



By Marti A. Hearst  
University of California, Berkeley  
hearst@sims.berkeley.edu

## Innovations in electronic academic publishing

Reports abound of the incipient implosion of academic libraries, caused primarily by the growing cost of journal subscriptions.<sup>1</sup> This is ironic, given the increased role of the researcher in the production of academic publications and the

ubiquitous access provided by the World Wide Web. Intuitively, the cost of academic journals should be *decreasing* rather than increasing.

At the same time, the number of scholarly journals whose primary venue is online has increased by an order of magnitude since 1994 (from 181 to 1093).<sup>2</sup> (The Information Economy Web site carries a list of electronic journals: <http://www.sims.berkeley.edu/resources/infoecon/EPublish.html>.)

A pressing question is: if online publishing of academic information replaces traditional print publication, will this result in significant cost savings? And if so, will the quality of published material be retained? Improved? And finally, what catalysts are needed to move scholars away from their traditional modes of publishing?

Andrew Odlyzko has written thoughtful discussions of roadblocks to and advantages of electronic publishing,<sup>4,5</sup> and Hal Varian has predicted the criteria that would have to be met to make academic electronic publishing viable.<sup>6</sup>

Publishers are exploring a wide variety of pricing models. The JSTOR project<sup>7</sup> imposes differential pricing by institution: those that are rated as having a larger need for the service are charged more.<sup>6</sup> Steve Harnad has suggested charging authors (or their institutions) a “page” fee when their paper is accepted for publication, as a way to help disseminate the work.<sup>7</sup> And at least one study has found that the better accessibility of online journals increased the number of subscriptions.<sup>8</sup>

Some publishers are now making online access available to journals at an additional

cost beyond that of the paper-version subscriptions. The ACM (<http://www.acm.org>) offers online access to a substantial proportion of the conferences and journals it sponsors. The charge is one reasonable fee for access to *all* online material in addition to the fee for a subscription to at least one paper journal. This strikes me as a worthwhile tradeoff, and I plan to subscribe to the ACM digital library service. Other publishers offer a much less inviting deal: institutions can pay an additional fee on top of the already inflated subscription price for online access to one journal.

### This installment

In this installment of “Trends and Controversies,” we examine three different innovations in electronic academic publishing of interest to practitioners in the field of intelligent systems.

The essay by Michael Wellman and Steve Minton on the status of *JAIR* supports many of Varian’s cost predictions. Wellman and Minton describe how this online academic AI publication works within existing social institutions to produce a high-quality, rapid-turnaround journal. The editors of *JAIR* note that electronic publishing allows for innovative policy choices: they have no pressure to publish papers just to fill out a required number of pages, and conversely, they have no space limitations impeding the publication of data, code, animated graphics, and other supplemental material. Another interesting policy is the refusal to review more than two versions of the same paper. If authors of rejected papers decide

to try another forum, they have received the benefit of useful reviews in a short amount of time (the turnaround for *JAIR* being 64.5 days on average). Wellman and Minton draw our attention to yet another advantage of online publishing: once a paper is accepted it can be published instantly.

Our next essay is highly relevant to the health care theme of this inaugural issue of *IEEE Intelligent Systems*. Ida Sim describes exciting possibilities for changing and improving the medical profession through innovative electronic publishing. She argues not only for convenient access to information, but also for using the electronic medium to enable entirely new forms of research and diagnosis. Sim’s PhD work in Stanford’s medical informatics department proposed a standardized knowledge-representation method for representing the results of clinical medical trials, and her essay paints a vivid image of how information sharing using this kind of representation could contribute to better health care in life-and-death circumstances.

Perhaps the most often mentioned roadblock to online academic publishing is the prestige factor: a “network externality” effect arises because scholars want to publish in the most prestigious location possible, and that currently the most prestigious journals in most fields are traditional print publications.

Although all three of our essayists mention the prestige factor, Michael Lesk’s contribution addresses this problem squarely. Lesk describes a radical new idea: partition up academic Web sites according to quality ratings (say, gold, silver, and bronze) and award monetary prizes of corresponding value for the best contributions to each area. He estimates the costs required would be less than that of the deficits of several university presses. He considers other approaches but surmises that in the end, money talks.

## Doing our bit

It seems to me devices like these can help lead to faster access for less cost, but that individual members of the scholarly community can also take steps of their own. For example, scholars can encourage nontraditional measures of impact,<sup>9</sup> and can think twice before publishing articles in expensive journals or accepting offers to join editorial boards of new journals that do not provide inexpensive online access.

As my own small gesture, I have persuaded the editors of this publication to make the full text of the “Trends and Controversies” column available online, free of charge. This can act as a testbed for the IEEE Computer Society; perhaps providing certain parts of the magazine for free can increase revenues overall. (Earlier issues of this magazine were freely accessible online. I placed two articles from a T&C column edited by Craig Knoblock into a collection of papers for one of my courses. If I had had to go to the library to get a one-day-only loan of the magazine, I probably would not have used the articles in my reader, and the publisher would have lost the revenue and exposure that the online version made possible.)

You can do your part to help break the online academic publishing logjam by visiting the site at <http://www.computer.org/intelligent>.

—Marti Hearst

## References

1. R.M. Berdahl, “Rising Costs are Challenges for Library,” *The Berkeleyan*, Jan. 14, 1998; [http://www.urel.berkeley.edu/urel\\_1/CampusNews/berkeleyan/01-14-98/library.html](http://www.urel.berkeley.edu/urel_1/CampusNews/berkeleyan/01-14-98/library.html).
2. P.F. McEldowney, “Scholarly Electronic Journals, Trends and Academic Attitudes: A Research Proposal,” masters project, Dept. of Library and Information Studies, Univ. of North Carolina, Greensboro, N.C., 1995; <http://poe.acc.virginia.edu/~pm9k/libsci/ejs.html>.
3. M. Odlyzko, “The Economics of Electronic Journals,” *First Monday*, Vol. 2, No. 8, Aug., 1997; <http://www.firstmonday.dk/>.
4. M. Odlyzko, “On the Road to Electronic Publishing,” *Euromath Bull.*, Vol. 2, No. 1, 1996, pp. 49–60; <http://www.research.att.com/~amo/doc/eworld.html>.
5. H.R. Varian, “The Future of Electronic Journals,” *Scholarly Communication and*



**Michael P. Wellman** is an associate professor in the University of Michigan’s Department of Electrical Engineering and Computer Science and also a member of its AI Lab. His research interests include decision theory, planning, knowledge representation, economics, probabilistic reasoning, electronic commerce, and distributed decision-making. He received an SB and an SM in computer science and a PhD in AI, all from MIT. He is the Executive Editor of the *Journal of Artificial Intelligence Research* and a member of the American Association for AI, the American Economic Association, the Association for Computing Machinery, and Informs. Contact him at the AI Laboratory, University of Michigan, 1101 Beal Ave., Ann Arbor, MI 48109-2110; [wellman@umich.edu](mailto:wellman@umich.edu); <http://ai.eecs.umich.edu/people/wellman/>.



**Steve Minton** is a senior computer scientist at USC’s Information Sciences Institute and a research associate professor in the USC Computer Science Department. His research interests focus on machine learning, planning, scheduling, constraint-based reasoning, and program synthesis. He received a PhD in computer science from Carnegie Mellon University and a BA in psychology from Yale. He was the founder and first executive editor of the *Journal of Artificial Intelligence Research*. He is on the executive council of the AAAI and a member of the ACM. Contact him at the Information Sciences Inst., Univ. of Southern California, 4676 Admiralty Way, Marina del Rey, CA 90292; [minton@isi.edu](mailto:minton@isi.edu); <http://www.isi.edu/sims/minton/homepage.html>.



**Ida Sim** is an assistant professor in the Department of Medicine of the University of California at San Francisco. Her research interests are medical informatics, evidence-based medicine, knowledge representation, medical publishing, computer-based reasoning, clinical trials, and meta-analysis. She received a BS in biological sciences from Stanford University and an MD and a PhD in medical informatics from the Stanford Univ. School of Medicine. She belongs to the American Medical Informatics Association, the American College of Physicians, the Society for Medical Decision Making, and the American Association for the Advancement of Science. Contact her at the Dept. of Medicine, Univ. of California, San Francisco, 400 Parnassus Ave., A-405, San Francisco, CA 94143-0320; [sim@medicine.ucsf.edu](mailto:sim@medicine.ucsf.edu); <http://www.dgim.ucsf.edu/People/Sim-profile.html>.



**Michael Lesk** heads the Division of Information and Intelligent Systems at the National Science Foundation and is a visiting professor at University College London’s Department of Computer Science. In the 1970s, he worked in the original group that developed Unix; he wrote Unix tools for word processing (tbl, refer), compiling (lex), and networking (uucp). He received a PhD in chemical physics from Harvard. He recently wrote *Practical Digital Libraries: Books, Bytes, and Bucks* (Morgan Kaufmann, 1997). He received the Flame award for lifetime achievement from Usenix and is a fellow of the ACM. Contact him at NSF, 4201 Wilson Blvd., Arlington, VA 22230; [lesk@acm.org](mailto:lesk@acm.org); <http://www.lesk.com>.

*Technology Conf.*, 1997; <http://www.sims.berkeley.edu/~hal/Papers/publish.html>.

6. K.M. Guthrie, “JSTOR: From Project to Independent Organization,” *D-Lib Magazine*, July/Aug. 1997; <http://www.dlib.org/dlib/july97/07guthrie.html>.
7. S. Harnad, “The Paper House of Cards (and Why It’s Taking So Long to Collapse,” *Ariadne (The Web Version)*, Vol. 8, Mar. 1997; <http://www.ariadne.ac.uk/issue8/harnad/intro.html>.
8. W.G. Regier, “EPIC: Electronic Publishing is Cheaper,” *Conf. Scholarly Communication and Technology*, Emory University, Apr. 1997; <http://www.arl.org/scomm/scat/regier.html>.
9. J. Ullman, “Research Publication Modes Need to Be Reengineered,” <http://www-db.stanford.edu/~ullman/pub/nopaper.html>.

## JAIR: an electronic journal by and for the AI research community

Michael P. Wellman, University of Michigan, and Steve Minton, USC/ISI

Progress in AI—as in any field of research—is marked by discrete reports of research results by individuals or teams, building on and relating to previously reported results of others. The accumulated record of these reports constitutes the field’s literature, which is integral to continuing academic research in the discipline, as well as to communication of the field’s contributions to other disciplines and to society at large.

Because of the research community’s centrality, over time it has developed a comprehensive system for building and

maintaining academic literatures. Many types of institutions participate in this system, including research organizations, professional societies, publishers, libraries, and universities. Together, these entities perform all the functions of publication—authoring, reviewing, distributing, archiving, organizing, and the like. The roles of this system's players have evolved into what was until recently a relatively stable configuration, where the mapping of activities to institutions was fairly clear-cut.

On occasion, technological innovation has altered the mapping between players and activities. For example, the proliferation of TeX and similar tools shifted much of the typesetting function from publishers to authors. Such shifts inevitably result from dramatic changes in the relative productivity of the respective players performing these activities.

By now, we can all agree that the Internet's advent is even more dramatically changing the shape of publishing and communication of all kinds (see Figure 1). This change manifests itself in an accelerated shift of functions among existing players—authors distributing their own literature via the Web, for instance—as well as the introduction of new players and new functions. For example, despite their broad scope, Web search engines—developed by institutions outside the conventional academic publication process—provide an important service for exploring academic literatures, and new electronic archives supported by institutions other than traditional libraries represent another new service and new kind of player.


Electronic journals play various roles for institutions and publication services. Many are simply online versions of existing publications, making them an expansion of services (often quite valuable) by existing players. Some new journals are primarily electronic, yet produced by existing commercial or society publishers. Yet others represent new entities—typically ad hoc formations within communities of academic researchers—brought into existence expressly to support an online academic journal.

## JAIR

*The Journal of Artificial Intelligence Research* is an instance of this last type (see <http://www.jair.org/>). JAIR is published by the AI Access Foundation, a nonprofit body

organized solely to produce the journal. Conceived in 1992 and publishing continuously since 1993, JAIR was one of the first serious academic electronic journals. Although quantifying success is difficult, we believe that most AI practitioners would say that JAIR has exceeded optimistic expectations, attracting a steady stream of high-quality submissions and distributing the best of these to the research community in a convenient manner, without fees. A recent survey of JAIR authors (admittedly, a biased sample) revealed a high degree of satisfaction with the JAIR process.

Perhaps even more difficult than evaluating the journal's performance is identifying the factors behind it. Nevertheless, we believe that the following characteristics have been instrumental in JAIR's success. Some reflect the journal's core values; others express design decisions at its founding;



*From its earliest days, JAIR received wide support in the AI research community, probably due mostly to the ideal of a free-access journal and excitement about the potential of electronic publication.*

and still others reflect fortunate coincidences.

*Free access.* JAIR articles are available online, free of charge. JAIR readers need no subscriptions, or accounts of any kind. Our aim is to eliminate all possible barriers to the dissemination of AI research deemed of sufficient quality by our editorial process. Of course, all editorial and reviewing labor—the major real cost of operating any journal process—is donated by individuals in the research community. We find it easier to obtain their services (an increasingly scarce commodity) because the journal is freely accessible.

We have been able to operate without fees by keeping our expenses extremely low. As Hal Varian (“The Future of Electronic Journals,” <http://www.sims.berkeley.edu/~hal/Papers/publish.html>) and others have pointed out, electronic publications

can avoid many of the costs associated with journals distributed and operated via paper media. Indeed, JAIR has achieved most of the cost reductions envisioned by observers. Distribution is handled by Web and FTP servers donated by our host institutions. These installations already support many servers and high-bandwidth connections, so the incremental cost is not prohibitive. We conduct the editorial process entirely via Web forms and e-mail, saving much administrative time (also donated by our host institutions) and completely eliminating postage. Our reviewing community—primarily computer scientists—naturally tends to be e-mail/FTP/postscript savvy, which contributes substantially to our process's efficiency.

*Resemblance to conventional journals.* A JAIR article looks exactly like a reprint from a conventional journal. The standard JAIR format is typical, and pages are numbered consecutively in a volume. Authors cite a JAIR article as they would an article from a paper journal. This approach has helped to overcome potential hesitation on the part of contributors concerned that their work would not be accorded the academic credit due standard articles.

*Hardcopy volumes.* Each JAIR volume is published in hardcopy by Morgan Kaufmann Publishers (<http://www.mkp.com/>). The existence of the hardcopy also relieves concern about the persistence of the journal (more an issue in the very earliest days than now), and provides further confidence with respect to archiving. Because most JAIR readers find it convenient to access the journal online, few request that their libraries purchase the hardcopy—although, given its low cost relative to most journals, many institutions would probably benefit from the hardcopy volume.

*Community support.* From its earliest days, JAIR received wide support in the AI research community, probably due mostly to the ideal of a free-access journal and excitement about the potential of electronic publication. Consequently, JAIR's editors and editorial board include some of the most prominent and respected members of the community, which was pivotal in establishing its credibility. To preserve vitality, we periodically rotate the editorial positions, and we continue to benefit from the

substantial participation of leading AI researchers (<http://www.jair.org/masthead.html>).

**Rigorous standards.** As spelled out in our editorial charter, *JAIR* strives to maintain the highest standards in quality research reporting. As a startup electronic journal, we were initially concerned that authors might assume relaxed standards, leading to either reluctance to submit their best work or willingness to submit subpar work, especially given the low cost of entry. Happily, neither seems to have happened, although we remain determined to maintain and improve the journal's reputation. Our publication process means that we have no issues to fill, no page targets to meet, and hence no pressure to accept articles we cannot enthusiastically recommend. Also, by avoiding dragging out the process over months and years, we are not compelled to accept articles out of implied obligation.

**Rapid turnaround.** One of *JAIR*'s major goals has been to reduce the time to publication. The long delays in reviewing and publication in many journals can significantly impede the communication of important research results. In most cases, we have cut the time from submission to an accept or reject decision to nine weeks or less. (In 1997, our median time from acknowledgment to decision was 64.5 days, a slight increase from previous years.) Conducting the entire process electronically and employing effective semiautomated tracking facilities saves considerable time. Perhaps more important, we have succeeded in establishing a new norm. Because referees now understand that we are a rapid-turnaround journal, they take the deadlines they agree to more seriously than they might otherwise (we never send papers to reviewers without prior consent). In fact, over 80% of our reviews arrive within a couple of weeks of their deadlines.

Our policy of reviewing at most two submissions of the same paper also reduces turnaround time. If an editor is not convinced by the second attempt, perhaps the author is better off trying another forum (presumably not a great deal of time was lost, and the authors benefitted from useful reviews anyway). And, of course, once we finally accept a paper, publication is immediate—there is no waiting for press runs,

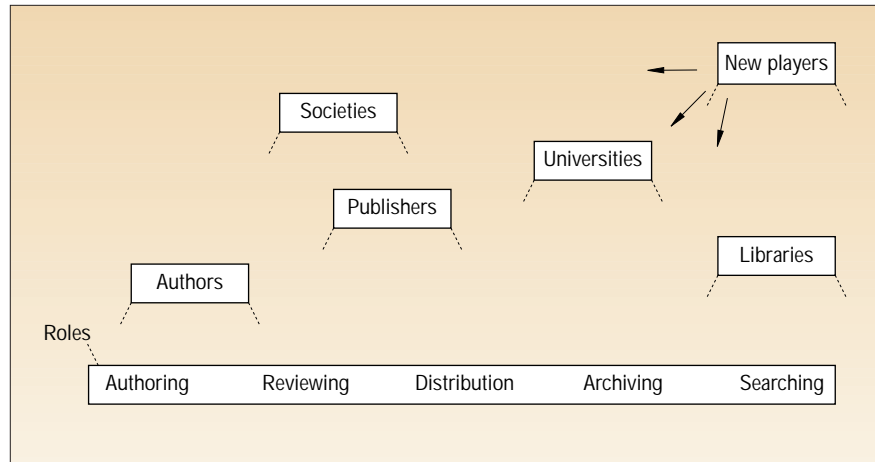


Figure 1. Technological change leads to shifting boundaries between the roles of participating institutions in the academic publishing process. In the highly dynamic current environment schematically suggested by this diagram, institutions are "floating" among many potential roles.

buffers, or synchronizations with other articles for an issue.

**Potential of electronic publication.** The electronic medium lets us stretch the concept of "article" to include many features not feasible in paper journals. Hypertext links, online appendices (containing code or data, for example), and multimedia enhancements (for demonstrations) have all appeared in *JAIR* articles. The *JAIR* site offers full-text search, automated keyword extraction, and other experimental visualization facilities. However, we believe that the potential remains largely unrealized, as authoring tools and other techniques that would exploit the possibilities are still in their infancy. *JAIR* is committed to engaging in occasional experimentation on innovative electronic features and to supporting techniques that will inevitably emerge through the thriving parallel exploration of the electronic publishing community.

### JAIR's future

Some observers have remarked on electronic publishing's potential for dramatically reshaping academic communication. For example, the new medium offers several alternatives to the conventional reviewing process. In our own field, the *Electronic Transactions on AI (ETAI)*—<http://www.ida.liu.se/ext/etai/> is an ambitious effort, led by Erik Sandewall, to restructure the standard cycle by instituting such innovative features as public comment before review. We find this kind of experiment very exciting and believe that process variations that capitalize on the electronic medium to gather evaluations can ulti-

mately substantially improve research communication.

The *JAIR* approach represents one model for community-supported electronic academic publishing, with a five-year track record. However, the electronic publishing environment is highly dynamic; we would be surprised if the model that works today is the ideal one for tomorrow. As Judith Axler Turner, the editor of the *Journal of Electronic Publishing* (<http://www.press.umich.edu:80/jep/>), pointed out recently, pervasive uncertainty renders production of electronic journals a relatively spontaneous affair.<sup>1</sup> Thus, we continue to watch our peers and look for opportunities—consistent with the core values of access and quality standards—to deliver higher levels of service to our authors and readers.

### Reference

1. J.A. Turner, "Mickey, Judy, Colin, and Me," *First Monday*, Vol. 3, No. 1, Jan. 5, 1998; [http://www.firstmonday.dk/issues/issue3\\_1/turner/](http://www.firstmonday.dk/issues/issue3_1/turner/).

### Acknowledgments

*JAIR* is the product of the dedicated efforts of many individuals, particularly the associate editors and board members. We would like to single out the contributions of Peter Turney, who developed *JAIR*'s core search and keyword features, and Jon Doyle, who enhanced our online presence in several ways. General opinions expressed about electronic publishing are those of the authors and do not represent official *JAIR* policies.



## Medical publishing meets AI

Ida Sim, University of California, San Francisco

You have died suddenly. In a drama befitting the best of television's ER, your doctors have restored your heart to its life-sustaining rhythm. For now, you are safe, but you are prone to dying suddenly again. Your doctors know of several new drugs that might keep you from suffering another fatal heart-rhythm abnormality. One drug, called amiodarone, has been tested in over 10 randomized, controlled trials. Together, these trials yield the best evidence on how well amiodarone works. Your doctors search the clinical literature to see if you should receive amiodarone. Will they find all the trials? Will the articles be in the library? Will the articles describe the trials completely and accurately? Will your doctors be able to properly synthesize the evidence from all the trials?

Every day, doctors try to practice evidence-based medicine: medical care that is based on the current best scientific evidence. Because academic medical journals are the primary vehicles for reporting new scientific evidence, the clinical literature is the foundation for the practice of evidence-based medicine. The evidence that is published in medical journals can truly be a matter of life, death, and everything in between: An electronic search that fails to identify the largest, most influential study on the efficacy of amiodarone may have far graver consequences for you than those of all your lifetime AltaVista searches combined.

Does the clinical literature effectively convey the evidence from medical research to the clinician? How might the clinical literature change with electronic publication, and will these changes help or hinder patient care? The short answer is that the clinical literature as it now stands does not satisfy the needs of its readers, and that electronic publication alone will not completely solve this problem. We need to augment electronic text publication with knowledge-base technologies for a complete solution.

The clinical literature is suboptimal

Scientific evidence can bear on a clinical decision in three ways, but the evidence is often so voluminous that frontline doctors cannot even determine which of the follow-

ing scenarios apply, let alone what the evidence says.

- *Good studies, easy answer:* In the ideal evidence-based-medicine scenario, well-conducted studies consistently show the superiority, or inferiority, of one therapy over another. Randomized trials, which randomly assign patients to their treatments, are generally the best studies because they are less subject to experimental bias.
- *Reasonably good studies, but interpretation requires advanced methodology:* In a more complicated but common scenario, a mix of higher- and lower-quality studies offer equivocal support for any one course of action. Reasonable specialists might interpret the evidence differently, and the interpretation requires both clinical and biostatistical expertise.
- *Lower-quality studies, offering little*

*An electronic search that fails to return to your doctors the largest, most influential study is a search whose consequences might be far graver than those of all your lifetime AltaVista searches combined.*

*useful information:* Less than 50% of clinical practice is supported by even modest scientific evidence. The majority of clinical practice has either not been studied at all, or has been studied with few or poor quality experiments.

As these scenarios imply, using the clinical literature is not simply a matter of retrieving the "right" article and doing what it says. Evidence from multiple studies must be carefully and properly synthesized. This time- and labor-intensive task is performed by *evidence synthesizers*, who generate systematic reviews of the evidence that frontline practitioners then read and apply to their patients.

Medical research should therefore be reported in such a way that the needs of evidence synthesizers are met. Unfortunately,

this is far from the case. Inconsistent indexing of articles results in electronic searches with poor recall and precision. Studies with negative results are disproportionately absent from the literature, thus introducing bias into evidence summaries. Study reports are sometimes incomplete, inaccurate, and unclear. Some studies have serious design flaws that preclude use of their evidence. The definition and reporting of treatments and outcomes are not standardized, and this makes it difficult to synthesize the evidence. The overall result is an inefficient transfer of precious scientific evidence from the research world to the clinic.

## Improving form, content, and medium

At the Third International Congress on Biomedical Peer Review and Global Communications last September, editors of major medical journals explored the ramifications of electronic publication (<http://www.ama-assn.org/public/peer/peerhome.htm>). An underlying assumption of the discussion was that the form and content of the articles will remain as they are now. But finding sub-optimal articles more quickly, accurately, or cheaply is only a partial improvement. Simply publishing articles in bytes rather than on paper will not repair the clinical literature's most critical deficiencies.

The clinical literature is too often wrong for everyone because it attempts to serve too many audiences simultaneously with short, single-format, text-based articles. Why don't we publish new knowledge as both prose discussions *and* as entries into knowledge bases? Knowledge that is fairly standardized will be amenable to such dual publishing, which has been called *electronic data publishing*.<sup>1</sup> For example, the journal *Science* recently published the entire genome of the bacterium *E. coli* as two components:

- an article describing the sequencing work and its implications,<sup>2</sup> and
- a set of entries into GenBank, a genomic-sequence database administered by the National Center for Biotechnology Information.

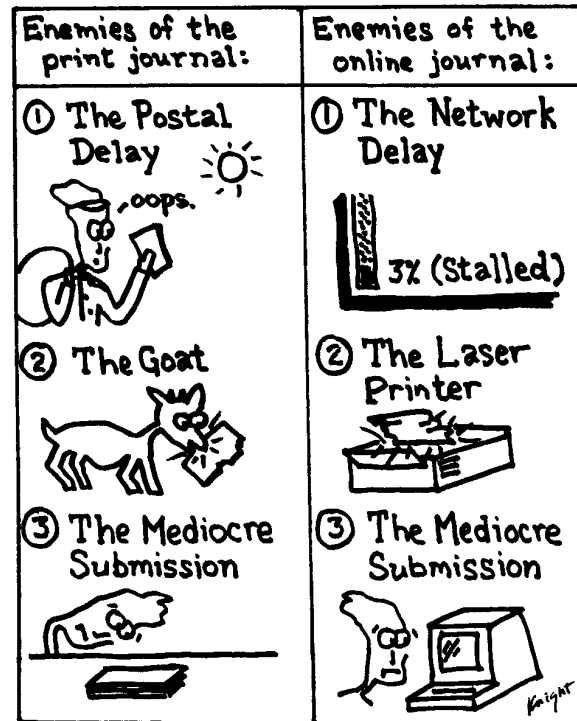
Because GenBank represents genomic-sequence information in a way that is generally accepted by the scientific community, researchers now publish most of their new sequencing data as GenBank entries

in conjunction with a journal article.

Randomized trials are also excellent candidates for electronic data publishing, because their structure is also highly standardized. Medical journals can publish randomized clinical trials in prose articles and also as instances in randomized-trial knowledge bases, or *trial banks*.<sup>3</sup> Just as in the electronic data publication of genomic sequences, journals can require prospective authors to submit for editorial review both a prose manuscript and a trial-bank entry. Authors can use Web-based authoring tools to submit accurate, complete, and standardized trial-bank entries. Accepted manuscripts can be published with a reference or a hyperlink to their corresponding trial-bank entries. Trial banks worldwide will interoperate so as to support comprehensive evidence synthesis.

What will publishing into interoperating trial banks allow us to do that we cannot with electronic publication of text alone? First, trial-bank-authoring software can help improve the quality of trial reporting. Also, we will be building knowledge bases of valuable medical evidence, knowledge bases that are designed specifically to support evidence-synthesis, and to support computer-based approaches to accessing and reasoning about that evidence. Nearly 9,000 randomized trials were indexed in the electronic bibliography Medline in 1996 alone; the availability of shared trial banks will markedly reduce the costs of building and maintaining expert and information systems that require up-to-date information on randomized trials. Furthermore, if trial banks are instantiated with a controlled medical vocabulary, search engines can exploit the trial-bank and vocabulary semantics to achieve more accurate retrieval. Finally, electronic searches of trials will be more comprehensive if trial banks interoperate worldwide, especially if the trial banks report on both planned and completed trials. Accurate and comprehensive searching is a prerequisite for proper evidence synthesis.

In contrast, if we published medical research only as electronic text, we would



remain subject to all the existing content problems of the literature. Also, we would miss the opportunity to leverage the resources of academic medical publishing to build sharable, standardized, up-to-date, self-sustaining knowledge bases of a key source of medical knowledge. In short, trial-bank publishing can lead to a foundation for medical expert systems that neither paper-based nor electronic text publishing can.

#### Knowledge technologies and scientific publishing

Trial-bank interoperation is where medical publishing meets artificial intelligence. To interoperate, trial banks must share a common ontology (or a common class definition) of the randomized-trial domain. I have defined an initial randomized-trial ontology for trial-bank interoperation in the Ocelot knowledge-representation system, a system that is compatible with the Generic Frame Protocol.<sup>4</sup> GFP is the forerunner of Open Knowledge Base Connectivity, the initial standard application programming interface for integrating knowledge systems developed under the DARPA-funded High Performance Knowledge Base initiative (<http://www.teknowledge.com:80/HPKB/>).

Scientific publishing can thus be a new demonstration area for knowledge-sharing technologies. Given the massive amount of information generated in domains such as

randomized trials and genomic sequencing, scientific work increasingly requires that this information be directly accessible for electronic computation. This real-world need for the knowledge representation and sharing of complex domains can anchor more theoretical work. If we hope to field new knowledge-base technologies in scientific publishing, however, we must design those technologies to be compatible with the social and economic realities of scientific publishing.

Dollars, stakeholders, and the promises of trial-bank publishing

The reality of medical publishing is that many medical editors are unfamiliar with computing, let alone artificial intelligence, technologies. Scholarly medical publishing is also subject to the interests of myriad stakeholders: publishers try to differentiate their journals so as to increase circulation and profits; editors are concerned about increasing their journal's quality, circulation, and reputation; authors seek academic credit and fairness, and want as little hassle as possible; clinician readers want information that is relevant, useful, conveniently accessible, cheap, and trustworthy. To successfully implement trial-bank publishing, we must ensure that our technical solutions fit in with the prevailing legal, social, and economic undercurrents, all of which are in increasing flux. Even if only text-based electronic publishing occurs, the power structure and economics of medical publishing are poised for significant change. With trial-bank publishing, even more uncertainties abound.

One of the primary uncertainties about electronic and trial-bank publication is its business model. Will publishers themselves administer trial banks, or will government, professional, or nonprofit organizations? Will journals contract with independent companies for trial-bank services? How will readers and information systems be billed for accessing text articles and trial-bank entries? My guess is that an assort-

ment of organizations will own and operate trial banks (thus making their interoperability even more vital). As for the out-of-pocket costs to the reader, we can only speculate. When our healthcare system finally competes on the quality rather than the cost of care, access to scientific evidence will be at a premium. Indeed, it might be health plans rather than libraries that will be the major future subscribers to electronic medical journals and to trial banks. Cost shifting in medicine is so pervasive and complex that the resulting prices for individual readers to access the clinical literature are hard to predict.

Of great importance in the prospects for trial-bank publishing is the role of academia in scholarly publishing, because it is academic physicians who generate much of the knowledge we need for evidence-based medicine. In medical academia, publication in respected medical journals is the coin of the realm for success and promotion. Given medical academia's conservatism, the traditional journals will be vetting academic quality for a while yet. The major medical journals will therefore play critical roles in leading the implementation of new modes of publishing. The challenge will be to introduce the tools and techniques of knowledge base technologies to those in academic medical publishing, and to ensure that the resulting knowledge bases all interoperate. The reasons for taking on these challenges are clear: evidence from medical research is far too valuable for it to be lost in mounds of paper or gigabytes of text files. After all, lives truly can hang in the balance of the newest research evidence.

## References

1. M.J. Cinkosky et al., "Electronic Data Publishing and GenBank," *Science*, Vol. 252, No. 5010, 1991, pp. 1273-1277.
2. F.R. Blattner, "The Complete Genome Sequency of Escherichia coli K-12," *Science*, Vol. 277, No. 5331, 1997, pp. 1453-1474.
3. I. Sim, *Trial Banks: An Informatics Foundation for Evidence-Based Medicine*, PhD dissertation, Stanford Univ., Palo Alto, Calif., 1997.
4. P.D. Karp, K. Myers, and T. Gruber, "The Generic Frame Protocol," in *Int'l Joint Conf. AI*, Morgan Kaufmann, San Francisco, 1995, pp. 768-774.

## How can we get high-quality electronic journals?

*Michael Lesk, National Science Foundation*

In 1993, Sir Peter Swinnerton-Dyer told an audience of British publishers and librarians that the United Kingdom's university system simply had no money to pay the journal price increases coming the next year. Speaking for the UK's funding agency for higher education, he gave them a choice: work with us to do something cheaper and better, or we will do something cheaper and worse.

Today, we use printing technology that is spectacularly efficient at turning out a million copies of a daily newspaper to print a tiny number of books or journal issues. Neither libraries nor publishers can afford the result.

For example, university presses now find that libraries will only buy about 250 to 300 copies of a typical scholarly mono-

*Once a university had a high-prestige section well-established, with proper attention from tenure committees, people would still submit to it even if the prize were eliminated.*

graph. To publish such quantities and still make money is no longer practical. Few if any university presses can break even, and yet there is increasing pressure from faculty members to publish their books. Despite the losses, university presses continue to publish books in increasing numbers (from 3,338 in 1975 to 4,601 in 1986 and 7,818 in 1994), adding to the pressure on libraries.

The situation in journals is no better. Serials prices rise steadily, with a 10.6% jump predicted for 1998 and the 1986-1996 period seeing a 147% increase. Every library is shedding subscriptions, dropping expensive journals and reducing monograph purchases. Fred Friend, the Librarian of University College London, claims "If I had a pound for every time I have written or said something about the

journal price rise problem, I would be so rich I could afford to buy quite a few journals!"<sup>1</sup>

Andy Odlyzko points out that Association of Research Libraries' statistics show the typical academic library spending about \$12,000 per full-time faculty member per year.<sup>2</sup> I suspect that if the faculty had this money, they wouldn't spend it all on the library. Everyone is tempted by the thought of simply putting the papers online, which would provide instant worldwide access and would seem to cost a lot less. Yet authors don't seem to want this.

## Why so little movement?

Why don't people put their best papers in online journals? We've had several years now of people such as Odlyzko and Steve Harnard pointing out how much money academia would save if scholars just posted their papers on the Web.<sup>3</sup> We've also had years of universities trying to deal with the financial support needed by libraries and university presses, as detailed by Anthony Cummings.<sup>4</sup> Yet scholarly journals continue to appear from a range of publishers, from those who have only one journal right up to giants such as Reed Elsevier Walters Kluwer—and libraries continue to buy them.

In studying the authors of scientific papers a few years ago, the Royal Society's Pam Waddell found that they did not wish to put their best papers into online journals.<sup>5</sup> They feared that they would not get adequate credit for such papers. Similarly, Mauri Collins and Zane Berge, surveying readers of an electronic journal, found that even they viewed academic administrators as less likely to give credit for electronic publication.<sup>6</sup> The fraction of their respondents who said either "an author receives *somewhat less*" or "an author receives *no*" [credit for online publications] was almost double the fraction who said an author receives as much credit for online publication as for print.

So, as we have known for a long time, unless the university reward system can be changed to encourage electronic publishing, people will continue to need paper copies of their work for review and tenure. What can be done to change the way faculty behave?

Peter Kirstein suggested to me that University College London was considering dividing its Web space into an official area,

in which material would be reviewed and approved, and a student free-for-all area. This could clearly be extended into bronze, silver, and gold areas. The idea would be to have a Web site having the prestige of a major journal, so that faculty reviewers for tenure and hiring would accept the presence of a manuscript in this area as comparable to publication by a major university press. Selection to the best areas of this Web site would be limited to refereed material, as with good journals today (and as with good electronic journals today as well). What is important is to persuade faculty to put their best material into a Web site rather than reserve it for print publication.

### Show them the money

How could a university convince people to put their best papers and books in this Web area? Once scholars start doing this, the tenure committees will immediately start to value it. The simplest way, in my opinion, is the economic solution: offer money. Suppose some university were to announce its gold area, setting up a review committee of faculty or outside experts, and offering each year a prize of \$20,000 for the best book manuscript selected for the gold Web site, \$10,000 for the best PhD dissertation, and two \$5,000 awards, one for the best faculty journal paper and one for the best undergraduate or masters research paper. This would immediately encourage faculty to supply their best manuscripts; a 10% royalty on a \$50 book that sells 500 copies is only \$5,000. A book might also appear on paper, but it would remain on the Web site. Some publishers might wish to change a paper publication of a prize-winning Web book, perhaps hoping for overseas sales.

The money involved is relatively small: \$40,000 per year. By contrast, MIT Press, a well-run and highly respected press, had a deficit of \$177,000 in 1995–1996. *The Yale Review*, just one journal, needs a subsidy of \$50,000 to keep going.<sup>7</sup>

Nearly every university already runs a Web site; the extra storage to hold even thousands of books is insignificant with disks now priced under 10 cents a megabyte. The peer review to select the prize-winning items is considerable work, but it is already being done for free. Journals such as Steve Harnad's *Psycoloquy* (<http://www.princeton.edu/~harnad/psyc.html>) already show that people will review for

online publications. For this to be successful, participants would need to accept that manuscripts would not be copyedited and that authors would need to format their own text. Although this might represent some decline in quality, authors would likely be careful about their presentation, especially if they thought it would influence the prize committee.

In fact, quality might well increase because of the increased capabilities of electronic publishing. Color pictures do not cost much more on the Web, unlike the often prohibitive costs for placing them in a printed book. Sound samples and animations can be put in a Web site; there is no easy way to do either in print. For music scholarship, as an example, the ability to include actual sound samples instead of scores should make works much more accessible to many readers. Thus, the new form of publication would have benefits to compensate the loss of professional copyediting.

Universities could try other ways to make Web publication as prestigious as paper publication. They could exhort people to believe that it was, and instruct tenure committees to value it equally. However, for several years, the UK has stated in its research-assessment exercise that online and paper publication will be valued equally, and that hasn't stopped the flow of papers in print journals from UK researchers.

More likely to succeed would be online publication (in a particular area or journal) by some highly prestigious, perhaps Nobel laureate, authors. This might cause the more junior faculty to wish to be in the same Web location. However, the senior authors involved would probably be busy with their own work and might resent being pressured to take a role in what they perceive as a dispute between publishers, libraries, and universities.

Greed is the more reliable motivator. If the prize amounts I have proposed are not adequate, they could be considerably increased and still be much cheaper than running a university press. With time, the system would be self-perpetuating; once a university had a high-prestige section well-established, with proper attention from tenure committees, people would still submit to it even if the prize were eliminated (or reduced to an unimportant amount by inflation).

So which university would like to try this first? ■

## Coming Next Issue

Automated Software  
Engineering for Building  
Intelligent Systems:  
Do We Need It?  
Can We Use It?

Guest edited by Wray  
Buntine, University of  
California at Berkeley  
and Ultimode Systems.

### References

1. F. Friend, "Why Are Librarians Talking to Publishers?" *Newsletter on Serials Pricing Issues*, No. 172, Jan. 1997; <http://www-mathdoc.math.sciences.univ-nantes.fr/NSPI/Numeros/1997-172.html>.
2. A.M. Odlyzko, "Silicon Dreams and Silicon Bricks: The Continuing Evolution of Libraries," *Library Trends*, Vol. 46, No. 1, Summer 1997, pp. 152–167; <http://www.research.att.com/~amo/doc/recent.html>.
3. S. Harnad, *Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing*, A. Okerson and J. O'Donnell, eds., Assoc. Research Libraries, Washington, D.C., 1994.
4. A.M. Cummings et al., *University Libraries and Scholarly Communication*, Assoc. Research Libraries, 1992; <http://www.lib.virginia.edu/mellon/>.
5. P. Waddell, "The Potential for Electronic Journals in UK Academia," *Libraries and IT: Working Papers of the Information Technology Subcommittee of the HEFC's Libraries Rev.*, Office for Library and Information Networking, Bath, UK, 1993, pp. 239–271.
6. M.P. Collins and Z.L. Berge, "IPCT Journal: A Case Study of an Electronic Journal on the Internet," *J. Am. Soc. Information Science*, Vol. 45, No. 10, Dec. 1994, p. 773.
7. "Writers' Campaign Saves the *Yale Review*: University President Changes His Mind," *San Francisco Chronicle*, Wed., Nov. 14, 1990, p. A22.