SLAC-PUB-8566
physics/yymmnnn
August 2000

The Horse That Drank: Electronic Communication and the High Energy Physics Community

Heath B. O'Connell[1]
Stanford Linear Accelerator Center
Stanford CA 94309
hoc@slac.stanford.edu
http://www.slac.stanford.edu/grp/lib/people/hoconnell.html

High energy physicists submit their papers to electronic archives, have them published in electronic journals and indexed in online databases. Despite this system now having been in place basically in its entirety since the middle of the '90's, other fields have been slow to follow. What makes high energy physics special (and what doesn't)?

[1] Work supported by the US Department of Energy contract DE-AC03-76SF00515.

1 Introduction

Paul Ginsparg's eprint archive [http://arXiv.org] shines like a beacon in the field of electronic communication for academic research. It serves the high energy physics (HEP) community so well, and has attracted so much interest from those who study scholarly communication, that one might wonder why this model hasn't been adopted by all disciplines.

To attempt to answer this question, we have to look at how the eprint archive came to be, what were the key ingredients in its success, and whether or not any of those are peculiar to HEP.

2 Historical development

"Eprints" is actually a play on words based on "electronic preprints," and paper preprints, articles made public before their publication in a journal, are certainly nothing new in HEP. As early as the 1960's large institutions were receiving thousands of these unpublished works per year in the mail. Two of these institutions, the Stanford Linear Accelerator Center in California (SLAC) and the Deutsches Elektronen-Synchrotron (DESY) in Hamburg, began to keep track of them in a fully searchable bibliographic database called SPIRES, the Stanford Public Information REtrieval System [1]. By the end of the '60's, this database was being used to generate a weekly list of all the new HEP preprints. This list, named PPF (Preprints in Particles and Fields), was then sent to hundreds institutions around the world each week.

The key ingredient here was centralisation. You only had to send your paper to a single place to get it seen all over the world and by subscribing to a single list you could keep track the latest research in HEP. With the rise of computer networking the service was further enhanced. By 1985 physicists all over the world could search the SPIRES database directly.

3 xxx.lanl.gov

The SPIRES database and PPF list made finding a useful paper, very easy. To actually read it, though, you would have to contact the author and request a copy of the paper. In the mid '90's it was still common to find a "preprint request card" in your mailbox asking for a copy of your latest paper. However, as everyone had an email address, you could simply send the author an email to request a paper.

The second method had an immediate benefit. Most papers were written on a computer, using the typesetting language, TEX, which allowed complicated mathematical expressions to be precisely rendered from a simple plain text source (and hence proved to be enormously popular with both the HEP and mathematics communities). As the TEX source file was simply plain text, it could easily be sent through email, so you could request a paper from someone thousands of miles away and be reading it in half an hour.

Before long, it was common to send TEX files to your peers (as you might once have done with paper preprints before the PPF list) leading to overstuffed electronic mailboxes for some. In 1991, Paul Ginsparg of Los Alamos National Laboratory (LANL) decided to do something about this. If authors could send their papers electronically to a central repository, with author-supplied bibliographic data, a list could be sent out of each *day's* additions and the "eprints" could be obtained directly from this archive.

The appearance shortly thereafter of the World Wide Web certainly made using LANL and SPIRES much easier (and indeed SPIRES had the first web server in North America), but both systems can still be used without it. What Web technology did do, however, was bring the internet to the general public, and in doing so made electronic scientific communication all the more visible and gave birth to the online publication of journals.

4 So why HEP?

Why did the High Energy Physics community provide such a fertile ground for this particular aspect of the so-called "Information Revolution"? We have touched on the reasons throughout this talk, but it is useful, perhaps, to summarise them to understand how this model might be more widely adopted (or why it might *not* be).

As can be understood from the previous section, with a thirty year development it has not been a revolution so much as an evolution. Generally free of commercial or governmental restrictions (with some notable exceptions), high energy physics has always had a strong culture of communication, as we might also expect with, say, the mathematics community. Indeed, at this level there is not much to separate the theoretical physicist from the mathematician. Both adopted TEX with a passion, and used email to send papers, important in fields requiring the rapid dissemination of completed work. In this climate, SPIRES, an endeavour run for physicists at a physics lab, provided a role model for a centralised information store that was emulated in the creation of the eprint archive for Physics. The new system was swiftly adopted by the close knit HEP community (at first the theory people), with appropriate checks to ensure the quality of the work posted. Before long mathematics and condensed matter (another TeX-friendly field) archives were established and also proved very successful.

On the whole, HEP journals have had little trouble dealing with the existence of eprint archives. The imprimatur of publication in a refereed journal is still highly important, and some 70% of eprints are eventually published (with another 20% appearing in conference proceedings). The American Physical Society's flagship journal, *Physical Review*, has enthusiastically entered the new age. It allows authors to submit their papers through the LANL archive, and considers the online version of the published document to be definitive [2].

Not all related fields of HEP, however, have been as quick to use LANL. Although true to some extent for the experimental community, this is especially true of the accelerator physics (AP) community, the people who actually build the giant machines used in today's experiments. There are really three reasons why they *might* not use it. The first is technical. TeX is not as dominant in for the AP community, since they are less likely to need a lot of mathematical expressions and more likely to need colour photos of their equipment. This reason does not hold though, as LANL accepts many other formats for documents. While TeX is preferred, it is by no means terribly difficult to use something else. Neither can one really believe that, formats aside, posting to LANL is simply too difficult. The second explanation is more compelling: awareness. Many accelerator physicists simply do not know of its existence. This is not as surprising as it might first seem, given the relatively low level of interaction between theoretical and accelerator physicists. The third is one of culture. The members of the AP community have historically relied less on preprints to discuss their work, and more on conference proceedings. Thus there wasn't the simple modification of existing habits that took place when HEP switched to electronic publication.

Using a recent conference organised by SLAC for AP, we have been encouraging this community to make greater use of the LANL archives. The goal is for the proceedings of this conference to be created from files that the authors have sent to LANL — a first for AP. It so far appears to be going well.

5 Beyond Physics

Two of the key reasons why the LANL archive is so useful are centrality (like SPIRES before it) and automation (which allows so much to be done with so little effort). It is useful to keep these two features in mind in any discussion of "author self-archiving." The World Wide Web, which has made the internet and the whole idea of paperless publishing so accessible can actually work against a centralised system! It is so easy to make your own web page, and post your papers there, that this looks like a tempting alternative to the LANL model. This has two problems. Firstly, unless you are a leader in the field, your personal web page is probably not going to be the most visible place to post your work. Secondly, there is the question of the archival storage of your work.

A half-way point that has been suggested to encourage the participation in author self-archiving in other fields is institutional archives. That is, the University of Ballarat, say, would require its researchers to post their work to a server. While this would certainly solve the problem of archival storage, I don't think such a service would be terribly useful on its own, as it fails to meet the centrality requirement. It is more useful to be able to go to the place where all the work in a particular discipline has been stored, rather than the place where all the work from a particular institution is stored (at least for active researchers). However, this wouldn't be a bad start.

Therefore, to make use of a large collection of institutional servers we would require a new system. The Universal Preprint Service (UPS) would seek to provide an interface that would allow users to search across a wide range of different archives [3]. Such a system would invariably require some degree of standardisation between the archives, and thus needs to be approached in a cooperative manner [4].

The fluidity of this new, electronic means of publication also has implications for the traditional research process. The author is free to replace the eprint at any time with a revised version, but it is also of utmost importance to establish who was the first person to present an idea. The LANL archive, has neatly addressed this issue in an appropriate manner, by archiving each version of an E-print with a date stamp, documenting any modifications in the paper. In some cases the author might remove the eprint altogether upon discovery of some fatal flaw. Once the article is published in a journal, it is generally regarded a finished piece of work, with any subsequent alterations handled via errata. Other authors may *comment* on the published paper, to which the author may *reply*, but the comment and reply are treated as new articles, and the original is left unchanged. One journal, however, has broken away from the traditional model. *Living Reviews in Relativity* [5] exists solely on the Web, and allows authors to constantly revise their articles. In part this is due to the pedagogical nature of the journal; it seeks to provide reviews that aid learning, rather than publish original research.

With seemingly no technical hurdles nor any real challenge to scholarly traditions, what still might prevent other fields from adopting such a system? Stevan Harnad of the University of Southampton has examined this question in depth [6]. One particularly complicated case is the eprint archive PubMed Central (formerly E-biomed) proposed by the U.S. National Institute of Health. In a field without a tradition of preprints, this proposal does not seem to have been warmly welcomed by the established journals (unlike the major physics journals). One objection deals with a very important difference between medical science and physics: an unrefereed repository for medical literature could present a danger to the public. However, this can be handled by the division of such an archive into a refereed and unrefereed section.

What is the role of journals in this new world? Would an eprint archive render them superfluous? While one might expect it could, this does not seem to be the case. The great majority of eprints is destined for the refereed journals and librarians are usually rather circumspect about cancelling print subscriptions for reasons of permanent access. The stability of this situation is possibly illusory, and maintained merely by institutional inertia. The importance of refereeing though, both for the reader who wants to be sure that a paper has already been judged worthy, and the author who needs to readily demonstrate a certain level of professional ability, is not going to lessen in any foreseeable future. Because of this there is a role for the continued existence of research journals. However, I feel those that take advantage of electronic communication either for lowering costs or improving services will be the ones to flourish.

References

[1] H. B. O'Connell, "Physicists thriving with paperless publishing," physics/0007040.

[2] http://www.aps.org

[3] H. Van de Sompel and P. Hochstenbach, "Reference linking in a hybrid library environment," D-Lib Magazine **5** no. 4 (1999) and D-Lib Magazine **5** no. 10 (1999) see also D-Lib Magazine **6** no. 2 (2000) [http://www.dlib.org]

[4] http://www.openarchives.org

[5] http://www.livingreviews.org

[6] http://www.cogsci.soton.ac.uk/∼harnad