4. ARTICLES

ELECTRONIC JOURNALS: OBSERVATIONS BASED ON ACTUAL TRIALS, 1987-PRESENT

Sandra L. Arlinghaus and Richard H. Zander.*

ABSTRACT

Electronic journals offer a 21st-century forum for the interchange of scholarly ideas. They are inexpensive, fast, easy to store, easy to search, and they have long-term archivability; these advantages easily justify the time spent learning to deal with the new technology. The authors, both editors of nationally-noted electronic journals, share with others their interdisciplinary experiences in dealing with this new medium for producing online, refereed journals.

During the past six years each of us has created and edited a successful electronic journal (E-journal) in our respective fields of geography (*Solstice: An Electronic Journal of Geography and Mathematics* first appeared in June of 1990) (Palca 1991; Peterson 1992) and botany (*Flora Online* first appeared in January of 1987) (Palca 1991). Both journals are peer-reviewed; both are available, free, over standard computer networks; and, both have editors who served as authors in early issues–to get the journal off the ground. E-journals provide an opportunity to share computerized information with others in an orderly and responsible fashion, within the context of current technology. They offer:

1. An inexpensive way to share information, quickly, with a large number of individuals;

   a. As direct, online, transmissions from editor to individual; in this case, the transmission should be free of charge, in much the way that a library card is free of charge. The editor/publisher bears the cost of journal creation and manufacture; the reader bears the cost of maintaining on online mail box;

   b. As direct transmissions to libraries – libraries should pay for diskettes, hard copy, online transmission, or whatever they desire. The cost to the library is generally greatly reduced from that of conventional journals, thereby freeing library funds for other useful projects. Funds generated from this source may make the E-journal(s) self-sustaining;

   c. As posted "messages" on an electronic bulletin board or files on an "anonymous FTP" server. The reader bears the cost of accessing the board or server and downloading the article.

2. When E-journals are highly specialized, they can serve as a more formal alternative to large (archived) data banks in the natural sciences and elsewhere. Indeed, when the E-journal is downloaded into a wordprocessor or a data manager, the content can be manipulated and edited carefully to fit the research needs of the individual user.

There are many systematic electronic communications already available and there are apparently more in the planning stages. The first edition of Michael Strangelove's "Directory of Electronic Journals and Newsletters" (1991) catalogues about 30 journals and over 60 newsletters. Major academic societies, notably the American Mathematical Society and the

American Association for the Advancement of Science, have announced far-reaching plans to produce other electronic journals: (Janusz 1991; Palca, 1991: 1480). A glance at a flyer for the Annual Meeting for the Society for Scholarly Publishing (July 1992) suggests that more than half of the four-day meetings will be devoted to issues related to electronic publication.

There are:

1. "Genuine" electronic journals.
2. Mere computerized versions of hardcopy titles.
3. Non-archived electronic databases that are not really citable in a scientific paper since the data used may have been changed or may no longer be available, even though these databases may be copyrighted.

What makes a systematic electronic communication a "journal" is a difficult issue (Nicholson 1992); concern for rigid, *a priori*, definition might better be replaced with open regard for all entries and suitable concern for the broad issues of journal production. For, an E-journal is first and foremost a "journal" that has simply been **modified** as "electronic," both linguistically and technologically, by the method of its transmission and production.

Thus, we offer a generalized summary of observations that have come from six years of actual trials with *Flora Online* and three years of actual trials with *Solstice*. It is useful to separate these results into three broad categories: content issues, production issues, and archival issues.

Content issues.

The most important concern is to obtain good manuscripts. And, to be acceptable as an outlet for scholarly publication, E-journals should approximate standard formats for professional journals, have high standards of scholarship, and be refereed. It does not matter how sophisticated the technological production becomes; if the journal does not have interesting and useful material of high quality, it will fail. This point should be obvious; however, it can become obscured, particularly in light of the exciting capability of the computerized format.

Thus, author perceptions of E-journals are critical; the most serious problem involves citation. Will others see the work? Will the work be taken seriously? The following strategies help:

1. the editor should see to it that the E-journal (and when necessary hard copy derived from it) is listed, housed, or otherwise recognized in

   a. standard reviews that are specific to the discipline of the journal;

   b. the usual indexing services (publications are often judged by the bibliographic and citation services that mention them — services that accept electronic files are particularly easy to deal with);

   c. news media, including field-specific conferences and meetings as well as mass media;

   d. standard book/journal registers of documents using conventional book/journal codes (such as ISBN and ISSN); and,

   e. library archives. Libraries apparently dislike the idea of downloading journals; they appreciate diskettes mailed to them. Archiving is important for E-journals so that data can be retrieved long after publication.

2. The editor should consider the unusual to boost regard and readership for this mode of journal transmission, such as:

   a. the use of reprints (with appropriate copyright permission) of hard-to-find works of field-leaders (prospective authors—of lesser fame—usually perceive some benefit-by-association and field leaders often are interested in participating in a different venture);

   b. the use of interactive review of material — post-publication review followed by online alteration of the original document as a later version (coded appropriately–original is version 1.0 and updates carry larger numbers according to the extent of change);

   c. the use of taxonomic, bibliographic, and other data sets consisting of long lists of records that can easily be downloaded and sorted according to user need. Several agencies are preparing monolithic data banks from which scientists can extract items of information using specialized data management programs. Unfortunately, such data banks usually employ in each different management system, complex and difficult for the scientist to learn, and the data banks give second-hand data (digested by those who run the data bank and who are not necessarily scientists). With the advent, however, of electronic publishing, information in the sciences developed by individual scientists can now be easily and directly shared;

   d. the use of novel typesetting or other electronic capabilities that display the power of the vehicle of transmission (Horstmann 1991); and,

   e. the sharing of experiences in E-journal editorship with others — through professional associations directly promoting electronic journal editorship (such as an E-journal editor's association) and with other organizations indirectly promoting it (such as the TEX Users' Group; "TEX" is a trademark of The American Mathematical Society).

Readers who are initial skeptics can become more receptive when they see actual output; hence, the early need for editor to become author. To increase E-journal availability, and to convert a wide variety of skeptics, E-journals should be distributed in more than one manner (e.g. diskette, File Transfer Protocol (FTP), Bitnet, on a listserv, U.S. mail, hardcopy).

When editor becomes author, then a mechanism for review is all the more important. Pre-publication peer-review by an editorial board or by other colleagues is effective and easy to achieve electronically; post-publication feedback in an open or closed forum is also simple electronically. In addition, it is important that the editor continue to publish in various other outlets held in high regard.

There are also a number of other reasonable, but less important, concerns that authors might have. These include:

1. Manuscript security; because E-journals can be forwarded easily, alteration of original manuscripts can occur. There are a number of ways to deal with this problem:

   a. Copyright a hard copy of the original transmission (thereby placing it in the Library of Congress);

   b. Advertise that the original computerized version or a hard copy of the original transmission is available (on-demand) to those wishing it–including libraries;

   c. Store single hard copies in selected libraries (including that of the author's institution);

   d. Transmit E-copies directly from editor to individuals, over standard electronic networks, using an electronic distribution list automatically marked with the sender's name

and time:

e. Download from an electronic bulletin board. A persistent worry here is that a file made available for downloading is not "published" in the sense of being distributed. This worry underscores, again, the need for adequate reviewing and indexing of the document. However, the prospective author should note that a file made available for downloading is in fact published because

i. this is the same way hardcopy books are published — they are simply advertised as available for purchase, and

ii. in bibliographic research, the date of publication is the date advertised as available, since it is impossible to track down the date of first purchase or first mailing of the book.

f. Copy-protect diskettes (using some sort of seal unique to the journal) to prevent unthinking abuse.

2. Virus and other crank programming prevention. Downloaded FTP or regular phone modem files from other computers can spread electronic viruses if they are "executable," and only if they are actually run as programs. Downloaded text files cannot spread viruses: downloaded executable files (.EXE, .COM in MS-DOS) can be examined by commercial programs for viruses before they are run. When the E-journal is made available through a network server, the E-journal's health is simply transferred elsewhere; the network supervisor has considerable responsibility in this regard. Of course, good backup habits and a procedure in place for dealing with viruses if they happen are a must in all workplaces that use programs obtained from outside the workplace.

Production issues.

Production issues generally appear to fall into one of two categories: Document manufacture and editing, and transmission. Warehousing is not an issue of any significance, nor is the sort of marketing that requires a network of publishers' representatives to sell hardcopy documents.

Document manufacture and editing.

The manufacture involves creating, or being supplied, electronic files. Editing at this stage in journal computerization generally requires in-house manufacture and distribution of files and their media. It is useful to aim for the lowest common denominator: currently, that means ASCII text and .GIF or .PCX graphics files if needed — such files are easily read on a IBM PC clone, a Mackintosh, or Unix machine (Xwindows or whatever), by any wordprocessor and most graphic file viewers. It would be nice if the files could be set up with the format of one of the new GUI wordprocessors (e.g. WordPerfect, MSWord) but it seems prudent to wait until a multiplatform wordprocessor that creates text files incorporating graphics images becomes commonly used. Most prospective authors can provide "manufacture-ready" copy in the form of an ASCII file sent over the e-mail or provided on diskette. Indeed, for MS-DOS environments DCA or RTF (Document Content Architecture or Rich Text Format) are also standard file formats retaining formatting commands; these may be used to transfer a formatted text to any of most commercially available major word processors. It is thus an easy matter to ship the E-file to referees and to provide authors with E-proof to check prior to final production.

There are a number of issues, found also in conventional publishing, that remain difficult. For this reason (also), it is useful for editors to be experienced as authors of conventional articles; it is additionally desirable for them to have had editorial experience in dealing with a conventional publisher.

1. When the ASCII file is typeset using TEX, mathematical notation, tables, and figures that are rectilinear in shape are easy to handle: otherwise, complex mathematical notation is difficult even to approximate in ASCII. The typeset TEX file is itself an ASCII file with ASCII formatting commands, and so can be transmitted easily.

   a. The computerized typeset TEX file is not strictly "what-you- see-is-what-you-get"; however, the file is of traditional quality typesetting, and the file of electronic text and notation can now be downloaded and cheaply typeset or printed in hard copy by the journal receiver at his or her expense. To typeset the file, the receiver must first convert the transmitted TEX file to a .DVI file and then print it on any available downloading device (such as a Xerox 9700 series machine or an APS phototypesetter).

   b. The receiver can view the transmitted TEX file on screen (with the TEX commands visible). The editor can right-justify the TEX file in a word processor (prior to transmission), and bitstrip it to retain it as an ASCII file, in order to produce a journal-like electronic page in the transmitted E-file without interfering with (or influencing) the typesetting of the hardcopy. Right-justified electronic copy tends to reduce the visual impact of the unnatural looking typesetting commands that appear in the TEX file as it is viewed online.

   c. TEX produces device independent files; however, because different installations of TEX support different features it is good, at present at least, to keep the typesetting simple. To this end, the editor should consider supplying a set of TEX macros to authors wishing to do their own typesetting using TEX; these can be supplied over the electronic mail in much the way that the American Mathematical Society encourages the submission of abstracts for its meetings.

   d. Not all individuals have access to TEX even though their university has it; individuals in mathematics departments generally do have access to it and know how to use it.

   e. Figures, charts, and tables that can be considered as a matrix (such as a crossword puzzle) can be typeset using TEX. Maps and non-rectilinear figures generally cannot.

   f. One approach to dealing with figures, that works easily, is to scan complicated maps and figures and to incorporate the scanned file into any distributed hardcopy by electronic cutting and pasting. The Xerox DocuTech stores scanned images as electronic files on a hard disk and permits such electronic editing. Hardcopy, complete with figures, can be produced in an on-demand fashion for sale to standing orders and to others who inquire. Warehousing is thus converted to a "just-in-time" approach requiring virtually no extra space or cost. Hard copies can then be made available in a variety of bindings.

   g. If the scanned electronic files are downloaded as part of a text file, then the reader's electronic cutting and pasting is unnecessary. The capability of future word processors holds the answer to the possibility of shipping mathematical notation, maps, and photos in a single easy-to-read, typeset, transmission.

   h. Graphics transmission can be executed immediately by making available for distribution binary files of graphics images on an Internet server for downloading via FTP (File

Transfer Protocol) or from a standard bulletin board.

i. Yet another approach to the graphics issue might involve linkage to a Geographic Information System to provide a procedure for creating compatible transmittable map files directly from data managers into a TEX-ed file. Data files are likely to be quite large; compressed files should be used with instructions for decompression and recompression provided online in "help files."

2. As above, TEX can be used to create an ASCII file that is typeset, including diacritical marks. If, however, the editor chooses not to use TEX, publishers can convert the formatting codes of other software such as Microsoft Word, XyWrite (Signature), and other robust word processors. If straight ASCII, perhaps employing the upper IBM ASCII set whenever diacritical marks are important, is used to transmit the electronic files, then another set of issues, some similar to and some different from using TEX, confront the editor.

a. At present it is important never to right-justify straight ASCII files. Right-justified text introduces extra spaces in word processors that produce straight ASCII files. To mend this, users must do a number of search-and-replaces, replacing double spaces with single spaces. They need to do this to make the text look like their own text so they can add items from a bibliography to their own bibliographies or add to other downloaded lists of subjects that are searchable with a word processor or data manager.

b. Data-intensive text files, either those for which it is difficult to find a publisher in hardcopy or, in particular, those that are suited to searching and other computer text manipulation (such as bibliographies or checklists), are well-suited to journals employing the straight ASCII format. Data files take two forms: article format, similar to paper publications – searchable with a standard word processor or with "text management" software, and, data base format, appropriate for importing into a standard data base manager. The latter should have data presented with an equal number of lines per record and information entered on the appropriate line for each field, or in another "delimited" format.

c. Large text files should be divided into smaller files each less than 300 kb in size. These can be uploaded as is, or first converted into smaller compressed (e.g., .ARC, .ZIP, or .LZH) files. Split text files can be downloaded and reconnected (through DOS copy command) by the user. Very large files may, for now, be more appropriately distributed on disk.

d. Foreign language characters, symbols, and graphics. Authors should expect that downloaders will generally use 8 data bits and an error-checking protocol, so binary files and text files with the IBM upper ASCII character set (foreign and special characters and graphics) can be easily transmitted. If the text is prepared in something other than a MS-DOS, pure ASCII environment (non-ASCII texts are created by many word processors), authors need to remove all software-specific formatting codes and type-style codes, before uploading. These can, however, be suggested — underlining codes, for example, might be represented by symbols like @ or | so downloaders can re-underline through search-and-replace.

Users of operating systems other than MS-DOS generally do not have access to the upper IBM ASCII set, which has foreign characters and symbols such as the degree sign

(°) and simple graphics. Also, because all users may not have MS-DOS microcomputers or compatibles, some authors may wish to substitute special codes for the IBM upper ASCII set used in MS-DOS. It is recommended that instructions for translating (by search and replace) the codes into the actual character be given at the beginning of the publication. Any system can be used; however, a simple system, which can be easily interpreted even before translation and may be easily used by non-MS-DOS systems, is the "backspace and overstrike" method: many foreign characters may be easily manufactured by causing the printer to backspace and overstrike a diacritical mark. Since some wordprocessors cannot deal with the ASCII backspace character (ASCII 8), substituting an unused lower ASCII character such as @ or | for the backspace character will allow search and replace for (1) the backspace character itself, (2) for an acceptable printer code substitute for it, or (3) replacement of the three characters with an IBM upper ASCII character. Examples of backspace substitution: a|' = à , A|o = Å, u|" = ü; and of direct substitution: deg. = °, u = μ (search for space-u-space and replace with μ). Graphics characters have little utility and cannot easily be coded for non-MS-DOS standard machines, so it is recommended that these be restricted to special applications.

There are a number of efforts at an enlarged ASCII set for foreign languages (Hayes 1992). The coming of Unicode or something similar will hopefully provide a complete set of multiplatform foreign characters.

e. In bibliographies, spell out all duplicate author's names (do not use a sequence of hyphens.) so that the author's names can be searched for. Begin each entry flush left and leave an empty line (two hard rights) between each entry.

f. Do not spell any words with all capital letters (this may make it difficult to search for them; it also looks bad).

g. If appropriate, present files in a "squeezed" form as an .ARC file or .ZIP or another 'archiving' utility file format. This allows faster and less costly downloading and keeps diskette files small.

3. File management seems to be relatively easy with an E-journal. Keeping track of manuscripts, and of who is refereeing them, and of their stage in the production process, is made simpler by the technology.

Transmission.

E-journals should have standard, and thus easy-to-use, modes of access. They should be transferable across different systems (e.g. various micro, mini and mainframe platforms). Alphabets should be standard (ASCII, ISO Superalphabet eventually) in order to be available to a wide number of users. Transmission can occur in a number of different ways and have various uses.

1. Issues may be obtained by "anonymous FTP" or downloaded via regular telephone lines by modem from an electronic bulletin board. An electronic bulletin board system is a computer and software system that can be accessed from outside by a caller, who likely has a number of options, including perhaps:

   a. Reading or leaving messages. These are typed while online and may be public or private (readable only by the addressee).

   b. Depositing or taking away data or text files. These are created with a word processor or data manager previous to calling and are "up-" or "down-loaded" as a unit.

c. Extracting information from a large data file. Authors can prepare compiliative publications that they use personally and wish to share. Then they may, if they wish, maintain the publications informally or formally as a series of versions in online data banks. Users of the bulletin board download online files, and use the files directly for searching for particular data or by copying portions to enlarge their own personal files, with due respect, of course, for copyright privileges of the original author.

2. A bulletin board can be of interest to scholars in the following ways:

   a. Messages - For exchange of ideas and information. Speed of contact is far greater than with regular mail. Special "Conference" sections allow public exchanges on single scientific topics that are equivalent to symposia at national meetings.

   b. Files — Electronic publications that may be cited in an author's curriculum vita. Such publications should be copyrighted. These include: original text material and computer programs; text or data files of an ephemeral or informal nature; and, previously published computer programs (of "reprint" value). With the eventual realization of a network of bulletin boards across the country, this method of transmission holds considerable promise.

3. Ship the E-journal across Bitnet or Internet to a distribution list of subscribers who ask to have the E-journal mailed to them. Some installations do not have the capacity to send files in excess of 25,000 characters. In that case, split the journal apart with instructions to the user to concatenate the files prior to downloading, printing, or typesetting.

Archival issues.

All journals are useful only for as long as they can be located in the holdings of some institution. As technological formats for producing journals change, it will be important to keep not only the new, but also the old — as back-up with a known life-span. Some of the issues that will confront archivists include those listed below.

1. Availability — the E-journal should be archived indefinitely in an institution willing to provide copies or the equivalent on request.
2. Durability — Archives should be maintained so as not to degrade with time, e.g. contents of diskette transferred to hard disk, then to optical disk, then to solid state or whatever future technology provides. Duplicates stored off-site, and EMF protection are also advisable in the long-term. Paper burns and degrades with age, but magnetic images can be maintained indefinitely if copied periodically onto new media (diskettes are said to have a maximum data retention life of 10-15 years).
3. Retrievability and salvageability - Standard operating system formats should be changed in a timely fashion: MS-DOS to Unix, etc. Standard word processing formats should be upgraded so they can be read decades hence. Database formats should be standard or also available in ASCII-delimited format. Any required programs (decompression programs, graphics viewing programs, special word processors) should be archived, too, along with necessary hardware platforms.

We have found that editorial and publishing problems can be overcome within the limits of existing technology such that electronic journals can be successful in transmitting and presenting information to scholarly readers. We foresee a significant upgrade in quality and

flexibility of electronic presentations with the advent of standard cross-platform graphics-capable word processors, standard export-import formats, and standard multi-language character sets. The advantages of electronic publication: inexpensive, fast, easy to store, easy to search, long-term archivability, easily justify the time spent learning to deal with the new technology.

References.

Hayes, Frank. 1992. Superalphabet compromise is best of two worlds. *UnixWorld*, January 1992: 99-100.

Horstmann, Cay S. 1991. Automatic conversion from a scientific word processor to TEX. *TUGBoat: The Communications of the TEX Users Group* 12:471-478.

Janusz, Gerald J. 1991. Reviewing at Mathematical Reviews. *Notices of the American Mathematical Society*, 38:789-791.

Knuth, Donald E. 1984. *The TeXBook*, Reading, MA: Addison-Wesley and Providence, RI: The American Mathematical Society.

Nicholson, Richard S. 1992. Data make the difference. *Science News*, March 28:195.

Palca, Joseph. 1991. Briefing. *Science*, November 29: 1291.

Palca, Joseph. 1991. New journal will publish without paper. *Science*, September 27:1480.

Peterson, Ivars. 1992. Math for all seasons. *Science News*, January 25:61.

Strangelove, Michael. 1991. *Directory of Electronic Journals and Newsletters*. Washington D.C.: Association of Research Libraries.

* Sandra L. Arlinghaus, Institute of Mathematical Geography, 2790 Briarcliff, Ann Arbor, MI 48105. Richard H. Zander, Buffalo Museum of Science, 1020 Humboldt Parkway, Buffalo, NY 14211.