

Citation Patterns to Traditional and Electronic Preprints in the Published Literature

Gregory K. Youngen

The number of physics and astronomy preprints available electronically has increased dramatically over the past five years. Internet-accessible preprint servers at the Stanford Linear Accelerator Laboratory (SLAC), Los Alamos National Laboratory (LANL), and elsewhere provide unrestricted access to citations and/or full text of many physics and astronomy papers long before they appear in print. Because of the timeliness of these papers, as well as the increasing demand for current research, physicists and astronomers have found it necessary to cite these preprints in their research articles rather than wait until they appear in print. This paper identifies the growing importance of electronic preprints in the published literature and addresses several areas of concern regarding the future role of electronic preprints in scientific communication.



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

The electronic preprint has become the first choice among some physicists and astronomers for finding information on current research, breaking scientific discoveries, and keeping up with colleagues

(and competitors) at other institutions. When mounted on servers connected to the Internet, preprints also allow free and unrestricted access to scientific information without concern for international or institutional barriers. In the pursuit of pure science, this is considered a good thing. However, the effects of the electronic preprint on research establishments, commercial and not-for-profit scientific publishers, and the researchers who write the articles are the subject of much controversy and skepticism with regard to this medium's impact on the future of scientific communication.

Definition of Preprints

There are several different definitions for the term *preprint*. In his 1996 article, David Lim

Gregory K. Youngen is an Assistant Professor in Library Administration and Physics/Astronomy Librarian at the University of Illinois at Urbana-Champaign; e-mail: youngen@uiuc.edu.

defines preprints as manuscripts that fall into one or more of the categories below:¹

- manuscripts that have been reviewed and accepted for publication;
- manuscripts that have been submitted for publication but for which the 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.

For the purpose of this study, items falling into the third category are the most important to identify. A manuscript cited as a "preprint" is most likely to be the earliest publicly available version of a work and thus the one to provide the most up-

Most traditional paper preprints are issued with a preprint number assigned by their host institution.

to-date report on the actual research.

Authors cite preprints in a variety of ways, depending on where the preprint is in the publication cycle and the editorial guidelines of the journal publishing the article. If a preprint has been submitted, but not accepted, citations usually refer to it as "submitted to" If the manuscript has been accepted for publication, citations usually refer to it as "in press." Problems encountered with the identification of a preprint's status are described later. The citations to preprints in this report were identified by using Institute for Scientific Information's (ISI's) SciSearch bibliographic database available through Knight Ridder Information Service, DIALOG.

History of Preprints in Physics and Astronomy

Almost 12,000 preprints are issued annually.² Until recently, most of these have been issued and distributed in paper by individuals or their institutions via mailing lists or upon request. High-energy physicists and astronomers have been at the forefront of using the preprint as a rapid communication tool due to the

timeliness of their research and relatively closed groups in which they communicate research results.

Physics and astronomy librarians also have been at the forefront of preprint management and control by establishing sophisticated in-house databases to manage bibliographic records to the preprint literature. E. N. Bouton and S. Stevens-Rayburn described two of the more comprehensive astronomy preprint databases and the impact of electronic preprints on traditional library service.³ Pat Kreitz described the Stanford Linear Accelerator Laboratory (SLAC) database of high-energy physics articles and preprints called SPIRES.⁴ The databases described in these articles track manuscripts submitted as preprints and then modify the bibliographic record when the preprints are published formally.

Development of the Preprint Server at Los Alamos

Whereas 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 (LANL) preprint server (<http://xxx.lanl.gov>) was founded by Paul Ginsparg in 1991 and is described in his 1994 article.⁵ Originally, the e-print archive was established to keep a small community of high-energy physicists up to date on one another's research. However, since then, it has grown in scope and use 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.

Significance of the E-Print Number

Most traditional paper preprints are issued with a preprint number assigned by their host institution. This number identifies the paper within the institution and

FIGURE 1
ISI Scisearch Record: Citation to Preprint

15625964	Genuine Article#: WV537	Number of References: 12
Title: Competing magnetic interactions in manganese perovskites		
Author(s): Fontcuberta J ; GarciaMunoz JL; Suaaidi M; Martinez B; Pinol S; Obradors X		
Journal: JOURNAL OF APPLIED PHYSICS, 1997, V81, N8,2B (APR 15), P5481-5483		
Cited References:		
DETERESA JM, 1996, V76, P3392, PHYS REV LETT		
FONTCUBERTA J, 1996, V34, P379, EUROPHYS LETT		
GOODENOUGH JB, 1967, V164, P785, PHYS REV		
HWANG HY, 1995, V75, P914, PHYS REV LETT		
MILLIS AJ, PREPRINT		
TOMIOKA Y, 1995, V25, P5108, PHYS REV LETT		
WOLFMAN J, J SOLID STATE CHEM		

distinguishes it from preprints issued by other institutions. Because the preprint numbers are not standardized, it is difficult to group and sort them in a database.

The e-print number assigned by the LANL e-print archive provides a standardized common number for preprints that allows each item to be uniquely identified regardless of its institution of origin. Moreover, the e-print number is useful for citing the work, as well as for serving as a common link between databases consisting of bibliographic information and full text.

LANL has established the following subject groupings and numbering system for incoming e-prints:

Astrophysics

astro-ph/9701001

Condensed Matter

cond-mat/9701001

General Relativity and Quantum Cosmology

gr-qc/9701001

High-Energy Physics—Experimental

hep-ex/9701001

High-Energy Physics—Lattice

hep-lat/9701001

High-Energy Physics—Phenomenology

hep-ph/9701001

High-Energy Physics—Theory

hep-th/9701001

Nuclear—Experiment

nucl-ex/9701001

Nuclear—Theory

nucl-th/9701001

Physics—General

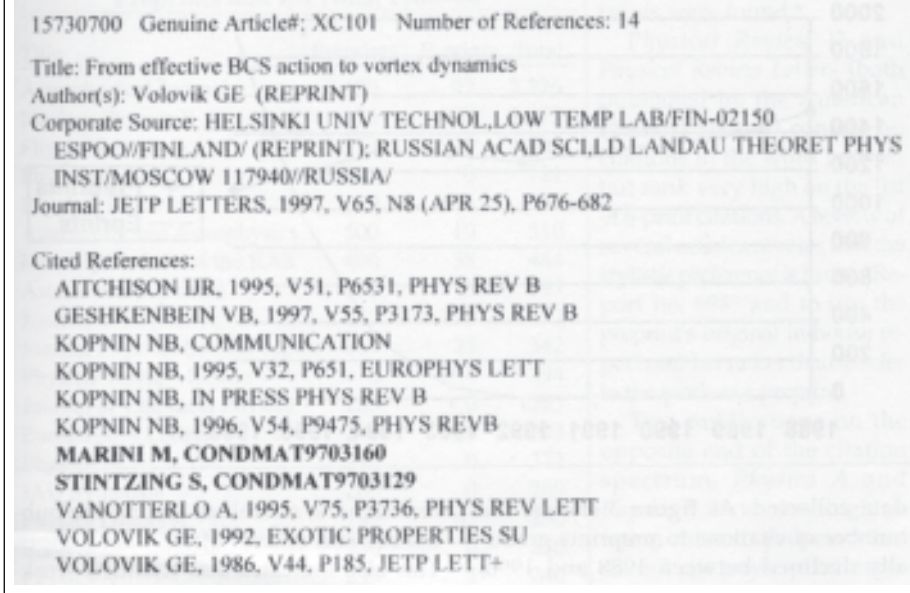
physics/9701001

Quantum Physics

quant-ph/9701001

LANL's alphanumeric code provides a broad subject categorization, year and month indicator, and an accession number. The e-print number is a useful form of identification and serves as a linking point for electronic publications. The SLAC SPIRES database and, more recently, the Astrophysical Data System (ADS) at Harvard use the e-print number to link their bibliographic (database) records to the full-text electronic version at LANL. Eventually, links could be established using the e-print number to track an article throughout its publication process, from inception to final publication to reuse of

FIGURE 2
Citations to E-prints in Scisearch



the work's data in future publications.

Difficulties and Inaccuracies in Preprint Citations

As mentioned previously, this study has used ISI's SciSearch database to identify journals that publish articles citing preprints and to obtain an overall number of citations to preprints within the past ten years. SciSearch is unique among commercial bibliographic databases in that its records contain not only bibliographic information, but also the citations used in the article.

The search phrase "Cited Work = Preprint" (cw = preprint) is used to identify citations to preprints. Many authors cite preprints in this fashion (see figure 1).

A problem identified early on was that this search strategy does not include citations to all works that technically could be considered preprints. For example, many authors cite works as: "Submitted to . . ."; "To be published in . . ."; "In press"; "In preparation"; or "Unpublished." However, citations containing

these words may or may not be referring to preprints. For the purpose of this study, only those citations specifically stating "preprint" have been analyzed. In reality, the number of citations to preprints may be much higher. However, using only the term *preprint* in the search nearly guarantees that an actual preprint is being referred to by the citing author.

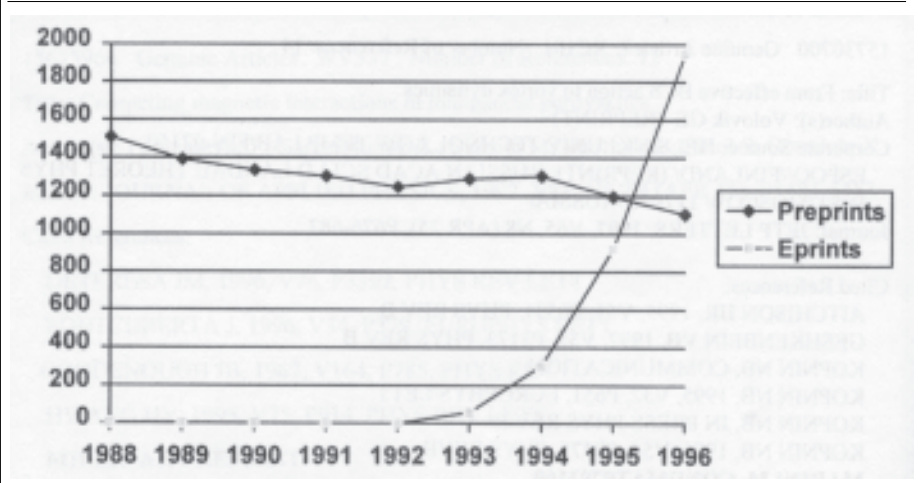
Citations to E-Prints

Citations to e-prints are more precise because their e-print numbers have become a de facto standard for identification (see figure 2). However, this is not true in all cases; some authors prefer to put the originating organization's preprint report number before the e-print number. Because SciSearch indexes only the first number, some e-prints go unreported in this study.

Analysis of Data and Summary of Implications

Despite the inherent problems with accurately identifying the total number of

FIGURE 3
Citations to Preprints and E-prints, 1988–1996



preprints (and to a lesser extent e-prints), certain trends can be identified from the data collected. As figure 3 shows, the number of citations to preprints gradually declined between 1988 and 1996, whereas the number of citations to e-prints nearly doubled every year since their introduction in 1992. Figure 4 represents the cumulative growth of citations to both preprints and e-prints between 1988 and 1996. The results seem to indicate that e-prints are becoming more ac-

cepted within certain sectors of the physics and astronomy community of researchers, as well as among the publishers and editors of the manuscripts. This also indicates that scientists working in subject areas that are more likely to use and cite traditional preprints are making the transition to electronic publications.

The journal titles with articles that most often cite preprints and e-prints are identified in table 1. An immediate observation is that journals tend to cite one type

FIGURE 4
Cumulative Citations to Preprints and E-prints, 1988–1996

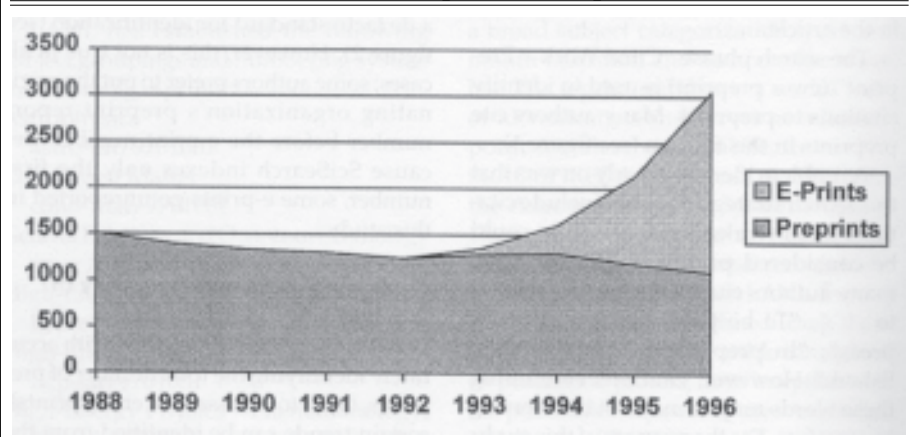


TABLE 1
Journals Ranked by Number of Citations to
Preprints and E-Prints, 1988–96

Title	Preprints	E-prints	Total
Astrophysical Journal	2,129	97	2,226
Nuclear Physics B	75	998	1,073
Physics Letters B	115	920	1,035
Physica C	721	0	721
Physical Review D	0	636	636
Astronomy and Astrophysics	500	10	510
Monthly Notices of the RAS	406	38	444
Astronomical Journal	392	5	397
Journal of Physics A	255	122	377
Journal Physical Society Japan	317	25	342
Physical Review Letters	0	294	294
Journal of Chemical Physics	285	0	285
Europhysics Letters	247	26	273
Physica B	272	0	272
IAU Symposia	260	0	260
Physics Letters A	195	65	260
Solid State Communications	250	0	250
Physica A	235	5	240
Journal Of Physics: Condensed Matter	223	4	227

over the other to a substantial degree. This may be because of editorial guidelines against citing preprints, preferring instead “submitted to,” “in press,” and so on. Another reason may be the subject scope of the journal. Clearly, the e-print phenomenon has been most strongly present in high-energy physics (especially particle physics) and astrophysics. This can be seen in the high e-print citation rates in *Nuclear Physics B*, *Physical Review D*, and *Physics Letters B*, all journals aimed at the high-energy physics and particle physics community. The other titles may not be publishing as many articles in these subject areas.

The instructions provided to authors by many of the journals and style guides of the publishers were reviewed (where available) to determine if there are any written guidelines on citations to preprints. With the exception of *Astrophysical Journal*, which states that “References to private communications, papers in

preparation, preprints, or other sources generally not available to readers should be avoided,” no firm rules for citing preprints were found.⁶

Physical Review D and *Physical Review Letters* (both published by the American Physical Society) contain no citations to the word *preprint* but rank very high on the list of e-print citations. A review of several articles reveals that the stylistic preference is to cite “Report no. ###” and to use the preprint’s original in-house report number rather than to refer to the work as a preprint.

Two publications on the opposite end of the citation spectrum, *Physica A* and *Physica B*, have very high citation rates to preprints but do not cite any e-print numbers. In private communication with the editors, they said there

is no rule against citing e-print numbers, but they just have not yet encountered citations to electronic preprints.

The very nature of e-prints—that they are somewhere between informal and formal publication—makes them difficult to classify.

Tables 2 and 3 show the number of citations in individual journals to preprints and e-prints, respectively. In both tables, journals publishing articles that frequently cite preprints and e-prints have remained consistent over the years.

Areas of Concern about the E-Print’s Role in Scientific Communication

As evidenced by the data reported here, e-prints have become an important part of the literature of physics and astronomy. The very nature of e-prints—that they are somewhere between informal and formal publication—makes them difficult to clas-

TABLE 2
Journals with articles That Most Often Cite "Preprints"

Citations to Preprints	1988	1989	1990	1991	1992	1993	1994	1995	1996	Total
Astrophysical Journal	253	231	228	236	230	247	220	238	195	2,078
Physica C	158	109	63	129	44	47	93	38	36	717
Astronomy and Astrophysics	53	48	49	50	54	59	73	47	53	486
Monthly Notices of the RAS	51	55	42	38	40	50	30	46	39	391
Astronomical Journal	36	44	41	40	49	43	52	44	30	379
Journal of the Physical Society Japan	38	43	34	25	24	39	36	33	41	313
Journal of Chemical Physics	46	38	41	27	20	22	34	26	22	276
Physica B	5	14	56	19	10	30	97	17	24	272
IAU Symposia	24	28	22	37	21	47	7	19	55	260
Journal of Physics A	31	30	28	45	16	33	31	22	19	255
Solid State Communications	72	43	31	23	24	23	8	13	8	245
Europhysics Letters	27	22	28	29	32	40	30	19	17	244
Physica A	6	17	40	28	35	35	23	25	17	226
Journal of Physics— Condensed Matter	33	34	21	18	15	19	30	23	24	217
Physics Letters A	27	19	34	19	24	25	16	15	12	191
Journal de Physique I	47	18	22	12	9	13	10	19	7	157
Zeitschrift fur Physik B	28	17	26	20	14	10	7	8	10	140
Nuclear Physics A	18	11	27	13	19	22	11	9	8	138
Publication of the Astronomical Society Pacific	25	11	14	14	17	13	10	12	8	124
Journal of Statistical Physics	25	17	15	22	9	10	9	8	0	115
Astrophysical Journal— Supplement Series	8	18	16	10	18	4	21	12	10	117
Physics Letters B	20	14	19	9	17	12	11	4	9	115
Surface Science	11	7	17	10	20	0	19	9	13	106
Journal of Low Temperature Physics	9	6	0	0	23	7	14	24	23	106
Macromolecules	5	6	8	9	11	16	19	13	12	99
All other titles	72	109	68	82	129	164	88	147	130	989
Total	1,128	1,009	990	964	924	1,030	999	890	822	8,756

sify. Managing the documents themselves has been accomplished quite admirably by LANL, SLAC/SPIRES, and the ADS. However, questions remain about the role of e-prints in the process of scientific communication and how much effort librarians and publishers should expend to incorporate e-prints into the mainstream publication and literature searching routine. Below are listed several areas of concern that need to be addressed in the near future as e-prints begin to play a more significant role in scientific communication:

- *Including e-prints in A&I services:* Because e-prints are easily accessible and relatively cost free, should abstracting and indexing services start including them in their coverage? After e-prints are published formally, 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 electronic journal citations with e-prints:* Several publishers, mainly professional societies, have started including links in their electronic journal

TABLE 3
Journals with Articles That Most Often Cite E-prints

E-prints	1993	1994	1995	1996	Total
Nuclear Physics B	13	61	215	430	719
Physics Letters B	21	81	200	421	723
Physical Review D	11	45	138	291	485
Physical Review Letters	3	17	61	151	232
Modern Physics Letters A	0	26	44	80	150
Classical and Quantum Gravity	1	14	39	52	106
Journal of Physics A	2	14	39	56	111
International Journal of Modern Physics A	0	11	19	54	84
Astrophysical Journal	0	4	20	54	78
Journal of Mathematical Physics	0	3	31	39	73
Nuclear Physics A	0	3	16	40	59
Physics Letters A	0	8	20	20	48
Progress of Theoretical Physics	0	0	12	36	48
Communications in Mathematical Physics	1	6	15	29	51
Physical Review C	0	0	14	25	39
Acta Physica Polonica	0	0	7	33	40
Monthly Notices of the RAS	0	0	7	22	29
Progress in Theoretical Physics—Supplement	0	0	6	23	29
Czech Journal of Physics	0	2	0	24	26
Europhysics Letters	0	0	7	14	21
Journal of the Korean Physical Society	1	3	6	16	26
Journal of the Physical Society Japan	0	3	0	17	20
Zeitschrift für Physik C	2	0	0	22	24
Theoretical and Mathematical Physics	2	4	6	0	12
Total E-Print citations	57	305	922	1,949	3,233

articles from e-print citations to the full text at the LANL server. Will commercial publishers follow suit? After the cited e-print is published formally, will the citation in the original citing article change? Should it?

- *Guidelines for withdrawal and revision of e-prints:* E-prints 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 e-print and that e-print changes dramatically before its final publication, what provisions can be made to

ensure that the changes are reflected in the citing article? Journals often publish errata for errors in articles. Electronic journals and e-prints have the capability to correct errors without having to go through the errata process. How should changes and errata be documented in the electronic publication?

- *Maintaining integrity of the e-print servers:* The major e-print servers exist at sites dependent on government funding. Although the xxx.lanl e-print server is funded largely through the National Science Foundation, both LANL and SLAC

are U.S. Department of Energy Laboratories. The future of the US DOE is debated in Congress frequently. If funding is pulled or transferred to another agency (the 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 e-prints are freely available now does not mean they always will be.

The time may be approaching for a more profound leap from the traditional paper-based format to the complete electronic storage and retrieval of scientific reporting.

- *Archival issues:* How will citations to e-prints in print journal articles be handled, say, 10, 25, or 100 years from now? Citations to print journal articles that old are found easily in libraries. Who will maintain the e-print 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?

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 are 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.⁷ However, often 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.⁸

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 under way. The time may be approaching for a more profound leap 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.

Notes

1. David Lim, "Preprint Servers: A New Model for Scholarly Publishing?" *Australian Academic and Research Libraries (AARL)* 27, no. 1 (1996): 21-30.

2. D. Dallman, M. Draper, and S. Schwartz, "Electronic Pre-Publishing for Worldwide Access: The Case for High Energy Physics," *Interlending and Document Supply* 22, no. 2 (1994): 3-7.

3. E. N. Bouton and S. Stevens-Rayburn, "The Preprint Perplex in an Electronic Age," *Vistas in Astronomy* 39 (1995): 149-54.

4. Pat Kreitz, "The Virtual Library in Action: Collaborative International Control of High-Energy Physics Preprints," in *Proceedings of the Second International Conference on Grey Literature (GL'95)*, ed. D. J. Farace (Amsterdam: Transatlantic, 1996), 33-41.

5. Paul Ginsparg, "First Steps toward Electronic Research Communication," *Computers in Physics* 8 (Jul./Aug. 1994): 390.

6. "The Astrophysical Journal Instructions to Authors," *Astrophysical Journal* 480 (May 1, 1997): ii.

7. Gary Taubes, "APS Starts Electronic Preprint Service," *Science* 273 (July 19, 1996): 304.

8. J. Hamilton and Heidi Dawley, "Darwinism and the Internet: Why Scientific Journals Could Go the Way of the Pterodactyl," *Business Week*. (June 26, 1995): 44.