

published as the ‘Back Page’ essay of the APS News, April, 2005:

Publication and the Internet: Where Next?

MICHAEL E. PESKIN

A part of the vision of the future of science enunciated by Vannevar Bush in 1945 was the ‘memex’, a machine that instantly retrieved any paper in the scientific literature [1]. Today, that device is at my disposal. When students come to my office with queries that might be about any topic in high-energy physics, I can put the answers in their hands with a few clicks of the mouse. This is the result of enormous effort in the past decade by many people who shared this vision, from the developers of the World-Wide Web to the authors of Web-accessible archives and search engines. It is time now to consolidate what we have learned and to pose clearly the next set of problems to be addressed.

Electronic dissemination of scientific information can decrease the cost of publication. So it is ironic that, in parallel with these technological developments, we are troubled by spiraling costs for scientific journals. Indeed, this ‘serials crisis’ is the primary concern of university and technical librarians everywhere. Some part of this crisis would be alleviated by authors choosing to publish in cost-effective journals. In my field of physics, the cost/paper to libraries varies among journals by a factor of 10, with the *Physical Review* at the low end of the scale. Over the long run, however, the physics community should move to new modes of publication that are facilitated by the new technologies. In this article, I will explain where I think we should go.

There are, I believe, two important components to a new publishing model. The first is to rely as much as possible on the authors of scientific papers to take over functions now carried out by scientific journals. Good software tools can facilitate this. The second is to recognize tasks that—irreducibly—require professional editors and staff, and to assign the real costs and collect revenues for these tasks.

In the following, I will quote costs for these services as a fraction of the present cost for the *Physical Review* to process a paper (‘PR cost’). Please realize that these costs depend on the exact level of service, and that I have assumed a minimal level, as will be explained. The numbers given should be taken only as rough estimates to guide the discussion.

My proposal is given in the belief that the individual scientific article will remain the basic element of scientific communication. Science grows when individual scientists have ideas which they support with evidence, polish, and defend. A scientific paper succeeds when it presents an intellectually coherent idea. The formulation of such ideas is not a community process. The judging of ideas cannot be automated. Still, we can present and exchange papers both more cheaply and more effectively.

What are journals for? In the traditional publishing model, scientists rely on journals to provide four distinct functions. These are: distribution, refereeing, archiving, and indexing. That is, the journal (1) makes the article widely available in a readable form, (2) performs an evaluation and gives its imprimatur to the correctness and importance of the results, (3) preserves the results for future readers, and (4) provides a basis for search and location of the contribution.

I believe that the roles of distribution and archiving can, to a great extent, become the responsibility of the author if appropriate technical means are provided. The roles of refereeing and, to a lesser extent, indexing, require professional services. Following this path, we can save costs and improve communication at the same time.

Distribution. The most elemental function of a journal is to present articles in a form in which they can be distributed and read, and to make these articles widely available. Until very recently, this required authors to interact with a publisher and an editor, laboriously exchanging and correcting proofs. Newton and Bohr famously agonized over their page proofs. I do too, but I have the luxury of desktop software that gives me control over every detail of how my paper appears. Indeed, the greatest success of the current phase of electronic publishing has been the triumph of the author-prepared manuscript. Anyone with a computer can now prepare a paper that is cleanly formatted, with equations beautifully typeset, and can convert it to an electronic file that can be transparently viewed over the Internet.

This has made it possible for physicists to communicate by posting papers on centralized electronic archives such as the Cornell e-print archive [2]. In some subfields of physics—mine is one—the posting on the Cornell archive is the publication of record. It is the place that my colleagues go to read the paper, even after it is accepted by a refereed journal. The posting date determines intellectual priority. Through this medium, our papers are brought immediately to the attention of our community world-wide. The cost/paper of arXiv deposition is less than 2% of the PR cost. These costs are currently paid by the National Science Foundation and the Cornell University Library.

I believe that the advantages of communicating through unrefereed e-prints is so great that eventually this method will be adopted by all communities in physics. Refereeing is not incompatible with this means of distribution; rather, it is a cre-

dential applied at a later stage of the publication process. It may be that new electronic archives will be created to better fit the cultural styles of other communities within physics (for example, solid state experimenters). It is only important that these archives are centralized for each subfield and that they are given permanence, e.g., through hosting by a university library. I will refer to these collectively as ‘the Archive’. Once a community has adopted Archive publication, more advantages follow, as we will see below.

Archiving. A centralized Archive can improve the presentation of papers by providing guidelines and templates for the creation of electronic manuscripts. In the process, it can collect ‘metadata’—explicit identification of authors, for example—that would be valuable in cataloguing. Eventually, the Archive might provide tools that allow authors to produce papers in a standard format for permanent archiving.

There is no universally accepted standard today for electronic archiving of papers. The choice should eventually be made by a consortium of institutions interested in preserving scholarly electronic materials, including scientific societies, university libraries, and the Library of Congress. The sooner this consortium is assembled, the better. In principle, pdf could be used as an archival file format, as long as a particular well-defined version of pdf is specified [3]. In this case, a rendering engine producing that version of pdf could be used to present papers at the Archive. Authors would then ensure that their papers posted on the Archive appear correctly and have the correct associated metadata. This would then leave to the Archive and the consortium of which it is a part only the task of preserving the electronic files indefinitely. This is a challenging task, but one that could be automated in a scalable way, without requiring further human intervention for individual papers.

In a system in which preparation for archiving is the responsibility of the author, the cost/paper of archiving could likely be brought down to a few percent of the PR cost. At this level, the costs might be paid by the institutions that support the Archive. If it turns out that, to produce archival-quality files, a central authority must reformat each article, the cost will be higher—more than 20% of the PR cost—and the payment for this service will need to be collected by journals.

Indexing. Once one has a centralized Archive containing papers and associated metadata, it is possible to search and index this database. The result would be a bibliographic record of the field, to the extent that its papers have been stored in the Archive. In high-energy physics and astrophysics, the libraries of the Stanford Linear Accelerator Center and the Harvard-Smithsonian Astrophysical Observatory already provide bibliographic search engines that are up-to-date within a few days of posting on the Cornell archive [4, 5]. Thus, these services are about three months ahead of refereed journals and six months ahead of commercial services such as ISI. They provide not only author, title, and textual searches but also forward and backward

citation linking. As more subfields adopt Archive publication, this capability could be extended to all of physics.

It is important to note that, for this service, it is likely not possible to give the full burden to the author. It is the current experience that the references in author-prepared manuscripts contain many errors and ambiguities. In principle, these can be corrected at a later stage by authors and readers. Still, someone must receive the notices and make the corrections. Both of the services named above have several full-time employees devoted to this task and incur costs/paper of about 4% of the PR cost. It is possible that similar services in other subfields of physics could be donated by other laboratories or universities. The multiple databases could be consolidated either by sharing of data through the consortium described above or by a higher-level service such as the recently announced Google Scholar.

Refereeing. I have now argued that the model of author-prepared papers published in an electronic Archive can fill the roles of presentation and archiving traditionally offered by journals and would enable improved indexing and search of the physics literature. The relatively small institutional costs for these services could be funded outside the system of journals. But the final role, refereeing, cannot be given over to this mode of support. It requires a substantial professional infrastructure.

In some models for electronic publication, the refereeing step is conveniently eliminated or replaced by automated download or citation counting. This, I feel, is a mistake. The essential feature of refereeing is that the original idea presented in the paper should be confronted intellectually by a knowledgeable reader. That ought to remain a necessary criterion for the acceptance of that idea by the scientific community. Such engaged reading should be part of any evaluation of the authors for grant funding or career advancement. The fact that refereeing in the real world often falls short of the ideal does not make this any less true.

Refereeing has costs that must be paid. Referees are typically individual scientist who volunteer their time. But a cadre of profession editors is needed to manage the dialogue between authors and referees and, ultimately, to take responsibility for the decision to accept or reject a paper. For the *Physical Review*, this task accounts for 30% of the PR cost. Adding the costs of overhead, financial services, and an editorial office, the cost per article would be almost 50% of the PR cost, even if the only product of the journal is the decision to accept or reject the paper.

Who would pay these costs? The model in which physicists pay for each paper submitted has been tried, and it was a failure. Authors migrated to journals with no publication charges but with much higher subscription costs to libraries. I favor the ‘institutional membership’ model, in which libraries pay a fee which allows authors from their institutions to submit papers for refereeing. These fees would be tiered for

research institutions of different sizes, as is done now for APS journal subscriptions. Such a fee would be easier to collect than page charges paid by individual scientists—but only if the library community understood and supported this model.

I believe that librarians would see the value of this model, even though the journals would not produce a product that libraries can purchase and own. The model fulfills their goals—access to scholarly communication, with evaluation, search, and permanent archiving. It provides some savings even from the current prices of low-cost journals. More importantly, by focusing on payment for essential services, it eliminates the niche held by high-cost journals, and this would bring libraries very significant savings.

Conclusions. I believe, then, that the new technologies have enabled a change in the way physicists publish that is more profound than simply making journals available on-line. If we use these technologies wisely, we can shift to authors many of the responsibilities now managed by journals. We must, at the same time, identify the irreducible part of the journals' task that requires a professional staff, and a means to pay for their service. In this way, we can remake the literature in a way that improves its accessibility and allows it to grow to accommodate the future development of science.

I am grateful to Martin Blume, Mark Doyle, Paul Ginsparg, Patricia Kreitz, and to Stuart Loken and the members of the APS Loken II Task Force on Electronic Information Systems for discussions of the points raised in this article. Of course, these people are not responsible for the personal opinions I have expressed here.

Bio: Michael E. Peskin is a professor of theoretical physics at the Stanford Linear Accelerator Center, Stanford University.

References

- [1] V. Bush, *Atlantic Monthly*, **176**, 101 (1945).
- [2] <http://arXiv.org/>
- [3] See <http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml>
- [4] <http://www.slac.stanford.edu/spires/>
- [5] <http://adswww.harvard.edu/>



Figure 1: The author discusses a classic paper from the literature: *Phys. Rev.* **179**, 1499 (1969).