

For better or for worse

Preprint servers are here to stay

by Peter B. Boyce

Where do preprint servers fit into the scholarly information system? Will they replace, as some of the more vociferous proponents claim, the entire system of scholarly publishers?

Let's try to cut through some of the rhetoric and take a look at preprint servers and the functions they perform, contrasting them with a fully interlinked electronic information system, which includes the better conventional journals.

As a background, consider the functions provided by the journals before the days of widespread electronic communication. I listed five roles for journals in earlier articles.^{1,2}

- Status: keeping the community abreast of where expertise resides, institutional activity, etc.
- News: disseminating the latest research results to the community.
- Information: providing a repository for the body of knowledge about a particular field.
- Author evaluation: providing a means for judging the competence and effectiveness of authors.
- Historical: maintaining a record of the progress of science through the years.

The status and news functions have always been popular with scientists. In the days of the paper journal, everyone used to wait eagerly for the arrival of the new issues of

the journals (and also for paper preprints) to see the latest results, to find out which people were working on the hot topics, and to learn what places might be moving into a new research area. Now they read the daily postings on the preprint servers to get this information.

Preprint servers and science

The other three functions are even more important. In many subject areas, particularly in the sciences, the main corpus of knowledge resides in articles published in the journals. Maintaining this knowledge is vital to the health of the discipline. The record of advances contained in the journals is also important for understanding the history of the discipline. Finally, everyone recognizes that publication in a well-respected journal is critical to advancement in one's career.

Electronic preprint servers fill the status and news roles very well. New results propagate through the community overnight. I believe this focuses more attention on what competing scientists are doing than is healthy. It seems to be encouraging premature announcement of results; being first now counts more than being thorough (or even completely correct?). However, peer pressure from colleagues does seem to keep the quality of the submissions higher than might have been anticipated. And competition is an excellent stimulus to work hard. In sum, the preprint

About the author

Peter B. Boyce is senior consultant for the American Astronomical Society and principal of P. Boyce Associates in Nantucket, Massachusetts, e-mail: pboyce@aas.org

Editors' introduction

In our inaugural column (*C&RL News*, January 2000) we promised to reach outside the library community to illuminate the changes taking place in scholarly communication from a variety of perspectives. This month's article does exactly that.

We asked Peter B. Boyce, longtime executive officer of the American Astronomical Society (AAS), to comment on the effect that preprint initiatives are likely to have on scholarly publishing. Are these initiatives a good or a bad thing from the perspective of a society publisher? How will they affect the publishing landscape? Is formal publication still needed, and if so, why?

Boyce has responded by contrasting the roles of preprint servers and formal journal publication, and goes on to envision a unified framework within which both communication modalities can coexist successfully.

Boyce was a natural choice to address this subject. During his 19-year tenure at AAS, he developed and led a program on electronic publishing and the use of the Internet.

He has been active in promoting collaboration and interlinking among the various information providers in astronomy, including the electronic journals, data centers, and the central database of astronomical abstracts.

Boyce has lectured extensively on electronic publishing in the U.S. and abroad, and is a frequent contributor to lists such as Liblicense-L.

Though officially retired and now living part of the year on Nantucket Island ("which," he writes, "almost qualifies as a foreign country"), he is still consulting on electronic publishing for the AAS and others. —*Ivy Anderson, Gail McMillan, and Ann C. Schaffner*

servers in physics and astronomy are, at this time, an excellent resource for keeping up with the latest news as well as for fostering competition.

Long-term effects

The other roles filled by journals have not been well filled by preprint servers, and may never be. As a repository of older information, the servers are incomplete, their internal storage format is not archivally robust, and the author is free to send in a new version of the paper at any time.

These destabilizing factors lead me to be very skeptical about the capability of preprint servers to serve as long-term archives. And without assured long-term stability and the imprimatur of a well regarded peer review process, self publication in the electronic preprint servers will not count for much toward academic tenure and job promotion.

When people claim that the preprint server will take over the full role of publisher, I again become skeptical. Good publishing is harder work than most people realize. Good authoring technique is also difficult. And copyediting is very important for most non-English-speaking authors, and even for a good percentage of native English-speaking ones. Consistency of expression and clarity of communication are what is desired. But more than that, a high precision of statement is required in the electronic world. References have to be entered correctly by the author to link correctly. Metadata has to be in a standard format to be useful. Keywords have to be consistent among different articles. Authors are not particularly good at doing these things correctly. More automated tools may help, but automation can not fix everything.

So, while electronic preprint servers have an important role to play in the scholarly information scheme, they fall short of filling all

About the editors

Ivy Anderson is coordinator for Digital Acquisitions at Harvard University, e-mail: ivy_anderson@harvard.edu; Gail McMillan is head of the Digital Library and Archives (formerly the Scholarly Communications Project) at Virginia Tech University, e-mail: gailmac@vt.edu; Ann C. Schaffner is associate university librarian for Research Services, Instruction & Planning at Brandeis University, e-mail: schaffne@brandeis.edu

the roles of the traditional journals. Preprint servers are designed for rapid and automated dissemination of individual articles, at the expense of ease of use, good linking to other resources, reasonable search capability, and integrity of the original submitted paper.

“ADS,” a linked, searchable abstract database

Many users, in their quest to keep up with daily advances, overlook these shortcomings. Astronomers, however, are fortunate to have the best of both worlds. Most users in astronomy, where good, well-designed electronic journals have been around for five years, go to the journals and to the “Astrophysics Data System” (ADS), a linked, searchable abstract database for older information, and use the preprint servers to keep up with the critical advances in their own area of work. Probably no other science has such a highly interlinked set of journal-based information resources available to the community.

“ADS,” a NASA-supported database of abstracts of all core literature, is a central resource in its discipline.^{3,4,5} The abstracts link to the full text, either in the electronic journals or to scanned page images, for the historical literature. The core literature in astronomy is available in page image format back to the first issue of the *Astronomical Journal* published in 1849. The abstracts also link to references, future citations, and published data tables, which are also online at various distributed international data centers.

Users can enter the database with the name of an astronomical object, get the published data on that object, and link to the articles where those data were published. The electronic journals not only have abundant navigational links to make them easy to browse on the screen, but they also link to the references through the “ADS” abstract database. PDF versions are available for local printing (although this does not capture the video), large data tables, or other electronic enhancements contained in the electronic version. Together this linked set of distributed resources offers much more than the preprint servers in terms of ease of use, available information, and breadth of coverage.

Unfortunately most traditional publishers have been slow to understand and to use the

full capabilities of the electronic environment. By and large, they are simply reproducing electronically the same old journals they had in paper—often just using PDF page renditions, which are nearly worthless for reading on the screen in comparison to a well-designed HTML article. Most publishers have been slow to realize the value of links and even slower to incorporate them, although that is changing now. In comparison to the average electronic version of a journal of today, preprint servers don’t seem half bad, and the information gets out rapidly. No wonder there is such enthusiasm for them.

An evolving system

It is obvious that we are in a period of rapid change. We are just learning to use the electronic environment effectively, and experimentation is important. The journal-based astronomical information system serves users well for now, but it will undoubtedly evolve. Richard Luce’s thoughtful article in this column illustrates how preprint servers are working to evolve.⁶ He notes a number of very important objectives, which are being addressed by the Open Archives Initiative:

- Integrating the preprint system into the larger system of scholarly information has to be done. The appearance of this goal on their list is important and welcome.
- Better search and retrieval capability, especially across disciplines, will be a boon to the busy scholar.
- Developing user friendly systems is imperative for preprint servers (which one astronomical user has dubbed “user belligerent”) and for the publishers, not many of whom put the needs of their users above their own considerations.
- Including the full range of metadata, full-text, and citation data. This complete range of information has been part of the astronomical journal-based system for five years. The resulting usage statistics, as well as abundant informal feedback, demonstrate the importance of including all of this information somewhere within a linked collaborative system.

All of this does not have to reside in one place. The concept behind the Web is that distributed resources can work as one system. Each contributing resource can be tended to by the experts who can make it

function best, and the whole community of users benefits. The scholarly information community is only beginning to make this a reality. The key, as Luce has said, is interoperability. My hope is that the political problems can be solved to make this a broad reality across the scholarly community. The technical problems pale in comparison.

In summary, we publishers have to admit that there is a role for preprint servers as vehicles for announcing new results. Some publishers have also re-engineered their procedures to incorporate new electronic tools and reduce costs. All of this is good for improving communication. Sometimes it seems as if progress is painfully slow; but since we are experimenting with the very fabric of our scientific knowledge, it is good to go cautiously. It would not do to inadvertently plunge our system of scholarly information into chaos.

My fear is that by the time the problems get worked out, the paper-based model of the article, or the monograph, as the fundamental vehicle for scholarly communication will be superseded. Perhaps in five years we will have a system of electronic "knowledge clusters" mined by a set of "bots" to produce, on the fly, a comprehensive and up-to-date set of distilled information in response to a

user's request. The long, linear, paper-based model of today's journal article or preprint would have no place in such a system. But that's a subject for another column.

And . . . by the way, I would prefer not to call the preprint servers "archives" until we are sure we will be able to read the material 50 years from now.

Notes

1. Peter B. Boyce, "A Successful Electronic Scholarly Journal From a Small Society" (presented at ICSU Press-UNESCO Expert Conference on Electronic Publishing in Science, Paris, France, February 19-23, 1996), <http://www.aas.org/~pboyce/epubs/icsu-art.html>.

2. Peter B. Boyce, "Building a Peer Reviewed Scientific Journal on the Internet," *Computers in Physics* 10, (1996): 216.

3. Peter B. Boyce, "What Does the Future Hold? Ask an Astronomer" (presented at NC Serials Conference, Chapel Hill, March 16, 2000), <http://www.aas.org/~pboyce/epubs/NCSerials2000/NC2000.html>.

4. Michael J. Kurtz, et. al., "The NASA Astrophysics Data System: Overview" p. 41 and ("For better or for worse" cont. from page 000) other articles in "The CDS and NASA ADS Resources: New Tools for Astronomical Re-

(continued on page 414)

Elegant Solutions for Preservation



A DIVISION OF LIBRARY BINDING SERVICE



- _____
Pamphlet Binder
- _____
Music Binder
- _____
Archival Folders
- _____
Manuscript Folders
- _____
Hinge Board Covers
- _____
Academy Folders
- _____
Newspaper/Map Folders
- _____
Bound Four Flap Enclosure
- _____
Archival Binders
- _____
Polypropylene Sheet & Photo Protectors
- _____
Archival Boards

ARCHIVAL PRODUCTS

A Division of Library Binding Service

**Call
for a complete
catalog**

P.O. Box 1413
Des Moines, Iowa 50305-1413
800.526.5640
Fax 800.262.4091
custserv@archival.com
<http://www.archival.com>

for the benefit of their readership. Janet Fisher, from MIT Press, spoke about the lack of conflict between ETDs and scholarly journal publications due to the major rewrites that chapters require to become journal articles and the importance of the academic journals' peer review processes.

When questioned about the likelihood of plagiarism when research is available through the Internet, all three panelists objected to the notion. Garfield pointed out that plagiarism is also easier to detect and Lewis added that those who are intent on doing it will. Fisher said there is no evidence that plagiarism has increased with the availability of journals online.

The librarians panel included Roy Tennant, California Digital Library Project; Julia Blixrud, ARL; and David Balatti, National Library of Canada. Tennant focused on the problems, reiterating some of Gladney's concerns, especially that libraries are not funded for digital archiving. He noted that ETDs are "a good thing" and that "barriers will fall in time." Blixrud addressed whether to adopt old strategies to new media. Balatti shared background information on the Canadian Theses Service.

Hands-on networking opportunities

There were many quotable moments, of course. Bruce Cochrane (University of South Florida, Evolutionary Genetics), for example, told us about being "digitally-abled" and the "quill pen problem." Daniel Ferreras (West Virginia University, Foreign Languages) implored us to empower students to show us the way. "We are preceding history in the making."

Concurrent presentations included case studies from many universities at varying stages of ETD projects. These included Harvard, Humboldt University (Berlin), MIT, Michigan State University, University of Arizona, University of Georgia, University of Iowa, University of Waterloo, Virginia Commonwealth University, West Virginia University, as well as Virginia Tech. Australian and German national projects were also described. There were hands-on opportunities to learn about a variety of topics including HTML, PDF, SGML, XML, XSL/T, and metadata.

The conference was also an opportunity to network informally, especially during Friday evening's reception which was accompanied by music performed by West Virginia

University's John Hagen (ETDs/Library) and Daniel Ferreras (faculty/Foreign Languages). The Dalí Museum was opened to us for an evening reception and self-guided tours. Lunches also afforded conference-goers the opportunity to enjoy the outdoors, especially the palm trees and the bay.

On the day before the conference, the Networked Digital Library of Theses and Dissertations (NDLTD) Steering Committee met to discuss accomplishments and to focus on future directions. Three subcommittees were established, with membership later drawn from volunteers among the conference attendees. In addition to planning next year's conference, the topics are training graduate students and standards and preservation. Each committee held initial meetings at the conclusion of Thursday's conference lunch, but new members are welcome. Progress reports will be prepared for the next Steering Committee meeting in Washington, D.C. on September 29, 2000.

NDLTD is a loose federation of 78 members from universities throughout the world, more than double the membership at this time last year when the conference was held in Blacksburg at Virginia Tech. At that time, two universities required that graduate students deposit ETDs. Four universities now require ETDs and there are more than 7,500 ETDs available from the member institutions. Next year the 4th International Symposium on Electronic Theses and Dissertations will be held at Cal Tech in Pasadena on March 22–24, 2001.

Note

1. See sites from the conference wrap-up session at <http://scholar.lib.vt.edu/presentations/ETDs2000/wrapup.pdf>. ■

(*"For better . . ." continued from page 407*)
search," a *Special Issue of Astronomy and Astrophysics, Supplement Series*, 143, no. 1 (April 2000), <http://www.edpsciences-usa.org/articles/astro/abs/2000/07/contents/contents.html> NASA Astrophysics Data System.

5. The NASA Astrophysics Data System is located at <http://adswwww.harvard.edu>.

6. Richard E. Luce, "The Open Archives Initiative: Forging a Path Toward Inter-operable Author Self-Archiving Systems," *C&RL News* 61 (March 2000): 184–186, 202. ■