Electronic Publishing at the American Astronomical Society

Peter B. Boyce and Heather Dalterio

American Astronomical Society

Abstract

Electronic communication is changing the way astronomers work and how they communicate. By forming teams of people with a wide variety of expertise, taking small steps, being flexible and soliciting the opinion of users, the American Astronomical Society has been successful in several electronic publishing projects. Experience has shown that electronic publishing entails as much effort as the paper counterpart, but that many benefits accrue to the user. Only by remaining open to new ideas and incorporating new tools as they become available will the true benefits of the connectivity provided by the World Wide Web be brought to the community.

1 Introduction

As publishers, societies, librarians and scientists, our purpose is to communicate clearly to our scientific colleagues new information about the universe. Scientific societies, in particular, were founded to foster communication within the scientific community and to see that such communication adheres to a reasonable standard of scientific rigor and quality. With the Internet now offering easy and effective connections among scientists, the possibilities for communication have expanded. The American Astronomical Society (AAS) has set out to harness the capabilities of the Internet and develop means for providing accurate, reliable and up-to-date information to the world-wide astronomical community. We have chosen a strategy that highlights the use of open standards and freely available software. We have avoided becoming dependent upon commercial products and large proprietary systems. Our relatively small team has been able to produce a comprehensive program of electronic publications and services.

In an environment which is changing as rapidly as the electronic network, it is virtually impossible to predict the form electronic publishing will take in the

Preprint submitted to Elsevier Science *5 July 1995*

future. The AAS has approached electronic publishing as an exploration of various tools and techniques. We are keenly aware of the experimental nature of what we are trying to accomplish. Only through continual modification in response to user's comments and suggestions will we eventually arrive at a system which is effective, useful and affordable for the working scientist.

As soon as Mosaic burst onto the scene three years ago, we realized that we now had a tool which would allow true publishing of scientific literature, complete with special characters, complex math, pictures and graphs. Before that, it was hard for even the most optimistic person to foresee just how rapidly the technology and software would advance. Such a rapid pace of change in the mechanism by which we communicate is unprecedented. Rapidly changing times require a flexible approach based on modular design and adherence to standard protocols. The use of small, modular blocks ensures that functions within your publishing system can be upgraded as new tools become available. Adherence to standard protocols for the exchange of information between the various software modules ensures that your system will continue to work as improvements are made, and will be able to communicate with software and databases outside your own system.

The philosophy of the AAS has been to set out a plan to accomplish the transition to electronic publishing, to take small steps to gain experience with this new medium, to retain the flexibility to improve various steps in the publishing process and to apply tools and techniques which are appropriate to the task at hand. In this way the AAS has gradually built up both the expertise and the system required to take on the task of handling the full scale production process of publishing 28,000 pages per year in our peer reviewed journals.

2 Electronic Publishing

Let's be sure we understand what we mean by "electronic publishing." As we developed our plans at the AAS we uncovered a number of misunderstandings of what it takes to have a complete publishing system. There was no general agreement of what "electronic publishing" means. As we have explained before, everyone tends to focus on that aspect with which he or she is familiar. Authors concentrate on electronic manuscript preparation and submission. Readers focus on getting information electronically from the Internet. Librarians see the delivery of information to users, and the often overlooked function of archiving material electronically, as important components of the overall process. Publishers deal with the process of handling electronic manuscripts, providing copy editing, formatting and typography services and producing versions suitable for delivery to the end users. They also worry about collecting the revenue needed to keep the whole operation financially viable. Electronic publishing of scientific journals embodies all of these functions. Electronic publishing is all this and more.

In our view, a full electronic publishing system needs to include:

Author preparation and submission of manuscripts. Use of the author's keystrokes will reduce the need for the author to proofread the manuscript.

Editorial process – peer review. At this stage, the peer review process helps to ensure scientific quality. While there are those who argue this point, there is a clear difference between unrefereed conference proceedings and the good, peer-reviewed journals.

Typography – copy editing, typesetting, page/screen makeup. One of the characteristics of a "good" journal is the clarity of the writing and the readability of the pages. The importance of these steps in improving the ease and accuracy of the information transfer is not fully appreciated by many.

Database preparation – conversion and insertion of the manuscript into the database. As described below, the database is the crucial, central element in an electronic information dissemination system. Developing and maintaining effective search and retrieval tools is included under this heading.

Production and dissemination in multiple output formats. The final step in the delivery system is the provision of the information to the user in a format suitable for the purpose and the medium of delivery.

Archiving – maintenance of integrity and version control. The permanent storage of the information for use in the future is a continuing task. The evolution of storage technology and the subsequent moving of the information to new generations of storage media will become a growing task for publishers and librarians.

The central feature of any electronic information distribution system is the database of information, coded and organized such that the desired information can be located and retrieved in a form that is most useful to the reader. Electronic publishing operations should be directed toward preparing information and putting it into the database, searching for and retrieving information from the database and maintaining the integrity of the database. From our early experience, we have found that readers want papers and other data in various forms. Output to a screen is important for browsing and searching but is not well adapted for extensive reading. Paper copies of selected information or individual articles will definitely be required. These have to be transmitted and printed out at the reader's local printer. Other societies are experimenting with the provision of journals on CD-ROM. Finally, it is clear that we are not yet ready to abandon altogether the traditional typeset, printed and bound journal. Except in the case of a new all-electronic journal, we have to ensure that we can produce the familiar, printed volumes from the same material and that all versions are identical.

Electronic tools provide expanded capabilities for search and retrieval of information. In addition, electronic distribution of information can extend to many different sorts of data. André Heck has encouraged us to expand our thinking; to go beyond traditional publishing to the provision of a wide variety of information in various forms. We can not restrict ourselves to just distributing articles. We also have to provide access to the underlying data and tabular material from articles and books as well as information about our organizations, our meetings, news items affecting our science and other material which the scientific community would find useful. As a database provider, Heck is insistent on the need to maintain the quality of information on the WWW, including not only careful preparation of the information, but also the necessity of maintaining files and archives in a usable state once you have them. It is becoming clear, as we accumulate more and more electronically accessible information, that the issues of quality, reliability and retrievability will become ever more paramount.

While publishers understand the following section very well, it is not generally realized either within the author or the reader community that effective publishing is hard work. While the rare exception will emerge, anyone who has had to read a number of manuscripts will agree that there is a wide variation in the quality of the works that authors produce. The quality of writing is very uneven. Some authors write very clearly, but many authors do not write either clearly or effectively. It is also surprising to find how careless some authors are. Helmut Abt in a recent study has found that 12 percent of the references cited in the *Astrophysical Journal* contain an error. While references to the *ApJ* are checked and corrected by the copy editors, it is just too costly to check all references to outside work.

It is also clear that authors make poor typesetters. Anyone who has had to referee papers or proposals where there is a page limit will agree that authors can become so entranced with their own words that they cram as many words on the page as possible, ignoring the factor of readability. This is not the way to convey information to the reader. With modern word processors giving the author control over the font size used, the problem has become even worse. We regularly see proposals written in 8 point font with one centimeter margins. It is at this point that the publisher and typesetter have to step in and improve the appearance of the page, enhance the readability of the paper and use their expertise to present the information in a clear but compact format.

As we have progressed into the transition from small projects to the full scale production environment, we have come to appreciate that the production of a major journal is an overwhelming task. The *ApJ* publishes 600 pages every ten days. That works out to more than 100 typeset pages every working day. This is a far cry from putting a few pages up on the WWW every six months or so. Informal procedures which involve a lot of manual intervention are not an option that any publisher can afford.

Finally, we have seen from our experience to date that the use of electronic tools require a higher standard of precision than in the days of paper manuscripts. A reference number implies a link to the list of references at the end of the paper. A superscript implies a math function. Even though they both may look the same on the printed page, you can not substitute one for the other. Authors (and typists) have to be careful to put in the functionally correct symbol, or else the reader will not be able to jump to the reference and see the abstract on line.

3 The AAS Program

The AAS has been providing electronic acceptance, manipulation and delivery of meeting abstracts since 1992. The abstract process has all the functions of a full-fledged electronic publishing system, including the collection of abstract fees. Our three years of experience with the meeting abstracts has been invaluable in helping to design our system for bringing the refereed journals on line. The AAS has published an electronic version of its monthly register of jobs in astronomy since 1993 and last fall launched a monthly electronic newsletter for members. Perhaps of most interest to this audience is the progress toward putting the AAS journals on line. Starting in September of this year, the AAS will publish an electronic version of the *Astrophysical Journal Letters*. A sample issue is available on WWW from the AAS homepage. As part of our learning process, and as a service to the scientific community, the AAS collects data tables from all the north American astronomy journals and publishes them in machine readable form on a CD-ROM every six months.

What have we learned from our experience so far? Most gratifying is the ready acceptance of the community to the electronic abstract submission service. Starting with slightly more than half the abstracts coming in electronically in 1992, we have achieved over 99.5 percent electronic submission for the last two meetings – there was just one paper abstract for the June, 1995, meeting.

Secondly, we have found that effective automatic processing is a critical part of our system. We could not handle the increased abstract load without automatic processing. Our experience shows that manual intervention is very costly. Mistakes in the electronic files can be very time-consuming to fix. How-ever, people do make mistakes, and some manual intervention is necessary. It is important to keep the number of mistakes as low as possible. We find that fewer than 10 percent of the abstracts have problems which require intervention, but fixing those files requires half the processing time.

Additionally, it has become clear that operations are more closely linked than before. It is more critical than ever to think through the entire process from input all the way through the production of the output, including the financial transactions required along the way. A poorly designed entry form will affect the entire process all the way through. Just as the authors have to be precise about what symbols they use, we find that we have to understand exactly what it is we want to accomplish at every stage in the process and work out effective procedures from the beginning to the end. This has required a major change in the way we approach projects.

We now use a matrix management approach. For each project we now build teams that include people with relevant experience from a cross-section of internal departments, and often include outside consultants.. Only in this way are we able to make the most effective use of the electronic tools available to us. It has been a difficult transition for the office, and we see similar transitions taking place as our publishers and editorial offices gear up to handle electronic manuscripts. Our publishing effort can no longer be thought of as a number of separate, independent offices that process the manuscript serially. Instead, each component of the process must be closely integrated with the others. Compatible software, standards for data interchange and full integration of the processes into the overall publishing stream have been an absolute requirement. This transition may be difficult for many organizations. We have found that it takes thought, planning, hard work and the dedication of everyone involved to make this possible. It is a paradigm shift, and the difficulty of making such a change within an organization must not be overestimated.

4 Electronic Peer Reviewed Journals

Turning to our peer-reviewed journals, we have been experimenting with an electronic version of the Letters section of the *ApJ*. We have chosen to experiment with the Letters for several reasons: First, it is a prestigious journal of short, four page papers and is one tenth the size of the main *ApJ*. Second, the Letters section comes as part of the subscription to the main Journal. Therefore, we can experiment with an electronic version of the Letters without jeopardizing the financial stability of the journal. Third, the Letters contain papers for which there is a desire to publish rapidly. The *ApJ Letters* provide an ideal vehicle to use in developing electronic procedures and trying new features.

We are well along in the process. As part of our philosophy to involve people from every stage of the process, we have assembled a team of AAS staff, outside consultants, editors and staff from the University of Chicago Press. To gather feedback from users and investigate various production issues, we have put together an electronic version of the November 1, 1994, issue of the *ApJ Letters*. Full production of the *Electronic ApJ Letters* will start with the July 1, 1995, issue. Thanks to support from NSF, the journal will be available on the WWW at no charge for the first year. During this year the AAS will be developing the mechanisms to charge for electronic subscriptions.

The *Electronic ApJ Letters* will consist of a central database of manuscripts, logically marked up in SGML. From this database it will be possible to browse the journal using a graphical browsing tool such as Netscape or Mosaic. Figures will be presented in the text as small "thumbnail" sketches, with larger versions available at the click of a mouse button. It will be possible to print postscript renditions of selected articles on the reader's own printer. Clicking on a reference will connect the reader to the Astrophysics Data System where the abstract (and, eventually the full text) of many articles will be available. Each article will also have an associated list of articles which have been published subsequently and which refer back to the article. We have given this feature the name of "forward referencing".

But, it is very clear that the real challenge on the WWW is to be able to locate the information of interest. It is in this area where we have put a lot of effort. We have been part of the process of refining an outstanding search engine, known as "ISITE," which has been developed by the group at CNIDR. ISITE has the capability to search the full text of everything in the database for any number of words, using synonyms, stems of words and specific combinations of words. Proximity searches (two specified words within a paragraph, for instance), boolean searches and author and title searches are standard. ISITE is a powerful engine, and several features have been added as a result of feedback from the users of *ApJ Letters* demonstration issue.

Finally, we can already guarantee that the electronic version of the *ApJ Letters* will be available a month before the paper version reaches the readers. It may be possible to save additional time as we become more proficient with the electronic production process. Even more time will be saved when we get a higher percentage of electronic manuscripts.

5 Other Projects

How does all this effort tie into other electronic distribution projects which are under way? There is a lot of money going into "Digital Library" projects in the United States which are setting out to produce a system similar to what the AAS has already achieved. Except for the scale of the projects, we have not seen any features in the larger projects which are not included in the AAS electronic journal plans. Indeed, with our small and expert team, the AAS has been able to react to changing conditions more rapidly than the larger groups. As the AAS gets into the production of the major journals, our flexibility will diminish, but our willingness to experiment with small segments of the journal will remain unabated.

There has been recent discussion about the use of "archives of electronic preprints" principally by the theoreticians among the high energy physics community. Originally, this seemed like the net-surfer's dream. Authors would put up their paper, and it would immediately become available to the WWW community for reading and discussion. Gone would be the expensive and time-consuming features of the scholarly journal. Gone would be the peer review process. Gone would be the need for copy editing. This would be instantaneous communication at its best. It doesn't seem to have worked out that way. Initially, people who look at the contents are impressed. For the first time, it is possible to achieve rapid communication with everyone on the net. It is possible to do limited searches for papers of interest, and it is possible to get printouts of the papers.

But the opinion of the physics community has reversed. Physicists are beginning to express dissatisfaction with the generally low quality of the postings. The readers, overwhelmed by the amount of data on the WWW, can not take the time to understand whether every author prepared paper is worthy of taking the time to read. What seems like a wonderful opportunity for authors is generating information which readers will, in the long run, generally not bother reading. This is not a way to maximize the exchange of information.

6 Predictions

We expect that growth of electronic distribution of information will continue. We also know from a survey done by Elsevier that readers want to read less. It is clear that any electronic publication system will have to maintain a high standard of quality, filtering out incorrect and less useful information. We hear from the AAS members that they will turn to the peer reviewed journals in preference to unrefereed conference proceedings. In the electronic world, where the tools now exist for nearly anyone to put a collection of papers up on the WWW, it will become even more important for the electronic publications to maintain, and even improve, the quality of their content. The editorial process will become more important than ever.

Nevertheless, the paradigm shift is upon us. The communication system is shifting rapidly. As publishers, with a responsibility for maintaining the flow of high quality information, we will have to continue to be aware of these changes and their implications. We predict that the WWW will force the growth of interdependence among researchers, institutions, libraries, publishers and vendors. No longer will it be possible for one group to exist and be effective without cooperating with the world outside. We see a trend, even as we write, toward effective distributed databases of important and useful information, such as the Astroweb Consortium.

Many WWW enthusiasts have assumed that publishing on the Web would be free and instantaneous, or nearly so. Our experience shows that the careful preparation of manuscripts needed to operate a successful electronic system will be nearly as costly and as time consuming as that required to produce a good journal on paper. Therefore, we do not anticipate any large reduction in the cost of the journal. It is impossible to maintain the high quality expected of the AAS scholarly journals without paying for it, both in money and time. If the authors are careful with manuscript preparation, the production time can be much reduced, but we cannot escape from the fact that good peer review and accompanying copy editing takes considerable time and effort.

Which brings us back to the preprint question. Right now the preprint archives are the only electronic way to reach a wide audience. We believe that, with the peer reviewed journals coming on line, the preprint archives will lose their attraction as places to publish. There will always be a need for rapid, unrefereed communication, but ways will have to be found to filter out the unreliable and unimportant (even crackpot) papers so the reader will not have to plow through everything anyone wants to post. We expect that the massive preprint archives will give way to more informal exchanges among workers within small areas of astronomy. Perhaps discussion groups with an editor/moderator will become important in advancing our science.

In summary, major change is in the air. We have to break free from the confines of the printed page. We must look to using the electronic medium and unprecedented connectivity in new and more effective ways. It is a truly exciting time, and we have barely begun to envision what is to come. But, in developing any new tools and methods, we must never forget that the most important attribute will be the quality of the information we make available.

7 References

[1] Boyce, Peter B. and Dalterio, Heather J. *Physics Today*, submitted, 1995

[2] Heck, André URL: http://cdsweb.u-strasbg.fr, 1994

[3] See WWW pages for the Optical Society of America, http://www.osa.org and the American Institute of Physics, http://www.aip.org, which will make the *Journal of Chemical Physics* available on CD-ROM in July 1995

[4] Abt, Helmut A. Private communication, 1995

[5] To see the AAS meeting abstracts, visit URL: http://aas.org/ and go to the link for "current meeting."

[5] http://aas.org/ and follow link to *Electronic Astrophysical Journal Letters*

[6] Chris Biemesderfer (ferberts associates, cbiemes@ferberts.com), Evan Owens (University of Chicago Press, eowens@journals.uchicago.edu), Archibald Warnock III (AWII Enterprises, warnock@clark.net) and Jim Fullton (Clearinghouse for Networked Information Discovery and Retrieval, jmf@cnidr.org)

[7] Publisher's Roundtable, April, 1995 Meeting of the Materials Research Society

[8] http://www.stsci.edu and other sites around the world