

ELECTRONIC PREPRINTS

HARRY E. PAYNE

Space Telescope Science Institute

3700 San Martin Drive, Baltimore, MD 21218, USA

Abstract. Paper preprints are in decline in astronomy, while the use of electronic preprints is on the rise, mainly *via* the LANL *astro-ph* preprint archive. I discuss the decline of the paper preprint, some preprint servers available on the Web and some general electronic preprint issues, the *astro-ph* archive and its response to these issues, and alternatives for the future of electronic preprints.

1. What are Preprints For?—Moving from Paper to Electronic

Preprints are an integral part of the astronomical research culture, serving two important purposes: as rapid communications of results, and as advertisements for the authors and their institutions. The interval of months or years between having a paper accepted for publication and having the printed journal reach subscribers could be dramatically shortened—in the electronic era, however, that interval itself should nearly be eliminated. Authors want to be sure that the senior astronomers who will influence tenure decisions are aware of their work, and the institutions want to be seen as scientifically productive.

Paper preprints have an impact that cannot be matched by asking someone to download and print an electronic preprint. And, there is some suspicion that senior astronomers do not download preprints. On the other hand, reproducing and distributing paper preprints has an identifiable cost, is slower, and greatly reduces the chance of discovery by someone whose name or library is not on the distribution list. So, while paper preprints are an effective way of reaching the intended audience, they exclude other audiences. In addition to impact *vs.* cost, the forces changing the balance between paper and electronic preprints are the move to electronic journals and the success of the Los Alamos National Laboratory (LANL) preprint archive,¹ claimed by its creators to be a desirable replacement for the refereed literature.

Paper preprints are in decline. In Figure 1a, I have plotted the rate at which paper preprints were added to two independent (but largely overlapping) lists: my institute's STEPsheet² (Space Telescope Exhibited Preprints) and the National Radio Astronomy Observatory's RAPsheet³ (Radio Astronomy Preprints) lists, maintained by the librarians at those institutions. I have also shown the total number of articles published in the ApJ, as determined by querying the NASA Astrophysics Data System⁴ (ADS). In 1996, ApJ articles made up more than 40% of the total number of articles published in ApJ, A&A, AJ, MNRAS, and PASP (Abt 1997), i.e., the main astronomical journals. The curves track each other extremely well, except since about 1994, when the article count is rising slowly and the preprint rate is declining sharply. A number of paper preprint series have disappeared—some were announced, others just stopped arriving. My own institution has twice discussed eliminating paper preprints, motivated by the potential distribution cost reduction.

The LANL *astro-ph* archive is taking up the slack, and more, caused by the decline of paper preprints, as shown in Figure 1b. The *astro-ph* submission rate has doubled in the last year and a half, after doubling during the year and a half before that. Cross-postings and updates are not

¹<http://xxx.lanl.gov/>

²<http://sesame.stsci.edu/lib/stsci-preprint-db.html>

³<http://www.nrao.edu/aolib/rapsheet.html>

⁴<http://adswww.harvard.edu/>

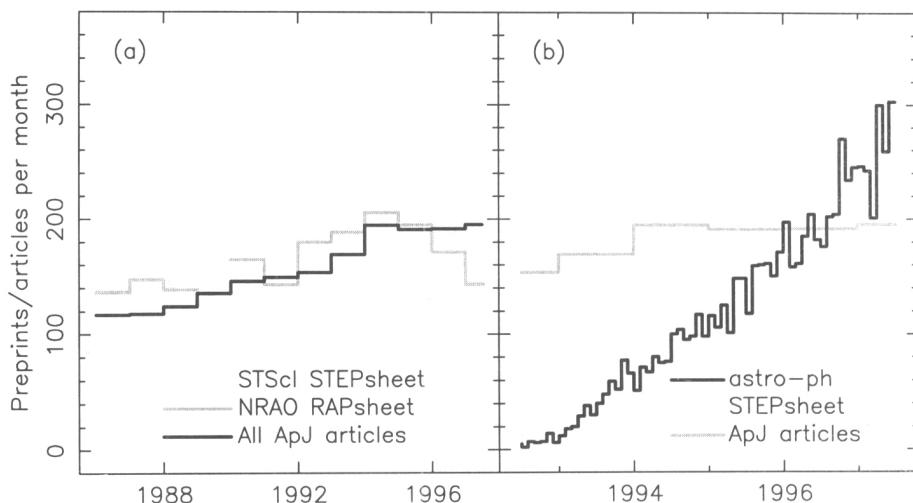


Figure 1. (a) A comparison of the rates at which paper preprints are added to STScI and NRAO preprint lists with publication rates in all sections of the ApJ. (b) The rate of LANL *astro-ph* submissions, compared to the STScI and ApJ data shown in (a).

counted in these numbers. At the moment, the area under this curve represents a small fraction of the literature—negligible compared to the ADS. But if the rate of growth continues, then most new journal articles will be found in the *astro-ph* archive.

2. Electronic Preprint Servers: What is Available Now?

Putting aside the exchange of preprints via e-mail, and focusing on preprint services available on the World Wide Web, we find that the field is dominated by servers or archives in various disciplines of physics and mathematics, with a sprinkling of computer science, biology, philosophy, and, of course, astronomy services. Prominent physics servers include the Los Alamos National Laboratory (LANL) preprint archive, the Stanford SPIRES high-energy physics database,⁵ maintained by SLAC librarians and containing over 345,000 entries, and the CERN preprint server.⁶ The American Physical Society preprint server⁷ provides an experimental platform for simplifying the communications between the author, referees, and the editorial office of the journals. The American Mathematical Society preprint server⁸ collects only bibliographic information, including an abstract, at their site, but authors are encouraged to provide URLs linking to full text versions of their preprints. Interesting features of the AMS server include asking authors to classify their contribution into one of nearly 100 standard subject classifications (e.g., *Number theory* or *Differential geometry*), and explicitly expiring links after two years.

In astronomy there is the *astro-ph* archive at LANL, a number of institutional, departmental, or group servers, and little in between. Of the 90+ preprint services listed by AstroWeb,⁹ most refer to departmental pages that actually have no links to preprints or no recent preprints. Exceptions include servers at the Australia Telescope National Facility¹⁰ and the McDonald Observatory Périquier Library.¹¹ The Princeton radio pulsar group maintains a directory of pulsar-related preprints¹²—

⁵<http://www-spires.slac.stanford.edu/find/hep>

⁶<http://preprints.cern.ch/>

⁷<http://publish.aps.org/eprint/docs/>

⁸<http://e-math.ams.org/preprints/preprints-home.html>

⁹<http://www.stsci.edu/astroweb/cat-preprint.html>

¹⁰<http://www.atnf.atnf.csiro.au/Preprints/>

¹¹<http://abstract.as.utexas.edu/abstracts.html>

¹²<http://pulsar.princeton.edu/rpr.shtml>

authors are asked to submit links to their own preprints. The Astrophysics Data System does not do preprints, other than to index the **astro-ph** archive.

Among the technically more interesting services is NCSTRL¹³ (pronounced “ancestral”), the Networked Computer Science Technical Reports Library, which consists of “a distributed network of interoperating servers.” Collections of computer science reports maintained independently by the participating institutions can be searched as if they were a single collection. Bibliographic records describing the items in each collection are exchanged between participating sites, so that each may host its own search engine. “NCSTRL-lite” sites consist of a simple ftp repository, with bibliographic records held at a central indexing site. This service is making use of important developments in the digital library field, and provides an interesting approach to distributed document collections of any type.

3. Electronic Preprint Issues

The purpose of the scientific literature is communication, and works that are published in electronic form but cannot be found do not advance that communication. Discoverability is everything. But as a paradigm for the future of publishing, having authors put up contributions on their personal, departmental, or institutional Web pages and waiting for readers to come to them is not working, because of the lack of discoverability. Using a global search engine, like AltaVista,¹⁴ turns up so much junk that nothing discovered this way has a scholarly aura, and no search engine targeted at distributed scholarly works exists in astronomy (or any other discipline). A distributed system with participation by institutions rather than individuals, like NCSTRL,¹⁵ would probably result in a better maintained, more homogeneous set of preprints, simply because of the extra attention they would get from their local librarian or site maintainer.

Preprints often change in wholesale ways, not amenable to corrigenda the way published papers are, and we expect them to go through several versions up to the published paper, which we then expect to be fixed. But it can be almost impossible to tell the lineage of an electronic document. And there is no system connecting preprints and papers. The librarians at my institution spend an hour or two with each issue of the major astronomical journals, trying to match the table of contents with their paper preprint list. The matching must be able to deal with re-ordered or different author lists, changed titles, and papers appearing in journals other than the one to which the preprint claims to have been submitted.

Authors relinquish their copyright when they publish, and they are entitled to keep their privately published electronic preprint only if the publisher does not object. In astronomy, the copyright holders have not yet decided how strenuously to object, but expiring electronic preprints does not appear to be common.

4. LANL’s astro-ph Archive

Astronomy has been fortunate in that timing and our relatively small size allowed winners to emerge early in the transition to electronic services. The Electronic ApJ¹⁶ is a model for other professional societies in how an electronic journal should work. The Astrophysics Data System¹⁷ is another treasure. And the LANL **astro-ph** preprint archive is the dominant source of pre-publication and unrefereed literature. The LANL archives are committed to the rapid exchange of information that is scholarly, but not refereed. They are very serious about being a preprint archive—preprints never expire. They are opposed to having the intellectual content retreat into the holdings of the journals, from which they can be retrieved only for a fee. Authors are encouraged to submit the final, published version of the paper, with an entry in the bibliographic record that gives the journal reference, and leave it there forever. That is the “subversive” nature of the archive. They challenge the existence of journals and the refereeing process.

The **astro-ph** archive is having a major impact on astronomy. Over 3000 astronomers subscribe to the **astro-ph** notification service, which brings a daily e-mail message containing the titles and

¹³<http://www.ncstrl.org/>

¹⁴<http://altavista.digital.com/>

¹⁵<http://www.ncstrl.org/>

¹⁶<http://www.journals.uchicago.edu/ApJ/journal/>

¹⁷<http://adswwww.harvard.edu/>

abstracts of new submissions. The ApJ Letters started encouraging authors to submit accepted papers to *astro-ph* in 1996, and *astro-ph* contribution numbers have become a *de facto* bibliographic reference. In at least one case, the availability of preprints from the LANL archive has caused a library to abandon the collection and cataloging of preprints. And *astro-ph* is having an effect on the research process: a provocative *astro-ph* contribution can create a flurry of activity where problems are hashed out on an impressive timescale of days to weeks.

The dominance of *astro-ph* solves the discoverability problem, they have effectively solved the version tracking problem, with the major exception of tracking all the way to the published paper, and there is a philosophical difference over the copyright problem. The failure of the *astro-ph* archive to link up with the astronomical literature is its greatest shortcoming. You can still retrieve the text of the first *astro-ph* submission,¹⁸ from April 1992, even though a search through the ADS¹⁹ shows that it was published in August of 1992. The *astro-ph* preprint has no links under the "Refers to"²⁰ and "Cited by"²¹ buttons, whereas the ADS²² entry for the corresponding paper lists 24 references²³ and 62 citations.²⁴

It may be because astronomy is so far out in front in the move to electronic publishing that the LANL shortcomings are so apparent. And that is a shame, since we have chosen to support a service which directly threatens our world-class electronic journals.

5. The Future

There are a couple of ways that the relationship between electronic preprints and the refereed literature could go. One is for electronic preprints to actually replace the journals. It would be straightforward to enhance the LANL server, say, to provide a mechanism for readers to record their comments as a running commentary, as a substitute for anonymous refereeing. The problems of the missing links could be fixed by checking references and citation against the ADS, contributions could be broken down by category, and our professional organizations could step in with some guarantee of permanence.

Another path for electronic preprints is a closer integration with the journals. That is the goal of our NASA AISRP²⁵ grant-funded electronic preprints project,²⁶ headed by Bob Hanisch (STScI). We would like to track a preprint through the publication process, so that it could automatically be identified with the published paper. If an author were able to tell the journal that this manuscript corresponded to a particular preprint (only the author can know, for sure), then, when the paper appears, the journal could notify some tracking authority to update the bibliographic record for that preprint with a link into the electronic journal. The system is a name resolver for manuscripts, returning the URL of the preprint until the paper is published, and the URL of the electronic journal article afterwards. At that point the preprint becomes undiscoverable, and can be removed. To make a complete system, we would collect and index metadata for each preprint, as in the LANL and NCSTRL systems. We would not archive most preprints, but instead rely on a distributed network of departmental or institutional preprint servers, but with centralized metadata and search engine.

If the journals were willing to accept an electronic preprint as a submission, or to produce an electronic preprint from a submission, then the journals could match preprints and papers without input from authors, and simplify distributing copies to the referees. Following the NCSTRL model, it should also be possible to decentralize the metadata, search engine, and name resolver, allowing each site a rapid local search limited to its own collection, to enhance the local preprint server. Or perhaps these elements should be merged into the ADS. Proper attention to the choice of metadata elements should eventually be rewarded by greater interoperability with other on-line document resources, in other disciplines.

¹⁸<http://xxx.lanl.gov/abs/astro-ph/9204001>

¹⁹<http://adswwww.harvard.edu/>

²⁰<http://xxx.lanl.gov/refs/astro-ph/9204001>

²¹<http://xxx.lanl.gov/cits/astro-ph/9204001>

²²http://adsabs.harvard.edu/cgi-bin/bib_query?1992ApJ...395L..83N

²³http://adsabs.harvard.edu/cgi-bin/nph-ref_query?bibcode=1992ApJ...395L..83N&refs=REFERENCES

²⁴http://adsabs.harvard.edu/cgi-bin/nph-ref_query?bibcode=1992ApJ...395L..83N&refs=CITATIONS

²⁵<http://www.hq.nasa.gov/office/oss/aisr/aisrp.html>

²⁶<http://doright.stsci.edu/Epreps/>

Acknowledgements

I would like to thank Helmut Abt, Gene Avrett, Ellen Bouton, Andy Fruchter, Evan Owens, and Archie Warnock, for helpful conversations and contributions, and especially my STScI colleagues Sarah Stevens-Rayburn and Edwin Huizinga for both discussions and a critical reading of the manuscript.

References

Abt, H. A. 1997, "Is the Astronomical Literature Still Expanding Exponentially?" submitted to PASP

Comments

EICHHORN: Preprints probably cannot replace the journals unless they provide the same connectivity as the journals, data centers and abstract services provide. Since they do not seem to intend to do that, they will not have the influence on astronomy that the E-journals have.

HJELMING: The LANL pre-print service and "personal publication" by email both address the fact that most papers are not worth permanent publication. The institutional publishers, librarians, etc. do not address this. They (LANL, Web publication) fulfill a perceived need that current institutions of publishing do not yet address.

FALCLEE: You mentioned two reasons for sending out preprints. 1. Advertizing. The first reason may disappear soon, since journals publish more quickly every year. The second reason will, however become more important since the journals grow bigger each year. Preprints allow you to target a very specific audience. We have had a very successful experience with launching a subject specific newsletter on the Galactic Center (GCNEWS). This newsletter distributes preprints (abstracts and lists) to a small but well selected list of readers - the majority of active researchers working in that field. This allows quick and efficient communication between scientists, helping, not threatening the role of major journals.

PAYNE: This service would be enhanced by the preprint to paper tracking I discussed.

BOYCE: A comment on unsolicited public comments serving as peer review. It hasn't worked. You get comments from people who don't contribute much to the field and the knowledgeable people don't have the time to read everything, much less comment.

As for Galactic Center and Be Star newsletters - this is one way a system "overlaid" on the current journal/preprint system can help sift out the relevant articles from the overwhelming mass of information available electronically.