Electronic Journal Publishing Seven Years On: Is the Revolution Over or Just Beginning?

Evan Owens

The University of Chicago Press, 1427 East 60th Street, Chicago, IL 60637-2954 USA
eowens@press.uchicago.edu

Abstract. The revolution that began with the first journals delivered via the Web in 1995 is now seven years old. Drawing on the experience at the University of Chicago Press in working with the American Astronomical Society, this paper examines technological developments in all aspects of the publishing process from author to reader looking at what has changed and what has not changed in last decade. It concludes by speculating as to whether this publishing revolution is over or just beginning.

1. Introduction

It has been just over ten years now that my colleagues and I at the University of Chicago Press have been working with the American Astronomical Society on electronic publishing issues. And it is almost exactly seven years to the day since we put the first issue of the *ApJ Letters* on the web in July 1995. A lot has happened in those years; this seemed like a good moment to look back as well as ahead. I can't possibly do what Virginia Trimble does in her legendary review articles in the *PASP*, but I hope that you will find my highly selective survey of the electronic journal publishing world as seen from a publisher's vantage point of interest.

I'm going to focus today on technology and the impact of technology on the various stages of the publishing process from the author to the reader. Along the way, I will also report on some recent developments with the journals of the American Astronomical Society and on some other publishing industry projects in which we at Chicago have been participating: citation linking and archiving. As those were the topics of my talks at LISA II in 1995 and LISA III in 1998, this will be a brief update in those areas as well. I'd like to conclude by speculating for a few minutes on the state of the revolution: is it over or just beginning?

2. Online Journals

I think that it is clear to all of us involved in libraries and publishing that something amazing has happened in the last few years. The growth of electronic journals and of electronic information resources of all sorts available via the web

has been truly fantastic. Getting precise numbers about all this content isn't easy; however. Here are a few indicators:

- At a recent Council of Science Editors meeting, Donald King claimed to have counted 17,562 refereed scholarly periodicals of which 10,195 available online. A representative of Elsevier commented that number was close to their own count.

- CrossRef, the publishers' association promoting the use of Digital Object Identifiers and cross-publisher linking, currently has 6,550 journals from 135 publisher with 5 million articles in its linking database; all available in full text online. I'll talk more about CrossRef later.

- Highwire Press at Stanford University hosts a collection of 335 primarily biomedical journals with 1,207,652 articles online.

- JSTOR, the journal scanning project launched by the Paul W. Mellon foundation, has digitized more than ten million pages of social science and humanities journals.

- Finally, the University of Chicago Press, one small publisher, has 16,000 astronomy articles, 205,000 digital object identifiers in our database (including content at UCP, JSTOR, and ADS) and approximately 1 million files on our web site for our collection of online journals. The number of astronomy articles on our site increases by about 3,300 per year.

Just thinking about how many files those ten thousand online journals represents makes me tired . . . and worried about the long-term maintenance of all this electronic content.

The usage of online materials has been growing as fast or faster than the growth of the available content: overall usage of our astronomy collection increased almost 50 percent in 2000-2001 and the increase in our medical journals was even greater. The demand for electronic articles in the sciences is very strong indeed. And we are just one publisher with a few dozen journals. Imagine how many articles are flying around the Internet from all those 10,000 titles now online. This is the really big news of the decade in electronic publishing: lots of electronic content is now available, lots of people are using it, and demand continues to grow. But is it a revolution? Or just rapid delivery of the same old stuff that we have been publishing for the last century or so?

3.   Publishing Technology

Online journals are only part of the picture of the effect that technology has had on publishing in the last few years. Following in the tradition of many talks by Peter Boyce, I'm going to march through the publishing process from author to archive.

4. Authoring

4.1. Text:

I must admit that authoring tools has not been an area of exciting developments in electronic publishing during the last decade, even though the quality of author's files is terribly important to those of us who have to work with them. For text, TeX macro packages such as AASTeX and others developed in the early 1990s continue to give good results, even if authors aren't wild about using them. Astronomy authors have always submitted reasonably good files; after a decade of AASTeX, almost all *ApJ* and *AJ* files arrive in great shape. Across all disciplines, it appears that the quality of author submissions has improved as general computer literacy has increased; even medical doctors can manage to create a reasonable text file most of the time now. The bad news is that file is likely to arrive in Microsoft Word format. There are no signs of widespread use by authors of SGML or XML anytime soon, much as we in publishing would like that to happen. Unfortunately, richly structured markup is a lot of work; probably more than we can expect from authors.

4.2. Graphics:

Astronomy graphics have always been sophisticated: figures of amazing complexity and beauty. It has taken time, but typesetters and printers have gotten better at dealing with the idiosyncrasies of astronomical PostScript. In most other disciplines, however, graphics are a big problem. Authors use a wide range of applications to create images: Power Point, Excel, graphics packages, word processors . . . And authors don't understand the difference between an output format such as TIFF or EPS and an application file.

4.3. Multimedia:

We have a long way to go before authors are fluent in the fancier multimedia formats. We have had authors try to send us two gigabyte video clips and wonder why we didn't think that was a good idea.

Overall, the general trend in publishing is to expect more from authors and to push more of the burden onto the authors. Authoring rich or complex content is still a severe challenge and a serious impediment to really interesting electronic publishing projects.

5. Submission and Peer Review

The most interesting application of web technology since the online journals of mid-1990s has been in peer review systems. Web peer review is really big right now, with lots of vendors and products competing in this space. Many publishers are licensing these products for use in their entire stable of journals. In the best astronomy tradition, however, the AAS identified the need in this area early on and commissioned us to develop a system custom built to their specification. We rolled that out in 1999 for the *ApJ* Tucson office and in July 2001 for the *ApJ Letters* Cambridge office. The advantages of web peer review are speed and reduced staff costs. The disadvantages are that most systems put

a considerable burden on the authors and the reviewers, if no more than learning how to interact with the system. For journals with high initial rejection rates, dealing with a lot of files that aren't going to ever be accepted is also a serious consideration. (Not an issue in astronomy!) The big long term advantage is that it makes it possible to contemplate the peer review of non-print content, material that can't be mailed to the reviewer in an envelope. Of course, you have to be publishing that sort of content and most e-journals aren't there yet . . . and may never be if it isn't appropriate to the discipline. Overall, a really interesting application of the web technology.

While we are on the topic of peer review, here are some interesting numbers from the *ApJ*. In the last three years, the approximate median time from submission to acceptance has declined from 150 days to 100 days. I would like to say that the credit for the improvements should go to our web peer review software, but in truth, it was mostly the strong management of Rob Kennicutt, the *ApJ* editor in Tucson and of Julie Steffen and Kerry Kroffe in Chicago. The *ApJ Letters* processing times have always been considerably faster than those of Part 1. But even the Letters office sped up slightly when we moved them to the new system - a pleasant bonus.

The interest in web peer review is one more manifestation of the desire for a shorter publication process. Authors, editors, readers: everyone now expects and demands "web time" rather than "print time."

6.   Editing, Typesetting, Printing, Distribution

The major event of the last few years in the production area has not been technological at all: it is the major movement from US and European suppliers to overseas vendors of keying, conversion, scanning, typesetting, copy editing, and proofing services. On the technological side, the industry has become reasonably proficient with SGML and XML over the last ten years. Of course, XML is now trendy and mass-market; even Microsoft appears to support it. It rather takes the fun out of it for those of us who were into structured mark-up when it was a radical new idea. Printing costs are down over the decade, primarily due to improvements in prepress technology. Mailing costs are up over the decade, an incentive to drop paper. Overall, these developments, along with the use of author's better prepared files, have combined to reduce conventional publishing costs.

7.   Subscriptions and Publishing

After adjusting for inflation and the increase in pages, the subscription price for the *ApJ* has stayed almost constant over the last twenty years. It is no secret that the AAS journals are run on a cost-recovery basis. Given that the ratio of page charges to subscriptions income has stayed roughly the same for many years, that all journals have seen a decline in number of institutional subscribers, and that all publishers have incurred increased costs for electronic publishing, you can safely conclude that there have been real reductions in production costs. From the subscriber's point of view, of course, the problem is still the top line: the growth of the size and cost of journals in current dollars.

The discussion of printing and mailing costs leads naturally to the question about the demise of print editions. At the National Academy of Sciences E-Journal Summit in January of this year, there was a lot of discussion on this point. For several years, publishers have thought that this year would be the year that libraries started switching from paper to electronic in significant numbers. It still hasn't happened. However, there is some thought that American public universities who are suffering from severe budget cuts, might finally be tempted to drop paper. The AAS has been offering quite a large discount for electronic only subscriptions: about a third. The 2002 rate for the combined *ApJ* and *ApJS* was $1200 for electronic access only; not bad for about 25,000 printed pages. The take up on this offer has been small but increasing. My colleagues at Chicago have asked me to ask you what your plans are for future paper or electronic decisions.

For individual subscribers, the AAS situation is interesting and atypical: After the entire *ApJ* went on line in 1997, a significant portion of the individual print subscribers switched to electronic subscriptions immediately. In subsequent years, there has been a continuing gradual decline in member subscriptions of any type (paper or electronic) — presumably as members have access through their institutional subscriptions. In the past year alone, the AAS lost between 10 and 15 percent of its individual member subscriptions. We have not seen or heard of the same kind of dramatic shifts in other society publications.

8.   Online Products

Moving now to the actual online products, there has been surprisingly little change in the basic product since the first web online journals in 1995. In 1995, we had HTML, PDF, and reference links to ADS (or PubMed) with the web presentation generated from full text SGML. In 2002, we still have HTML, PDF, and reference links to ADS. We have added machine-readable data tables and a small amount of multimedia; some disciplines have done more in this area than has astronomy. The limitations of the HTML product that were obvious in 1995 haven't really been solved yet: We still don't have robust support for math and special characters at the browser level, and there are still significant incompatibilities between browsers and browser versions. The STIX project to develop a comprehensive set of Unicode symbol fonts for scientific publications may finally deliver in the next year or so. With STIX fonts, XML, and style sheets, we could finally create a really good looking browser-based product. At a purely technological level, this is the project that interests me most in the next few years.

For amusement, we dug out the HTML from the *ApJ Letters* prototype from July 1995 and put it back up on a server. Everything still worked, though I didn't have a pre-version 4 browser available to recreate the exact look and feel of 1995. Around 1998, we regenerated the early *ApJ Letters* articles from the SGML source to a different HTML design that more closely imitated the paper journal design more closely—a bad idea as it turned out. Other changes included use of HTML tables, and somewhat more liberal use of graphics. Our current HTML design no longer attempts to mimic the appearance of the print

so closely. Aside from the use of HTML frames, the feature set is much the same as in 1995.

9.   New Online Services from the AAS

We do have some new product developments from the AAS to announce. Beginning July 1, 2002, the *ApJ* and *ApJS* will be posting the accepted versions of papers as preprints, with the consent of the authors. The preprint is the PDF of the last version of the paper, the one that the editor accepted for publication. After editing, typesetting, and proofing, the final versions will replace the preprints on the site. The same Digital Object Identifier will be used throughout. A journal subscription is required to access the preprints, but the email alerting service will be free. The AAS thought that this service might be useful to some parts of the community; we will all be interested to see what the usage actually turns out to be. As the system is only going live this week, it will take a while before authors are aware of this new service and have granted consent for their accepted papers to appear in it. By September, this should begin to be interesting.

Some other projects underway for the AAS:

- DOI citation links via CrossRef (low priority because of utility of ADS linking)

- Funding source / instrumentation keywords (AAS has committed to working on this with interested parties)

- STIX / XML version of journals (my particular favorite project for the next few years)

- Data center linking and National Virtual Observatory (a major long-term interest of everyone in the discipline); I'll return to this later on.

10.   Reader Behavior: Links and Linking

After the product, the next step is the reader and what the readers are doing with all this stuff that we are creating and making available.

I can't pretend to answer that question definitively, but I have brought along a few statistics that might be of interest. Here are some numbers that indicate movement to and from our site for the month of April 2002 (all numbers are rounded):

- Links to UCP: PubMed 61K, Google 48K, ADS 47K, Yahoo 28K, DOI links 2K, JSTOR 0.5K

- Links from UCP: JSTOR 9K, ADS 7K, PubMed 5K

It would appear that many of the bells and whistles that publishers are adding to their sites aren't being used extensively. It also appears that a significant portion of reader traffic originates at secondary services; possibly as much as 1/3 for medical journals and 1/2 for astronomy journals. (Don't quote me

on that yet.) We also know that a significant portion of readers still download the PDF of most articles that they look at, presumably printing and reading offline. This is consistent with results of a survey by Boyce and Tenepir, that a great majority of users of online journals print out anything in which they are seriously interested.

Despite evidence that readers don't actually read online yet, the publishing industry is working together to provide more facilities for online research . . . And to add value to the electronic editions. The CrossRef project is about creating direct links from citations in reference lists to articles on other publishers sites using the Digital Object Identifer system (DOI). You can read more about it at www.crossref.org (`http://www.crossref.org`). Of course, ADS has provided this functionality for years for astronomy at the small cost of an extra click or two. This cross-publisher linking is very important in disciplines that don't have an ADS equivalent or for cross-disciplinary research. This feature will begin to appear in AAS journals later this year as we implement it for all our entire site. The CrossRef system is very new and usage is still small, but growing quickly.

CrossRef has another big project in planning stage: a cross-publisher, cross-discipline full-text search and linking system. I am currently on the board of directors of CrossRef and would be happy to talk to anyone about CrossRef off line.

11.   Archiving Electronic Content

The last stage of the process is what happens to all this content in the long run. Unfortunately, this is a difficult problem and no grand solutions have appeared since I spoke about this at LISA III. The Andrew W. Mellon foundation funded a series of planning grants on e-journal archiving that were to lead to large scale implementations. Chicago participated with Wiley and Blackwell in a project led by Harvard University Library that proposed to tackle the problem of archiving all the electronic content including the SGML / XML source with a high level of richness preserved. Unfortunately the Mellon foundation has chosen not to go ahead with the implementation. Some of the work has been salvaged, however; the National Library of Medicine is funding the further development of an XML DTD for archiving journal content. Another interesting project is LOCKSS, out of Stanford University Library (`http://lockss.stanford.edu`), that takes the opposite approach: instead of trying to archive the publishing raw materials, it tries to make many copies of the web sites. This is a clever low cost, short term solution. Clearly there is a lot more to be done in this area, particularly with electronic only content.

12.   Revolution Over or Just Beginning?

So where are we now, seven years into e-journals delivered online via the web? Is there a revolution yet? Is all this electronic publishing just "distributed printing"? Is our expensive e-publishing infrastructure under utilized? Will electronic publishing ever live up to its potential? When?

My answers to those questions are No; Yes; Unfortunately Yes; I certainly hope so; Not until we actually read and work on screen! That may take a

generational change in computer hardware and software and the users of the hardware and software.

There are some scattered signs of revolutions out there, but most of the innovation is not coming from traditional publishers. It is my personal opinion that the next round of real innovation in electronic publishing will come in discipline-specific ways that will be about not formats but content. Format will follow content, as it were. It will not be merely repackaging an existing mode of communication — the scholarly article — and delivering it on a new platform.

I have a one very modest example to describe from the journal *Current Anthropology*, a journal where the editors are eager to experiment with the new media. Many years ago, this journal pioneered the practice of publishing articles accompanied by invited replies and commentaries . . . a technique adopted in the electronic media by Stevan Harnard's journal *Psycoloquy*. The editors decided that what they wanted to do in the electronic world, was to provide supplemental cultural materials to enrich and illuminate the text of the article. An example of these "enhanced" articles is "Ghost Dancing the Grand Canyon: Southern Paiute Rock Art, Ceremony, and Cultural Landscapes" by Richard W. Stoffle et al (*Current Anthopology*, volume 41, no. 1) (`http://journals.uchicago.edu/CA/journal/`). In the electronic version of this article, there are additional images of Indian Rock Art, sound clips of oral histories, photo essays, maps, and sound clips of music. In effect, this is the content of a small multimedia CD-ROM along side an article. The disadvantage of this sort of publication is that it has all the production costs of a small multimedia CD-ROM. The advantage is that it is of interest to scholarly and lay readers and that it makes great teaching material.

I'm not saying that this is an answer for astronomy; to the contrary, astronomy will have its own discipline-specific innovations. And where will those innovations come from? I don't know. I'm not an astronomer . . . only a publisher. But I am fascinated by the work of the data centers and information service providers in astronomy. And I'm excited by the potential of the National Virtual Observatory and similar projects to encourage new forms of scientific communication. At the moment though, I'm still waiting for the revolution.