

Citation Patterns to Electronic Preprints in the Astronomy and Astrophysics Literature

Gregory K. Youngen

Physics/Astronomy Librarian and Assistant Professor of Library Administration, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Abstract. The explosive growth in the number of citations to electronic preprints (eprints) in the literature of astronomy and astrophysics is documented in this paper. Eprints are the electronic versions of papers that have been submitted for comment and review among peers, for publication in journals, or prior to presentation at conferences. Internet-accessible eprint servers located worldwide provide unlimited free access to these publications long before they appear in print in journals or as conference proceedings. Because of the timeliness of these papers, as well as the increasing demand for current research, astronomers and physicists alike have found it necessary to cite these eprints in their research articles rather than wait until they appear in print.

This paper documents the increasing reliance on eprints in the journal literature and in conference proceedings by tracking the citations to eprints in articles and papers over the past five years. Along with identifying and documenting the trend, issues for concern over the growing number of citations to eprints, as well as areas for further study are discussed.

1. Why study preprints?

Scientists in physics and astronomy have been sharing the results of their research via preprints for many years. Since the Internet has improved the speed and efficiency of communication, preprints have become a much more common form of scientific information exchange. The results of this study indicate that electronic preprints in the fields of astronomy and astrophysics are becoming an increasingly important tool for the dissemination of primary research information.

The electronic preprint servers are often the first choice of physicists and astronomers for finding information on current research, news on breaking scientific discoveries, and keeping up with colleagues (and competitors) at other institutions. In addition to these benefits, electronic preprints allow the free and unrestricted access to scientific information without concern for international or institutional barriers. In pursuit of pure science this is considered a good thing. How this improved communication medium interacts with research establish-

ments, commercial and not-for-profit scientific publishers, and the researchers who write the articles is still in the process of working itself out.

2. Definition of preprints

There are several different definitions for the term "preprint". In a recent article, David Lim (1996) defines preprints as manuscripts which fall into one or more of the three categories listed below:

- manuscripts that have been reviewed and accepted for publication
- manuscripts that have been submitted for publication but for which a decision to publish has not been made
- manuscripts that are intended for publication but are being circulated among peers for comment prior to being submitted for publication.

Electronic preprints can fall into any and all of these categories. However, most of the eprint manuscripts posted to the servers are fairly complete reports ready for submission to publications and/or conferences. Authors cite traditional (paper) preprints in a variety of ways, depending where the preprint is in the publication cycle and the editorial guidelines of the journal publishing the article. If preprints have been submitted, but not accepted, citations usually refer to it as "submitted to..." If the manuscript has been accepted for publication, citations usually use "in press..." Problems encountered with the identification of an electronic preprint's status are not usually so ambiguous, thanks the eprint number, described later. The citations to preprints in this study were identified by using Institute for Scientific Information's (ISI) *SciSearch* bibliographic database available through Knight Ridder Information Service, DIALOG and the Stanford Linear Accelerator (SLAC) SPIRES database.

3. History of preprints in physics and astronomy

Close to 12,000 preprints are issued annually (Dallman 1994). Until recently, most of these preprints have been issued and distributed in paper by individuals or their institutions via mailing lists or upon request to the authors. High-energy physicists and astronomers have been at the forefront of using the preprint as a rapid communication medium due to the timeliness of their research and relatively closed groups in which they communicate research results.

Physics and astronomy librarians have also been on the cutting edge of preprint management and control by establishing sophisticated in-house databases to manage bibliographic records to the preprint literature. Bouton and Stevens-Rayburn (1995) describe two of the more comprehensive astronomy preprint databases and the impact of electronic preprints on traditional library service. Kreitz (1996) describes the Stanford Linear Accelerator Laboratory database of high-energy physics articles and preprints called SPIRES. The databases described in these articles track manuscripts submitted as preprints, then modify the bibliographic record when the preprints are formally published.

4. Development of the preprint server at Los Alamos

While traditional bibliographic databases were developed to provide access and accountability of paper-based preprints, the advent of electronic preprints provided the opportunity for development of full-text access to the papers, not just their bibliographic citations. The Los Alamos National Laboratory preprint server was founded by Paul Ginsparg in 1991 and is described in his 1994 article (Ginsparg 1994). The *eprint archives* were originally established to keep a small community of high energy physicists up to date on one another's research. The archive has grown in scope and in use since that time to include many other areas of physics, astrophysics, and mathematics. The LANL site is mirrored in several other countries throughout the world to provide improved access internationally.

5. Significance of the eprint number

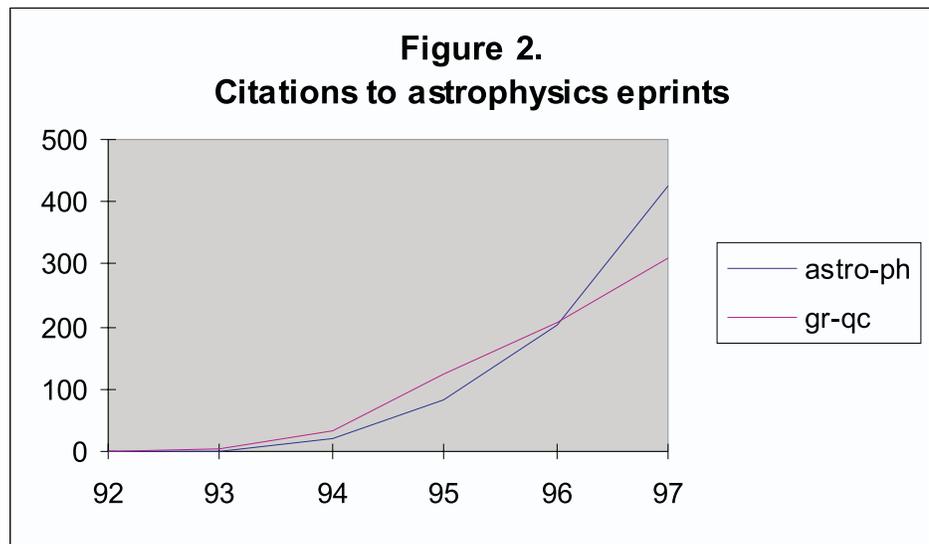
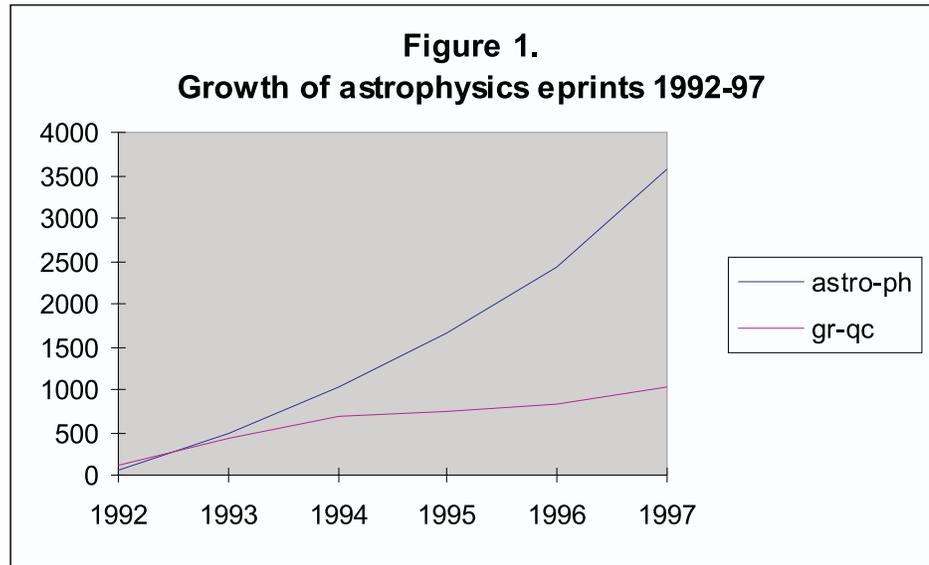
Most preprints are issued with a preprint number assigned by the author's host institutions. This number identifies the paper within the institution and distinguishes it from preprints issued by other institutions. The preprint numbers are not standardized, so it is difficult to group and sort them in a database.

The eprint number assigned by xxx.lanl.gov (the LANL preprint server) provides a standardized common number for preprints that allows the item to be uniquely identified regardless of the institution from which it originated. The eprint number is also useful for citing the work, as well as serving as a common link between databases consisting of bibliographic information and the full text of the article.

LANL's alphanumeric code provides broad subject categorization, year indicator, and accession number. The eprint number is a useful form of identification and serves as a linking point for electronic publications. The SLAC SPIRES database and the Astrophysical Data System (ADS) at Harvard use the eprint number to link their bibliographic (database) records to the full text electronic versions at LANL. Eventually, links could be established using the eprint number (or some other mutually agreed identifier) to track an article throughout its publication process, from inception to final publication, to reuse of the work's data in future publications.

6. Analysis of data and summary of implications

Despite the inherent problems with accurately identifying the total number of preprints, and to a lesser extent eprints, certain trends can be identified from the data collected. As can be seen in Figure 1, the number of astrophysical (astro-ph) eprints has been doubling every year since their introduction in 1992. The number of General Relativity / Quantum Cosmology (gr-qc) eprints are growing as well, but at a slower rate. This would lead one to believe that eprints are becoming more accepted within certain sectors of the physics and astronomy community of researchers as a means of quick dissemination of information. This would also indicate that scientists working in these subject areas are making the transition to electronic publications.



Another indication of the growing acceptance and importance of eprints is reflected in Figure 2. This graph shows the number of citations to astrophysical and gr-qc eprints in the printed literature continues to increase at an exponential rate. The growth rate in citations reflects not only the authors' acceptance of the eprint, but the publishers and editors of the manuscripts as well.

In addition to collecting data on the number of eprints published and documenting citation rates, information on the journals publishing articles that most often cite eprints was also collected. Results for astro-ph are reported in Table 1 and gr-qc eprint in Table 2.

Table 1. Top 10 Journals citing astro-ph eprints

1.	Astrophysical Journal	206
2.	Physical Review D	159
3.	MNRAS	68
4.	Physical Review Letters	55
5.	Physics Letters B	54
6.	Astronomy & Astrophysics	53
7.	Nuclear Physics B	37
8.	Astronomical Journal	20
9.	New Astronomy	13
10.	Astroparticle Physics	12

Table 2. Top 10 Journals citing gr-qc eprints

1.	Physical Review D	159
2.	Classical & Quantum Gravity	129
3.	Nuclear Physics B	65
4.	Physics Letters B	54
5.	Physical Review Letters	35
6.	Modern Physics Letters	26
7.	Journal of Mathematical Physics	21
8.	Physics Letters A	16
9.	General Relativity & Gravitation	14
10.	Int. J. Modern Physics D	11

Not surprisingly, eprints are being cited in the most important and influential journals in their respective fields.

Do highly cited eprints eventually become highly cited journal articles? It's really too soon to tell definitively. Table 3 lists five (relatively) highly cited eprints. While three of the articles have not been in print long enough to be cited, two have been. Early results indicate, at least anecdotally, that these documents will continue to be frequently cited as articles. This fact should be remembered by librarians and others using the citation databases for assessing an article's impact or importance in the field. Citations to published articles could be considerably higher if the citations to the articles' preprint are also factored in. The search results were also checked for the occurrence of self-citation. Only one self-cite to the Alcock article was noted.

Table 3. Tracking several highly-cited astro-ph eprints

Eprint	eprint cites	article cites	Current status
9601063 (Gorski)	7	0	Not yet published
9606080 (Jedamzik)	6	0	PRD(57)1998
9610162 (Kunic)	6	9	ApJ(482)1997
9606165 (Alcock)	6	16	ApJ(486)1997
9702100 (Bond)	7	0	MNRAS(11/97)

7. Types of papers citing eprints

It is assumed that reports, conference papers and other eprints will cite eprints more often than published, peer reviewed journal articles. Since conference papers, brief communications and technical reports are written for timeliness and are subject to a less-thorough review process, eprints would more likely be cited in these formats. This assumption is confirmed with a comparison of citing documents in the SLAC/SPIRES database and ISI's *Scisearch* database.

Scisearch, which only covers journal literature, lists 738 published articles citing astro-ph and gr-qc eprints (1992-1997). SLAC/SPIRES on the other hand, which includes many conference papers, and excludes articles once they are published, reports 4,190 documents citing astro-ph and gr-qc eprints for the same time period. It should be noted that the SPIRES database covers a much smaller percentage of the total astrophysics literature than Scisearch, but contains a much larger number of documents citing eprints.

8. Areas of concern

As evidenced by the data reported above, eprints have entrenched themselves in the literature of astronomy and astrophysics. The very nature of eprints, that they are somewhere between informal and formal publication, makes them difficult to classify. Managing the documents themselves has been accomplished quite admirably by LANL, SLAC/SPIRES and the ADS. Questions remain about the role of eprints in the process of scientific communication and how much effort should librarians and publishers expend to incorporate eprints into the mainstream publication and literature searching routines?

Some issues needing to be addressed include:

- *Including Eprints in A&I Services*
Since eprints are so easily accessible and relatively cost free, should abstracting and indexing services start including them in their coverage? Once eprints are formally published, it would be simply a matter of updating the database record to include place of publication and even a link to the full text.
- *Connecting Ejournal Citations with Eprints*
Several publishers, mainly professional societies, have started including links in their electronic journal articles from eprint citations to the full text at the LANL server. Will commercial publishers follow suit? Once the cited eprint is formally published, will the citation in the original citing article change? Should it?
- *Guidelines for Withdrawal and Revision of Eprints*
Eprints are by nature, documents submitted for review. They often change before the final version or formal publication. How should these changes be documented and tracked? If a journal article cites an eprint and that eprint changes dramatically before its final publication, what provisions can be made to ensure those changes are reflected in the citing article? Journals often publish errata for errors in articles. The electronic journal

and eprints have the capability to fix errors without having to go through the errata process. How should changes and errata be documented in the electronic publication?

- *Maintaining Integrity of the Eprint Servers*

The major eprint servers exist at sites dependent on government funding. Both LANL and SLAC are U.S. Department of Energy Laboratories. The future of the US DOE is frequently debated in Congress. If funding is pulled or transferred to another agency, Department of Defense, for example, how would that affect the preprint servers and the people who operate them? Would a commercial concern be willing to take over, and at what price? The fact that eprints are freely available now does not mean that they always will be.

- *Archival Issues*

How will citations to eprints in print journal articles be handled in the future, say 10, 25, or 100 years from now. Citations to print journal articles that old are easily found in libraries. Who will maintain the eprint files in the future? Will the electronic media be constantly upgraded from pdf and postscript to the new and enhanced formats of tomorrow? Who will be responsible for the cost and quality of the conversion?

9. Conclusion

The impact of electronic preprints on the future of scientific and technical publishing should be of interest and concern to scientists, publishers, and information professionals alike. Today scientists from around the world have free access to the most current research findings and reports weeks and months before the final products end up in print or presented at a conference. Several publishers have responded with their own initiatives to produce electronic preprints of articles to be published in their journals before they are sent to press (Taubes 1996). However, often times, these services are for subscribers to their print journals only. Other journal publishers have refused to accept manuscripts if they have appeared on the Internet (Hamilton & Dawley 1995).

The scientists publishing in the fields of physics, astronomy, and mathematics have a long history of sharing preprints among their peers. This tradition has laid the groundwork for the sharing of that same information in a new and improved format. Whether the rest of the scientific community, as well as scholars in the social sciences and humanities, adopt these practices remains to be seen. The movement toward electronic journals is already well underway. The time may be approaching for another paradigm shift from the traditional paper-based format to the complete electronic storage and retrieval of scientific reporting. Librarians, publishers, and the scientists themselves all have a stake in the outcome of this evolutionary shift. Laying the groundwork for a smooth transition will help everyone cope with the changes that are inevitable.

References

- Bouton, E. N. & Stevens-Rayburn, S. 1995, The preprint perplex in an electronic age, *Vistas in Astronomy*, 39, 149
- Dallman, D., Draper M. & Schwartz, S. 1994, Electronic pre-publishing for worldwide access: the case for high energy physics, *Interlending and document supply*, 22(2), 3
- Ginsparg, P. 1994, First steps toward electronic research communication *Computers in Physics*, 8 (4), 390
- Hamilton, J, & Dawley, H. 1995, Darwinism and the internet: why scientific journals could go the way of the pterodactyl, *Business Week* 26 June, 44
- Kreitz, P.A. 1996, The virtual library in action: collaborative international control of high-energy physics preprints. in: *GL'95. Proceedings of the Second International Conference on Grey Literature*, Amsterdam, Washington, DC., 2-3 November 1995, D. J. Farace (ed.), Amsterdam: Transatlantic, 33
- Lim, D. 1996, Preprint servers: a new model for scholarly publishing? *Australian Academic and Research Libraries (AARL)*, 27 (1) (March 1996), 21-30
- Taubes, G. 1996, APS starts electronic preprint service. *Science*, 273 (19 July 1996), 304