
Common and contrasting online-repository effects across two disciplines.

Issues of Online Research Repositories from the Perspective of the Biomedical Sciences

David L. Armbruster
Health Sciences Library
877 Madison Avenue
University of Tennessee, Memphis
Memphis, TN 38163
darmbruster@utmem.edu

Abstract

This commentary on Joseph Y. Halpern's proposal for a computing research repository discusses difference in traditions and practices of online publishing and repositories between computing and biomedical sciences. Issues of accessibility and archiving are also discussed.

H.3.7 Digital libraries—biomedical sciences, publishing, online, archiving, accessibility

Introduction

Between October, 1999, when I first read Halpern's article, and January, 2000, when I completed this commentary, so much happened (and will continue to happen) in the area of online repositories that my comments may be of little import, even passe, as they are read in May, 2000.

With that caveat, let me state that the purely computer-related issues of CoRR are not within my area of expertise. Nonetheless, Halpern identifies issues relating to online repositories that clearly transcend the computing world. I want to comment at some length on two of these: (1) online publishing and repositories in the biomedical sciences and (2) archiving and accessibility. I will conclude with comments on a few additional topics relevant to online scientific communication in general.

Online Publishing and Repositories in the Biomedical Sciences

Halpern is correct to point out that different disciplines have different traditions in publishing and in sharing information, even in the world of paper publishing, but especially in the new world of online publishing.

The tradition of publishing in the biomedical sciences is considerably different from that of the preprint world of high-energy physics or of computer science, at least as described by Halpern. Biomedical journal publishers, both for-profit and not-for-profit, have stated clearly in instructions to authors that they are not interested in publishing results that have appeared previously (prior publication). In the late 1980s, a few publishers even went so far as to state that a conference proceedings abstract of more than, say, 700 words would be considered prior publication. That specific practice has changed, and most instructions to authors today explicitly state that publication in the form of an abstract is not considered prior publication.

However, many journals include in instructions to authors an explicit warning that work published on a Web site, including an author's personal Web site (assuming it is available to the world), is considered prior publication, and such an article will not be considered by the journal. For example, the *American Journal of*

Physiology, in its December, 1999, instructions for authors, states "...the journals of the APS [American Physiological Society] will not accept manuscripts in which...a significant portion of the data in the form of figures and tables has been published elsewhere. Prepublication in ...electronic media is also considered duplicate publication." *JAMA [Journal of the American Medical Association]* states in its instructions for authors (January 5, 2000), "Manuscripts are considered with the understanding that they have not been published previously in print or electronic format...." In the January 1, 1999, issue of *Science*, information for contributors states, "Science will not consider any paper or component of a paper that has been published.... Distribution on the Internet may be considered previously published material. We reserve the right to determine if such an Internet posting compromises the originality of the paper as a submission to *Science*." Many other journals simply state that prior publications will not be considered and leave it to the potential author to figure out whether that prohibition includes online publishing.

On the other hand, in an editorial of January 16, 1999, *BMJ (British Medical Journal)* editor Richard Smith (1999) wrote: "Communication in science and medicine will not be well served by standing in the way of publication in many versions, and the BMJ is willing to consider for publication eprints that have been posted on Web sites so long as their status as eprints is clear."

It's interesting to note that at the end of 1999, BMJ launched clinmed.netprints.org, "an electronic archive where authors can post their research into clinical medicine and health before, during, and after peer review by other agencies. Resulting from a collaboration between the BMJ Publishing Group and Stanford University Libraries, it will allow researchers to share their findings in full, for free, and as soon as their studies are complete" (Delamothe, T., Smith, R., Keller, M. A., Sack, J., and Witscher, B., 1999).

Despite publishers' restrictions, online repositories and similar ventures are burgeoning.

Project MUSE, begun in 1995 by Johns Hopkins University Press, is a database of electronic journals published by 10 university presses; others are expected to join. Although subscription fees are charged, they are being kept to a minimum. Additional information can be found at <http://muse.jhu.edu>.

PubMed Central, announced in August, 1999, by Harold Varmus, former director of the National Institutes of Health, has been both highly praised and vilified. A much more comprehensive and inclusive resource than the National Library of Medicine's PubMed and announced earlier as E-biomed, PubMed Central will be a "Web-based repository for barrier-free access to primary reports in the life sciences" (National Institutes of Health, 1999).

In the fall of 1999, an international group of academic and government scholars proposed Open Archives, a free online repository for the archives of the universities throughout the world (Kiernan, 1999). Like PubMed Central, Open Archives would include both peer-reviewed and unreviewed articles.

The immediacy and ease of communicating through the Internet have certainly been compelling reasons for online repositories. But I wonder whether scientific publishers themselves—especially those opposed to the idea of free and open online repositories—may have provided impetus for developing repositories and communicating more and more science via the Web.

Throughout the 1990s, librarians, publishers, researchers, and university administrators, among others, have argued long and vehemently at times over the high cost of library periodical subscriptions (e.g., about \$17,000 per year for a subscription to *Brain Research*). Librarians and university administrators have complained about the cost of library subscriptions for scientific journals, and publishers have tried to justify their pricing structures. On average, subscription costs have risen 10% per year since the late 1980s. Regardless of the rationale for such pricing, the library community has become increasingly vocal about the financial problems

resulting from these increases. Administrators often don't fully appreciate that libraries and their attending costs for space, staff, and collections are not going away simply because "everything is now on the web." So requests for increased funding for simply maintaining the subscriptions the library already has, much less adding to the collection, go unfulfilled.

Researchers grumble loudly about being required to spend research dollars on interlibrary loans, because their library can no longer justify subscribing to expensive research or clinical journals needed by the researchers. While some universities earmark a certain percentage of the indirect costs from research grants for library books and subscriptions, some do not. So being the generally creative types they are, researchers have joined forces with technology to begin finding ways around the system. Organizations and publishers like the BMJ are realizing that "if we can't lick 'em, we'd better join 'em."

As an editorial aside, I strongly believe that within the next few years, for-profit (and some not-for-profit) publishers will figure out a way to continue making money even while authors submit their research findings to online repositories, and many of the problems and restrictions we see today in the biomedical sciences will diminish. In the Information Age, we must look at new ways of doing business, and scientific research and communication are included.

Accessibility and archiving

Accessibility has several components, in addition to the technical issues addressed in Halpern's proposal. Others include subscription vs. free access, open vs. controlled access, and stand-alone vs. networked (e.g., online repositories of repositories much like bibliographies of bibliographies), and these must be decided at the outset by those developing and establishing

The benefits of an open...
online scientific
communication process and
the mechanisms for archiving
that communication are
enormous and obvious, but
none of these is free.

online repositories. In addition, robust search engines and relational databases are essential to information access in online repositories.

Another aspect of accessibility is information that used to be very difficult, if not impossible, to find can now easily

be contained in accessible online repositories. An example is electronic master's theses contained in the Networked Digital Library of Theses and Dissertation (NDLTD). Although we have been able for a long time to locate most Ph.D. dissertations through *Dissertation Abstracts* (now online so we don't have to wait for the annual print publication) and to acquire them in print and other forms from Bell and Howell Information and Learning (formerly UMI and University Microfilms, Inc.), most theses and, importantly, the information (e.g., data, bibliographies, and literature reviews) contained in them have languished in university libraries; often only the author, faculty advisor, and cataloging librarian knew of their existence. With the advent of the NDLTD and the pioneering work done at Virginia Tech, electronic theses (and dissertations) are becoming accessible worldwide through the Web.

The bottom line to easily accessible online repositories may be financial. Who will pay initially to develop, implement, and maintain the online repository? Who, if anyone, will make money from the venture? Who may lose money in the process? The benefits of an open, free, online scientific communication process and the mechanisms for archiving that communication are enormous and obvious, but none of these is free. While governmental agencies are likely candidates for providing resources through grants to develop, implement, and maintain online repositories, their inherent bureaucracy dictates against entrusting the entire venture to them. The less-bureaucratic,

dynamic, for-profit publishing community may well find that adding value to basic governmental online repositories can provide new avenues for revenue.

Archiving, also, is a multifaceted issue. I would certainly agree with Halpern's recommendation of using the latest version of a document as the default; this approach could, in fact, be critical in a clinical situation, for example. However, using pointers to earlier versions should also be a requirement. Although Halpern discusses input logistics, he doesn't discuss the matter of what to include. Drafts? Preprints? Nonpeer-reviewed manuscripts? Peer-reviewed manuscripts? Published papers? A combination? While these decisions are important and certainly must be defined for contributors and users, even potentially more important aspects of archiving are permanency and future accessibility.

Halpern's hope that translation programs will be written to maintain access to (or viability of) older online documents and information is one that certainly has more chance of occurring in the CoRR community than in some others. This archive issue is of grave concern to the library community in terms of archiving electronic theses and dissertations. Institutions that have existed for more than 25 years, including libraries, often have electronic files that can no longer be accessed, because neither the software nor hardware is available (except, perhaps, in a museum or historical collection) to access the data. This issue is perhaps less important in rapidly changing sciences than it is for the humanities, where very old work is often researched.

Another important access and archiving issue in the library community regards whether a library that has a subscription to an online journal will always have access to the subscribed issues in the event the subscription ceases. Currently, this question must be handled by each library with each vendor or publisher through contract negotiation. We are seeing library consortia proliferating in response to the publishers' divide-and-conquer approach to dealing with libraries.

A companion question that must be answered is: What happens to information contained in online journals that cease to publish? Will researchers and libraries have access to the information? How will access occur? Through CDs? Through online repositories? Through other media or mechanisms? An online repository such as CoRR would obviate this problem, assuming it (or other online repository) has ensured open and free access.

Additional Issues

Halpern cites several value-added services publishers might provide: summary and abstracting services, advanced searching tools, current awareness reports, and so on. Value-added initiatives seem to be appearing with increasing frequency. Although all may not survive, they are at least exciting experiments for the online publishing environment.

McCollum reported (1999) on a value-added initiative by 12 scientific journal publishers. Sometime in 2000, the coalition plans to link footnotes in their online journals to at least the online abstract, if not the complete article, in the journals of other participating publishers. The publishers involved are John Wiley and Sons, Academic Press, American Association for the Advancement of Science, American Institute of Physics, Association for Computer Machinery, Blackwell Science, Elsevier Science, Institute of Electrical and Electronics Engineers, Kluwer Academic Publishers, Nature, Oxford University Press, and Springer-Verlag.

The issue of CoRR's difficult user interface resulting from the fact that CoRR is currently a volunteer effort is important to anyone considering a similar (ad)venture. All too obviously, many Web site were started as a seat-of-the-pants experiment by volunteers with little or no support. While the number of online repositories is increasing, government and private funding must be available for research and experimentation into various ways to create and maintain them. Usability research—in many respects a "brave new world"—is burgeoning. In the world of special-

ized online repositories, successful paradigms are probably few and depend for their success on a relatively small, focused, sophisticated user group. The attributes of successful paradigms will probably increase quickly, because users of the Web are becoming more numerous and less focused.

Finally, Halpern's vision of an online repository for computer science research is one that should be pursued vigorously, just as similar efforts in other disciplines are being pursued. Different disciplines will no doubt create different models and mechanisms for online repositories. In the process, creators of these repositories are going to trip, scrape a knuckle, fall, even fail. But better models and better mechanisms will result. Some of the issues that were relevant in 1999 have already been resolved in early 2000; other issues will take more time to solve; new ones will appear in 2001. The Electronic Information Age is creating a very different world than existed even in 1990. The academic, governmental, and business sectors involved with scientific communication must find common ground from which to build the online infrastructure described by Halpern.

References

- Delamothe, T., Smith, R., Keller, M. A., Sack, J., and Witscher, B. (1999). Netprints: the next phase in the evolution of biomedical publishing. *BMJ*, 319, 1515-1516.
- Kiernan, Vincent. (1999). 'Open archives' project promises alternative to costly journals. *The Chronicle of Higher Education*, December 3, 1999, 46(13), A43-44.
- McCollum, Kelly. (1999). Publishers of on-line journals plan to link millions of science footnotes. *The Chronicle of Higher Education*, November 26, 1999, 46(14), A21.
- National Institutes of Health. (1999). <http://www.nih.gov/welcome/director/pubmedcentral/pubmedcentral.htm>. Accessed January 27, 2000.
- Smith, R. (1999). What is publication? *BMJ*, 318, 142.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that all copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. © 2000 ACM 1527-6805/00/05—0049 \$5.00