quandt, eds., Univ. Calif. Press, 1999

[Okerson] Back to Academia? The Case for American Universities to Publish Their Own Research. Logos 2, no. 2 (1991): 106-112. Ann Okerson.
<http://www.library.yale.edu/~okerson/case.html>

[OJS] <http://pkp.sfu.ca/ojs/>

[OpenArchive] <http://www.openarchives.org/>

[PDF/A] ISO 19005-1. Document management - Electronic document file format for long-term preservation - Part 1: Use of PDF (PDF/A)
<http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml>

[Portico] <http://www.portico.org/>

[ScholarOne] <http://www.scholarone.com>

[SE] <http://www.scholarlyexchange.org>

[SE2] <http://www.scholarlyexchange.org/journals.html>

[Sosteric] Electronic First: The Upcoming Revolution in the Scholarly Communication System. Mike Sosteric, Yuwei Shi, Olivier Wenker <http://www.press.umich.edu/jep/07-02/sosteric.html>

[Wikipedia] <http://en.wikipedia.org/wiki/Pdf>

[Yale] Yale News Release February 23, 2001: Yale Library to Plan Digital Archives with Elsevier Science
<http://www.yale.edu/opa/newsr/01-02-23-02.all.html>

Table 1**Out-of-Pocket Cost Estimates for a Facilitated Open-Source Publishing Initiative**

Task	Estimated cost / year / journal with 50 published articles	Estimated cost / year / journal with 100 published articles	Estimated cost / year / journal with 250 published articles
Supported Platform	0 - 1500	0 - 1500	0 - 1500
Author Submissions	0	0	0
Review Functions (PDF creation @ \$2/ manuscript)	100	200	500
Editorial Decision-making	0	0	0
Copy editing @ \$50 / article (optional)	2500 (optional)	5000 (optional)	12,500 (optional)
XML conversion and tagging	250 (optional)	500 (optional)	1250 (optional)
Display (reformatting @ \$5/ manuscript + PDF creation @ \$2/ manuscript)	350	700	1750
Basics (required)	450-1950	900-2400	2250-3750
Options	2750	5500	13,750

All costs in \$ US

Submitted to and accepted for publication December 20, 2006 to Science and Technology Libraries, Haworth Press