

Preprint Servers: Status, Challenges, and Opportunities of the New Digital Publishing Paradigm

Sharon M. Jordan

U. S. DOE [Office of Scientific and Technical Information](#)

InForum '99, May 5, 1999

The digital revolution that Dr. Warnick was just talking about is also challenging how the scientific community deals with preprints. Two years ago, we heard here at InForum from Dr. Paul Ginsparg about the [Los Alamos National Laboratory \(LANL\)](#) e-print archive. His pioneering effort was a significant turning point for scientific publishing. Nowadays, you can hardly find an article about preprint servers or e-prints or electronic journals that does not mention the Ginsparg or the LANL archive.

There has been explosive growth in electronic preprints and preprint servers. Let's review this growth. How far has it gone? In reviewing the "literature" -- that is, conducting an online survey of preprint servers on the web -- there was observable change from two years ago.

Background

Scientific progress is largely based on the open exchange of research data and results among scientists. Besides the publication of technical reports that we are familiar with -- and that are now solidly making the transition to electronic access -- the traditional scientific publishing paradigm for research papers has been the long-accepted model of the scientific journal. Journals contribute to the scientific process as the means to not only communicate but also to certify the validity of the information through a peer review process and to establish a prestige ranking order among scientists. The journal publishing world is changing too; this will be discussed in a session on e-journal initiatives here this afternoon. So what is the status of preprint servers?

What Is a Preprint?

Preprints are manuscripts that have not yet been published, but may have been reviewed and accepted; submitted but with no publication decision; or intended for publication and being circulated for comment.

Preprints available on the web may also be referred to as "e-prints." Most e-prints are fairly complete research reports in electronic versions of papers that have already been *submitted*: (1) for dissemination, comment, and review among peers; (2) for publication in journals; or (3) prior to presentation at conferences.

Similarly, the [American Physical Society \(APS\)](#) defines e-print as a preprint (pre-published article) in electronic form. The APS also states that the concept is now more nebulous and is general enough to include any electronic work circulated by the author *outside of the traditional*

publishing environment. On many preprint servers, it means any electronic (not necessarily printable), research-related information provided by the author.

In our own Guide to STI Management issued last year ([DOE G 241.1-1](#)), we defined preprint as a document in pre-publication status, particularly an article submitted to a journal for consideration for publication - a definition that may have been too narrow in scope. We also took care to define a “postprint” as a document in post-publication status, particularly an author's article or paper, after it has been published in a journal. However, most of the sites I reviewed recently did not make that distinction, but some services do add information about where the preprint has been used.

Thus, the practice in general is that un-refereed preprints are initially deposited at preprint servers and the author may replace them with revised versions that have been accepted for publication (that is, a refereed, published reprint in the terminology of some and the STIP equivalent of a postprint). At the LANL e-print archive, the fact that every paper posted is intended by the author to be submitted to a journal and thus subject to peer review motivates the author to be accountable to certain standards.

Preprints are available on the web long before the actual published paper, and increasingly the e-print versions are being cited in scholarly journals because timeliness of information paces scientific advancement.

Making them available on the web only makes sense if you consider why scientists publish:

- To ‘advertise’ or make their work known (i.e., to be *visible*)
- To contribute to or build upon the body of scientific knowledge (i.e., to be *used*)
- To receive credit or be acknowledged for their work (i.e., to be *praised*)

Visible, used, and praised. That sounds familiar. Wasn't that the theme of last year's InForum and isn't making R&D results visible, used, and praised still a goal for us?

The Status of Preprint Servers

The raw numbers of “hits” indicate significant growth and availability of preprints. A simple web search for “preprints” will yield tens of thousands of items. The number of preprints submitted to the LANL site each month now averages over 2000. In astrophysics alone, the number has doubled every year since introduced in 1992. Narrow a web search to “preprint servers,” and the result is just hundreds, though of course, you have to refine your search to get to the actual preprint servers. Still the number of hits reflects a lot of literature available on the subject. The finding: there is a growing acceptance and importance of e-prints; they have become a more common form for exchanging scientific information.

Disciplines

Physics: In terms of the scientific disciplines covered, the predominance is in physics. Scientists in physics as well as astronomy have shared their draft papers (that is, preprints) with trusted colleagues for years. They've simply changed the manner in which they shared as new technology has allowed: from mail, to fax, to Internet - quickly adapting to the latest and fastest communication technology. Thus, physicists have learned in recent years that, to obtain the most current information, they must go to preprint servers.

For some fields of physics, publication on the Web (by posting to the preprint archive at LANL for open public comment) has become a required prelude to publication in the journals. Only after the manuscript has received comment by others in the field, whose remarks are electronically associated with the original, do researchers in theoretical particle physics or related fields submit a "finished" manuscript to the printed physics journals.

Mathematics: Certain philosophers of the preprint revolution envisioned a couple of years ago that "unless there is something unaccountably different about physics, the other disciplines should follow the same pattern..." And, in fact, in January 1998 LANL added a mathematics archive, followed by computer science.

Mathematics preprints had been covered by the [American Mathematical Society's \(AMS\) Preprint Server](#). However, the AMS has restructured this past year, closing down its preprint server but now maintaining a listing of all math preprint servers in a new directory. It is noteworthy that 14 preprint servers previously listed on the AMS page have now been incorporated in the LANL preprint site.

Biology: Although the practice of sharing preprints with colleagues is seen less often in biomedical research, biology is another discipline now turning to the web as a reasonable way to share new data in certain areas where it is beneficial to visualize large sets of data and concepts and build upon collective information, such as in the area of genomics.

One year ago, the [British Medical Journal \(BMJ\)](#) opened an online "Preprint debate." The result was announced by the editor as a decision to "dip their toes into the water" of online peer review by posting preprints for online commentary.

BMJ determined that the long process of formal journal publishing is detrimental to science and the public health. Their position is that the medical community in particular would benefit from the ability of researchers to include "unpublished preprints" in their review of the literature, and that any way of getting scientific advances in the public domain fast is work-supporting. The observation was that researchers will use preprints because of their benefits. To paraphrase the editor: "The web gives us the opportunity to decide how we add value in the dissemination of STI... Perhaps we're not adding enough."

Also about one year ago, in June 1998, human geneticists and molecular biologists were able to post their preprints on the Web. The service is part of a nonprofit web site (called **HUM-MOLGEN**) run by U.S. and European scientists-editors. Currently, there are over 5000 subscribers and over 40,000 participants. It is interesting too that, reportedly, Dr. Ginsparg noted that the biologists could have joined his infrastructure rather than establishing a separate one.

Some Key Preprint Sites

Hundreds of the “preprint servers” located in my review were the sites of individual researchers or departments of universities. A few examples are: Preprint Archive of the Institute of Particle Physics and Astrophysics at Virginia Tech (<http://www.phys.vt.edu/>); UC Berkeley Astronomy Preprints (<http://astro.berkeley.edu/preprints.html>); University of Stony Brook Department of Applied Mathematics and Statistics (<http://www.ams.sunysb.edu/papers/papers.html>); or University of Wisconsin Mathematics Department Preprints (<http://math.wisc.edu/Preprints/>).

A fair number hosted international preprints within a scientific discipline. Many of those “cross-linked” to other sites. Here are some of the key preprint sites:

LANL’s E-Print Archive (xxx.lanl.gov) was started in August 1991. It is a well-organized and administrated infrastructure that abides by certain rules and standards and has advisory boards. It now covers physics, mathematics, nonlinear science, and computer science. Full-text preprints are made available in electronic formats. The site has 14 mirror sites in 14 countries, and this archive is frequently viewed by others as the model. Usage statistics are available online (http://xxx.lanl.gov/cgi-bin/show_monthly_submissions and http://xxx.lanl.gov/cgi-bin/show_weekdays_graph).

CERN Document Server: Preprints (<http://preprints.cern.ch/>) keeps monthly listings from 1994 onwards of preprints from CERN, preprints received and scanned at CERN, and some of the Los Alamos e-print archives. Full text is provided. Another CERN site, HEP preprint servers and databases (http://wwwas.cern.ch/library/preprint_servers/hep_servers.html), provides “HEPDOC” - this provides searching of CERN, SPIRES, and KISS sites in one search.

The CERN site also links to ten other significant preprint sources, including four DOE laboratories: BNL (Brookhaven); Fermilab (Batavia, IL); SLAC SPIRES (Stanford); and LANL (Los Alamos).

SLAC SPIRES-HEP (Stanford Public Information Retrieval System - High Energy Physics)(<http://www-slac.slac.stanford.edu/find/spires.html>). The HEP preprint database contains almost 400,000 bibliographic entries with 180,000 entries linked to full-text (hosted by various sites) and grows by about 20,000 documents per year. Included in that number are preprints, journal articles, technical reports, theses, and other documents. The SPIRES site notably points out that preprints are a component of the collective scientific and technical information (see <http://www.slac.stanford.edu/slac/sciinfo.shtml>). The collection is searchable by author, title, report number, institution, collaboration, and more. Postscript versions of selected preprints are

available for viewing or abstracts of the e-print archive papers may be read. Pat Kreitz, one of the STIP participants here today, has a key role in the SPIRES site and has set the example for preprints being an STI function. She noted recently that most preprint sites are used by researchers at their desktops, mostly for current awareness, rather than being accessed by library users.

American Physical Society (APS) E-Prints (<http://publish.aps.org/eprint/>) was launched in prototype form in July 1996, with full-fledged service reached a few months later. The APS E-print server now regularly receives 40-50 papers per month for public display. While this is only a few percent of the volume available at the Los Alamos e-Print archive, the two servers do not completely overlap, and the APS server may provide an outlet for authors who perhaps would not have used e-prints otherwise. It has been an important proving-ground for a variety of web-based technology at the Physical Review publications office, particularly in relation to continuing innovations in web-based submissions. According to information on the site, the server will probably continue, but will likely be more closely tied to the journals themselves, since APS views these technologies as central to the future of communication and publication of physics information. The site points out that the e-print system is NOT a publication of the American Physical Society, and therefore no editorial control is extended to the content.

American Mathematical Society (AMS) Global Directory of Preprint and e-Print Servers (<http://www.ams.org/global-preprints/>) has been created by the AMS as part of the e-MATH site, located in the Publications and Research Tools section. The goal of the Directory is to maintain an updated directory of all mathematical preprint and e-print servers worldwide so that mathematicians can browse through their rich content or post their own research on the server of their choice. The directory does not offer full-text but is a tool to locate the various available servers, some of which do host full-text preprints.

Chemical Physics Preprint Database (<http://www.chem.brown.edu/chem-ph.html>) is a fully automated electronic archive and distribution server to host full-text preprints for the international theoretical chemistry community. The database purportedly provides rapid and efficient preprint distribution within the international chemical physics community. The project is a joint effort by the Department of Chemistry at Brown University and the Theoretical Chemistry and Molecular Physics Group at the Los Alamos National Laboratory. The web site acknowledges the assistance of Paul Ginsparg at Los Alamos National Laboratory, where the full-text preprints are hosted.

Recent Initiatives

The **National Institutes of Health** announced this spring that it is considering a proposal for web-based publishing of biomedical research papers. The fact that the Director of NIH, with the huge resources at his command, would support the idea excited researchers and database specialists who are advocating a revolution in scientific publishing (perhaps they are not content with the pace of a peaceful revolution). The NIH proposal for an e-print repository would be modeled after LANL's e-print archive. On April 22, NIH issued a plan that would be implemented by the **National Library of Medicine**.

A significant addition in their model, however, is that it would include a means to conduct peer reviews. Peer review has been the sacred “ace” that journal publishers have held. Thus, if an e-print service is able to include peer review, it will be another very significant development that will again shift the digital publishing paradigm.

The [Scholarly Publishing and Academic Resources Coalition \(SPARC\)](#), affiliated with the Association of Research Libraries, is also supporting web-based publishing of research. The aim is to develop electronic repositories where research can be posted, seen, and discussed by scientists worldwide, without the wait and cost of journal publications. The first partner with SPARC was the American Chemical Society -- the largest scientific society -- to publish an organic chemistry research.

Objections/obstacles echoed for any proposal have been the same:

- Scientific societies rely on their journals for income
- Role of Government questioned in publishing ‘journals’
- Journal publishers feel revenue-threatened
- Measure, or lack thereof, of prestige for publishing e-prints

Speaking on the NIH proposal earlier this spring, a Director within the [Federation of American Societies for Experimental Biology](#) was quoted as seeing a risk of “destroying the scholarly journal system that has served science so well for centuries.” However, if the availability of the information serves the interests of scientists and the advancement of scientific discovery, then access will be created, because the technology will allow it, regardless of the bureaucratic or others barriers. NIH has responded to the need and interest of scientists.

The Future of Preprints

The impact of preprints on scientific and technical information publishing should be of interest and concern not only to scientists but to the information professional, like us.

Some common threads should be noted. Preprint servers are:

- Primarily for STI, in most cases allowing online access to full-text research papers in specific disciplines.
- Most commonly hosted at universities, government labs, or - to a lesser degree - by professional societies.
- Well established in physics, astronomy, and mathematics. Chemistry and biology are now beginning to appear.
- Modeled for DOE STI at DOE labs LANL, SLAC, and Fermilab.

Electronic texts enable not just fast access but also new features and enhancements not possible before and likely not envisioned by the those of us in STIP.

Many are convinced that Internet publishing of scientific research papers is an inevitable new paradigm. I think we have to agree, based on our observation of the swift and significant impacts

already of the digital revolution. The line is also becoming very fuzzy between preprints and e-journals and how the two might eventually converge.

With the technology at hand, the process and the manner in which scientists publish and are served by publishers will continue to evolve. Is it not reasonable to anticipate - envision or even plan - preprint servers be established for each of the scientific disciplines of interest to DOE? We already cover Physics, Astronomy, Mathematics and Computing. Biology and Medicine are now in process. Other opportunities remain for areas such as Fuels and Energy Research; Chemistry; Materials Science; and Environmental Science.

In sum, whatever the scientific community decides it needs will shape the nature of electronic access to research papers. We in STIP are part of that scientific community and ought to be prepared for it - not just responding and reacting, but, as information professionals, also to anticipate and assist in the process. As Dr. Warnick mentioned, we at OSTI do not see any immediate prospects for getting involved in preprint servers, as we are fully preoccupied conquering other forms of text. Although DOE laboratories are involved, at least in some cases, the effort is little connected to the STI community. STIP may have the opportunity to get involved. If the openness of online publication appears to be helping scientists better understand what they and the public really need, it seems like a good thing for us in the STIP community to take part in some way.

Bibliography

Delamothe, T. "Electronic Preprints - What Should the BMJ Do - Clear Labeling Might Be the Answer," *British Medical Journal*, March 14, 1998.

Elliott, Sir Roger, "The Opportunity and Challenge of New Communications Technologies," *Nature*, World Conference on Science, January 1999.

Harnad, Stevan, "Learned Inquiry and the Net: The Role of Peer Review, Peer Commentary and Copyright," *Beyond Print*, 1997.

Harnad, Stevan, "The Invisible Hand of Peer Review," *Nature*, Web Matters, Nov. 5, 1998.

Kreitz, Patricia, "RE: Impact of Preprint Servers," e-mail to the DOE Library Operations Working Group, April 2, 1999.

Kreitz, Patricia, "The Virtual Library in Action: Collaborative International Control of High-Energy Physics Pre-prints," SLAC-PUB-7110, Feb. 1996.

Marshall, Eliot, "NIH Weighs Bold Plan for Online Preprint Publishing," *Science*, Mar. 12, 1999.

McConnell, J., "Having Electronic Preprints Is Logical," *British Medical Journal*, June 20, 1998.

Taubes, Gary, "Science Journals Go Wired," *Science*, Feb. 9, 1996.

Lewin, David I., "Online Science Journals: A Net Gain?" Columbia University, 21st C, Issue 3.4 (<http://www.columbia.edu/cu/21stC/issue-3.4/lewin.html>).