

ELECTRONIC PUBLISHING IN ASTRONOMY

PETER B. BOYCE

American Astronomical Society

2000 Florida Ave., NW, Washington, DC 20009

Abstract. Links are important. The providers of electronic astronomical data are linked together the way no other science is. Electronic journals are part of this distributed, linked resource. Permanence is important. The AAS journals use a new publishing process which ensures the ability to maintain permanent access to our electronic journals. Process is important. With the right process, we have been able to add links automatically to our journal; links to references, links to citations (where the electronic material exists) and internal links for ease of navigation within the journal. Readers like this. Links are important.

1. Introduction

As of late 1997, astronomy still leads other sciences in the quality and number of electronic journals. Beginning in 1992, the American Astronomical Society (AAS) has pioneered the successful electronic publishing efforts in astronomy. Boyce and Dalterio (1996), Boyce (1996), and Boyce, Owens and Biemesderfer (1997) have summarized the AAS electronic publishing program. Calvin (1996) has given the AAS electronic journals a positive review. To date, we have found no other electronic journal which provides the number of features found in the electronic version of the AAS's *Astrophysical Journal* (ApJ), and no other science provides as much as forty percent of their literature in a good electronic format; HTML presentation with copious links to references and citations, backed by a robust, electronic archival version. By 1998 we expect that 75 percent of the astronomical peer reviewed literature will be on line in such a format as described below.

The Letters Section of the ApJ has been available electronically since mid 1995,¹ and the main journal joined it on line in January 1997. The *Astronomical Journal* and the *Publications of the Astronomical Society of the Pacific* will have electronic versions by January 1998. All three journals are now published by the University of Chicago Press. The Elsevier electronic journal, *New Astronomy*,² started in late 1995 and includes all the features of the ApJ plus the inclusion of video clips which will also become part of the AAS journals in January 1998. *Astronomy and Astrophysics Supplements* have been on line since 1997 and the main *Astronomy and Astrophysics* journal is scheduled to be on line in 1998.

2. Links

What makes a good electronic journal? After two years of development and a further two years of publishing, it is clear which features are appreciated by readers. By far the most important feature is the presence of links, both internal, for navigation within the document and external, to the references and citations which can be retrieved by the reader instantly. The internal links make it possible to start browsing an article by looking at the figures, and then to jump right to the text where the figure is discussed. Or a reader can start with the reference list, find a reference and jump into the middle of the article where the reference was cited.

We produce two versions of the journal in HTML format, designed for on-screen browsing and effective linking. Special symbols and mathematics are inserted as graphic images using a specially

¹<http://www.journals.uchicago.edu/ApJ/>

²<http://www1.elsevier.nl/journals/newast/>

designed library of GIF images which are inserted in the proper places. Although the *Astrophysical Journal* requires the use of about 1,200 special characters, the processing has been automated to the point where this poses no particular problem. One of our HTML versions includes the whole article in one file. In the other version, the article is broken into smaller chunks, allowing a user with lower bandwidth connections to only download the abstract, or to look at only the references or the figures. Of course, each chunk has the same navigational links. About 20 percent of the HTML readers use the chunky version of the articles.

However, readers need to do more than just read articles on the screen. They want to be able to print out articles of particular interest to them. For this, we produce a page image version of the journal using Adobe's PDF format. Many other publishers have begun publishing PDF articles as their only electronic version. They lack the links, machine readable tables and other electronic-only features which our readers find indispensable. An analysis of our access site shows that readers access the HTML version six times more often than the PDF version, a ratio which indicates that readers browse the material in HTML and only use the PDF version for printing out the articles of interest.

The real strength of the electronic journals in astronomy lies not in the articles themselves, but in the links to the outside references as well as to the articles which cite the one being read. Readers say these external links are the single most important feature, but they do not come automatically. First, the reference material has to be available on line. Second, the information has to be identified using a uniform naming scheme so that we know how to access the specific information. Third, to ensure longevity of the links, they have to be made using the uniform names coupled with name resolvers which translate the logical names to the actual "hard coded" URLs. Each information provider must run a name resolver. Fourth, the linking system has to be designed so that the links may be added automatically during the production process. Adding links by hand, indeed doing any action by hand is prohibitively expensive.

Astronomy is fortunate that much of the required material was already available on line. The abstracts of the core journals in astronomy already existed in the NASA-supported Astrophysics Data System (ADS), a searchable database of abstracts.³ At the same time as the AAS started on the electronic publishing project, the ADS began to accumulate the full text page images of the material in addition to the abstracts. Full text presentations of much of the older astronomical literature are now available on line with accompanying lists of references and citations.

Astronomy was also fortunate to have several data centers in which published tabular data and catalogs were available in electronic form. Ten years ago, the data centers had already devised a system of for assigning logical names to published articles. It was a simple matter to expand the use of these names, known first as "refcodes" and now called "bibcodes," to the electronic journals. The material in the data centers was organized by astronomical object name and provided pointers to the paper literature. With the appearance on the scene of the electronic journals it was simple to tie into the existing system, providing another entry point by which to link to abstracts of articles and to the full text page images at the ADS. By officially linking together these internationally distributed *electronic resources, and smoothing a few rough edges at the interfaces, astronomy now has a full fledged, linked data system which operates in a collaborative mode.*

To recount, astronomers worldwide now have access to much of the core literature, along with published data tables and data on astronomical objects. All of the material is linked together through a common naming convention and using name resolvers to ensure robust links. We call the whole collection, the *Urania* resource.⁴ Astronomy now has a very powerful electronic information resource. As she looks down from Olympus, the Muse of astronomy has to be proud that we chose her name for this research tool.

3. Permanence

The inherently impermanent nature of electronic information is a major drawback for scholarly material, much of which is expected to retain value decades and even centuries into the future. Garrett and Waters (1996) authored a report on archiving Digital Information which lays out the problem. Many of us still remember eight track tapes and 45 rpm records, yet it is virtually

³<http://adswww.harvard.edu>

⁴<http://www.aas.org/Urania>

impossible to play music on either of those media today. In the scientific world, much information stored on 7-track tapes was simply abandoned because it was too difficult to read the tapes. Changes in technology make our storage media obsolete. But, software advances and changes in storage format also render material inaccessible after a relatively short lifetime. In either case, with today's pace of change, scholarly material with expected lifetimes of more than five years must be designed with the archival storage requirements in mind.

Electronic archiving implies that users have continued access to the material. For simple page image documents with no links, this is still fairly simple. But, based upon the reception given the electronic ApJ by the community, storing the pages without also storing the links, machine readable data tables, video clips and other electronic-only enhancements will not suffice. Keeping electronic material accessible implies that the material will have to be managed during its lifetime. Translations to new storage media and new formats will have to be done. We are not talking just about saving the material, we are also envisioning updating the material to incorporate new technology, keeping the material available to one new generation of Web browsers after another, or even making the material accessible as we inevitably move beyond browsers.

How do we do this? The answer lies in preparing the original material in an open, standard format which incorporates logical markup. At present publishers have generally agreed to use the Standard Generalized Markup Language (SGML). The ApJ archival database, which the public never sees, is composed of manuscripts coded in SGML. From this database, we can automatically derive both HTML screen versions and the PDF version.

Updating the public versions now becomes an automatic, almost trivial, process. We have updated the HTML versions of the electronic ApJ three times since its inception. Originally we presented tables as graphics. All the electronic issues now support the HTML table specification, even though we started publishing before the specification was developed. Our HTML versions now take advantage of the expanded HTML character set, reducing the number of character graphics we have to use. Perhaps the most noticeable change has been the incorporation of the citation list with every article. All this, plus a host of smaller improvements, have been done by modifying the translation program which produces the HTML version, and reconstituting the whole corpus of publicly available material. The success of this translation leads us to believe that we can do the same with the complete archival database when the need arises, translating it to whatever new standard is appropriate. As of late 1997, XML appears to be emerging as the new standard. By planning for a robust, long-lived archival format from the beginning we have taken a large step toward ensuring the effective access to our scholarly journal far into the indefinite future.

As we have demonstrated, maintenance of an electronic archival journal is far different from simply putting the material into a storage medium and leaving it forever. Maintaining successful and effective access over decades requires active management of a well prepared database. Since the archive management function is tied so closely to the original preparation of the database, it seems logical that the responsibility for maintaining the journal archives should rest with the publisher.

How to pay for the necessary active management is a problem. Commercial publishers are loathe to take on the responsibility because they envision an increasingly heavy burden to maintain material which is used less and less frequently. The AAS is a non-profit publisher with some moral obligation to maintain the availability of the research corpus in astronomy. Still, we have to worry about funding the necessary management activity. We have chosen to establish a journal archive fund into which we will put an amount of current revenue sufficient to enable us to completely translate our archival database every five years. We fully expect that the growth in computing capability will outpace the growth of the size of the journal archive. Our experience leads us to expect that, if we maintain our own archive, we can do so at less than one percent of the cost of the journal per year. Whereas, if any other organization tried to take over archival maintenance it would cost substantially more.

4. Process

Producing an effective electronic journal requires approaching the task with a new mind set. Publishers have been putting words on paper for hundreds of years. But an effective electronic journal has to go beyond the printed page. The reality of that statement has been demonstrated by the community's acceptance of, and demand for, links, machine readable tables, videos and other ma-

terial which can not be rendered on paper. Yet, most publishers are still concentrating on creating a typeset paper page first and generating an electronic rendition after the fact; a process which includes all the costs of the traditional journal plus additional costs for the electronic version.

Since the UCP and the AAS have been partners from the beginning in the development of our electronic publishing process, we have been able to retool the entire publishing process and to take advantage of some of the economies which an electronic environment can provide. Our process now focuses first on producing the electronic database of manuscripts in a robust, richly-tagged, SGML format. Then we derive both the paper and public electronic versions from the electronic archival database.

We are realizing a number of advantages. We accept electronic manuscripts and translate them into SGML automatically. We use carefully designed, electronic tools to copyedit on the screen. We use electronic transmission of material. We create links automatically and check that they work by using tools maintained at the ADS. In short, we have abandoned the traditional paper-based process while still keeping the rigor of our peer review and the high quality copyediting which has been a hallmark of the ApJ for 100 years. There are some cost advantages to our new process, too. We are still in the development phase, and it is too early to know how much we are going to save in the long run. But, we do know that we are producing both paper and electronic versions (including the contribution to the electronic archive maintenance fund) for less than we were producing paper three years ago.

The new electronic process makes it possible to shorten production time. Our prime example is the electronic ApJ Letters. By typesetting them in house, we are able to produce and post an electronic version without page numbers within two weeks from when the scientific editor accepts the paper and six to eight weeks before the paper issue date. We are posting individual papers as they are completed and only collecting them into volumes when the paper version is ready.

5. Summary

Astronomy is well ahead of most other disciplines in providing accessible, interconnected electronic information. The scholarly journals, while important, make up only a part of this information resource. The power of the full system lies in combining the electronic journals with the bibliographic databases, electronically accessible historical literature and a full range of observational data ranging from the raw observations to the published catalogs. Yet, all of us associated with the electronic delivery of information feel we have barely begun to understand how to make the most effective use of the new technology. The future will surely see impressive new developments which will help accelerate the pace of discovery.

6. Acknowledgments

Evan Owens of the University of Chicago Press and Chris Biemesderfer of ferberts associates are responsible for making the AAS program a success. Our work has been supported in part by a grant from the National Science Foundation and a cooperative agreement with the National Institutes of Standards and Technology.

References

- Boyce, P.B. and H. Dalterio (1996), *Physics Today*, Vol. 49 pp. 42
Boyce, P.B. (1996), *Computers in Physics*, Vol 10, pp. 216
Boyce, P.B., C. Biemesderfer and E. Owens (1997) *Ser. Rev.*, Vol 23, No. 3, pp.1
Calvin, W. H. (1996), *Science Surf*, 96, 2 <http://WilliamCalvin.com/column02.htm>
Garrett and Waters (1996) <http://lyra.rlg.org/ArchTF/>

Comments

GENOVA: A comment on a question about the cost of the efforts asked of authors in electronic publishing. The possibility of checking the contents is increased - e.g., electronic tables are better checked than by the referee, tagged object names, having been checked, will be better recognized ten years from now than in the present situation.

BOYCE: You are right that better tools for the author will relieve some of the burden of preparing electronic manuscripts. I should add that in the electronic era, where we want to take advantage of the automation, we, as authors, have to be more precise about references, tables, links, etc.

EICHHORN: PASA intends to build links into reference lists. They plan to ask the authors to provide the links. When the *ApJL* appears in electronic form, will the page numbers be known?

BOYCE: No, it will not. This means that we will need to have different identifiers for articles before pagination. When the papers are posted one by one the page numbers are not known. The publisher will have to supply a document ID. It should be published with the paper, and there will be tools to provide that information upon query as well. There had to be a correspondence map to equate the various article IDs.

SEIDELMANN: What is the future of costs for the libraries? Currently costs are going up, so subscriptions must be cancelled. What is the future of libraries?

BOYCE: Our experience shows that by retooling the production process cost savings can be achieved. Other publishers seem to be charging more for the electronic journals, but they have not retooled their process yet. In astronomy the combination of pressure from the low-priced AAS journals and the preprint archives should help to contain costs.

CHRISTENSEN-DALSGAARD: Comment 1. This is a comment relative to publishers being responsible for the historical record. Libraries now collect past issues and are willing to handle electronic journals more than eight years old. Comment 2. As a society space is in a unique relation to standards. You have mentioned SGML, HTML and XML. There are more, such as those for access and query: Z39.50, for storing information: VICML, and NYTIME.

BOYCE: Concerning comment 1, the publishers are better able to handle the links but there are cost problems for commercial publishers. Concerning comment 2, We have tried Z39.50 access for remote queries to our journal, but that standard is not implemented with access control so we can no longer use it. The situation is changing so fast that many of the older standards are, for the moment, irrelevant. It is a problem, but we have had to make our choices based on things which we know will provide a graceful path for migration into the future. The critical item is to use open and public formats, and protocols.

HJELMING: Your talk deals with the present and the future. What is being done to include the many decades of journals in electronic form?

BOYCE: The ADS is doing that for major journals as you will hear about in a coming talk.

DE GEUS: Regarding the archivability, the commercial publishers take this issue very seriously. We don't consider it to be entirely the role of publishers, but libraries will need (and want) to continue the ability to archive materials. We talk to many libraries to find out what their wishes are in this respect, and try to provide as good a service as possible.

SORENSEN: I am worried about your comment that the publishers need to be the keepers of information in the electronic community. I fear for the continuity and access implications. Also, you have not addressed the cost implication of maintaining a fully transparent system as new technologies become available.

BOYCE: We have done two things to make archiving affordable. First, we prepare from the beginning for the ability to update and move to new technology in a way which does not use a lot of resources, i.e., automatically. Second, we have established a fund for archiving into which we put a small portion of the present income. This will allow us to update the archive completely every five years.

ABT: With regard to finances, traditionally most of the cost of producing a journal is in type setting. By convincing authors to submit electronically we provide savings to finance the on-line editions. All manuscripts submitted electronically are used in that form. However, receipt of electronic manuscripts is often not trouble-free, so we hesitate to ask volunteer science editors and referees to do the extra work. But the techniques currently being explored is to post the manuscripts on the Web with passwords so that the scientific editors and referees can call them up. Peter, how long do you think that CD-ROMS will still be useful?

BOYCE: CD ROMS will lose their attraction very rapidly. I put the lifetime at a few years at most.