BEYOND THE CUTTING EDGE, ELECTRONIC PUBLICATIONS IN THE 21ST CENTURY

Jennifer Pattison Rumford and William R. Riedel

ABSTRACT

The Ocean Drilling Program (ODP) has amassed large databases (including paleontological) as a result of its research on the geology of the ocean floor. ODP has begun to present its research results in electronic proceedings that will soon completely displace paper publications. **Palaeontologia Electronica** (PE) has pioneered the concept of an all-electronic paleontological journal with no paper equivalent.

The publication and distribution of electronic data and scholarly papers are no longer the cutting edge of publishing technology; the cutting edge now moves to the next generation of electronic documents. The wide experience gathered by ODP and PE, from the collection of drilled cores and data, through preliminary reports to final scholarly publication, has provided the historical background. The next generation of electronic publications will include refinement of some of the present elements supported by technological improvements to create a product that meets the full spectrum of needs within the scientific community.

The goals for these publications should include the creation of a truly dynamic platform for the exchange of information and ideas; the permanent, long-term archive of published materials and the data on which they are based; improved world-wide access; and a realistic and improved presentation of text, data, graphics, and other media.

Jennifer Pattison Rumford, Ocean Drilling Program, TAMU, College Station, TX 77845-9547
William R. Riedel, Scripps Institution of Oceanography, UCSD, La Jolla, CA 92093-0220.

Keywords: electronic publications, future, world wide web, publishing, CD-ROM

Submission: 1 October 1998, Acceptance: 22 February 1999

INTRODUCTION

The development of a communal hypermedia world of electronic publications is no longer a matter for debate. It exists. How successfully this medium surpasses the world of traditional print journals (that remain static with frozen information), depends on how authors and readers make use of electronic documents, and what tools are available for the creation and use of hypermedia documents.

There can be no doubt that electronic publications will be part of scholarly life in the 21st

century, but there are serious drawbacks to electronic publishing with the technologies now available. In particular, the move from a printed document to a web-based electronic document entails a number of inconveniences. Paper holds more information in a given area than a monitor can, and because of the nature of electricity and the technologies of cathode ray tubes and liquid crystals. Moreover, a reader cannot process up to 40% of the information presented on current computer monitors. These same readers also find it slower to read on-screen than from paper (Tufte 1991).

More serious impediments to the advance of on-line scholarship can be found with potential authors, though. Older, established authors are, in general, wedded to traditional paper publication channels, and want assurance that their work will be safe from plagiarism. Young authors are concerned with the need for maximum credit (ideally a job, and academic advancement) for their publications. All authors are concerned with the overall reliability of the on-line material (Harnad, 1998).

However, with the explosive growth of the Internet and, specifically the World Wide Web (WWW), during the past five years it is obvious that there is no turning back. Within the U.S., funding organizations, such as the National Science Foundation (NSF), are actively promoting electronic-over-paper distribution of information. The Digital Libraries Initiative, sponsored by the NSF and five other agencies is an attempt to develop the next generation of digital libraries and the applications used therein.

Tracking Readers' Use

The automatic collection of data on a readers' use of the individual files constituting an electronic publication can yield information of substantial value to publishers and to authors.

This information can include who accesses the electronic publication files (by country and organization) and when the file is accessed (by date and time of day). By using this information, the publisher can determine how well various documents, and individual parts of those documents, fulfill marketing goals.

Although web statistics offer an amazing amount of information, it is not the same information as found in a citation index, and should not be treated as such. The tool used for collecting web statistics, including tracking access, must be robust enough to delineate WWW users stumbling into a site, and bouncing right out, and serious readers, who arrive at the site, read and digest material. These users will probably comes back to the site multiple times, and may eventually refer to the published material in future publications. The information collected with such a system is acutally much superior to what is found in a present-day citation index.

For the technical specialist, who may or may not be an editor, readying documents for web publication, reports generated on previous traffic to the site will be invaluable as they can provide information as to how the reader came to the document (what the entry and exit points were), and what path the reader took through the document (how the document was read). Such information can help eliminate the stumbling blocks that

can be encountered in navigating through a non-linear document. Also, the technical specialist can help an author to ensure that readers readily encounter those elements of a document deemed to be most crucial. At this point, Internet browsers have different capabilities. User statistics can determine what browser is most commonly used used to access files in an electronic publication series. They can also aid in determining how best to prepare the files. These statistics can also be used to determine what files contain the most, and least, accessed information. The words used in searches can also be identiifed to determine what readers are looking for. Further downstream, authors can use a similar methodology to ensure that they get their message across to readers effectively.

In order to accomplish these desiderata, electronic documents should consist of many small files rather than fewer, larger ones, so that details of their use can be better tracked. Smaller files are also faster to load, and they also facilitate targeting a specific piece of text or graphic during a searching or indexing procedure.

The software that tracks file usage on the network should gather data in as much detail as feasible - even to the timed progress of an individual user through the files constituting a document. A consequence of this is that we will need a kind of meta-software, to help us find what we are looking for in this mass of detail.

LIFE-SPAN AND GROWTH: LIVE INFORMATION

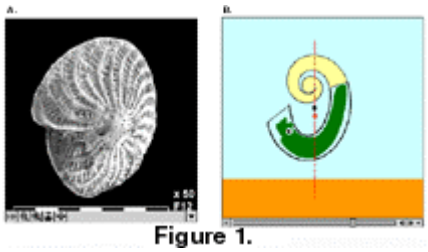

Figure 1.

In the area of scholarly publishing, electronic books and journals have already adopted a number of enhancements made possible by global access to computers, the Internet, and CD-ROM readers. **Palaeontologia Electronica** exploits non-linear access to various parts of its publication, unlimited use of color, removal of limitations on article size and number of graphics, the use of movie clips (Figure 1), and the linking of references (backward) to earlier abstracts and full-text articles on the WWW (Figure 2).

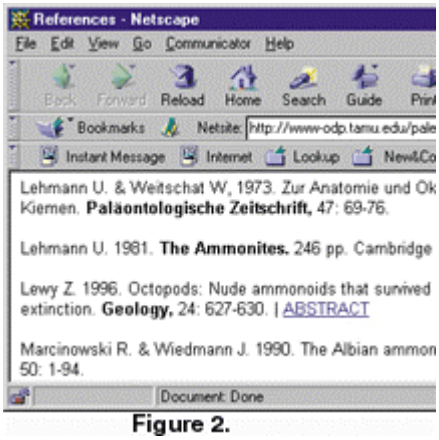

Figure 2.

One aspect of fully electronic publication with no printed equivalent, that will certainly develop early in the next century is the addition of dynamic elements to traditionally static articles. An obvious example is the provision of a discussion space that permits readers to attach comments or elaborations, thus bringing added value to an article. Such a discussion space could remain open for as long as the article resides on the Internet (Figure 3).

Another development, which is foreshadowed by the paper-based **Science Citation Index**, will be forward referencing. Forward references provide links to citations that appear subsequent to an article's publication. These would enable a reader to track a topic forward in time, through later publications that cite the original article, and are being incorporated into journals such as PE as its articles are cited. **The Astrophysical Journal**, for example, makes extensive use of forward referencing.

The Ocean Drilling Program is an unusually tightly integrated, clearly circumscribed program, covering a broad spectrum of the geological sciences. As such, it provides an ideal environment for the handling and dissemination of information. If this opportunity is fully grasped, ODP will become a model for other programs and disciplines as they develope electronic publications. Key components of this well-structured information flow at ODP have already developed-- the **Initial Reports** (IR) and **Scientific Results** (SR) volumes, and the JANUS database. The transition from paper to electronic publication, now underway, invites closer integration of the existing components, which can be viewed as a succession.

Figure 3.

1. Observations made, and data collected (JANUS database
2. IR volumes (description and preliminary interpretation)
3. SR volumes (further interpretation and preliminary synthesis)
4. Continual development in the scientific literature - continuing publication within the ODP structure (see The Issue of Archive)

5. Beyond to outside journals, books, etc.

As an example of the utility of such an integrated system, electronic links could be provided from the IRs to corresponding elements of the JANUS database. From the SR volumes, links might be provided back to the corresponding parts of IRs, and, if necessary, to additional JANUS elements, as well as forward to articles in outside journals. Through these links, readers would be able to follow the development of interpretations and syntheses, and to see easily the data and observations on which they are based. This process is essential for critical science, but is presently time-consuming, and therefore too rarely done. This is because at present, the described procedure involves using several different paper publications and difficult-to-access databases. The linking of electronic publications and databases will thus be invaluable to conscientious research.

Stanley Chodorow writes, "In medieval intellectual culture, works of literature - history, theology, law, medicine, and literature in the strict sense - flowed from author to author, across generations, growing and changing as individual contributors worked on them" (Chodorow, 1997). This was true both figuratively and literally, as a scholar might scrape a parchment sheet clean of ink and replace parts of a document. The physical reality of a document on the WWW allows it to be collaborated on, annotated, and shared through time in much the same way. Successive versions can be tracked and cited, because of the stability of file names and contents, and identification of commentator or reviser.

MIXED FORMATS: WHY THEY ARE NEEDED

The beauty of the HyperText Markup Language (HTML) (or the up-and-coming Extensible Markup Language [XML]) lies in the rapid access, painless browsing, and search capabilities within a set of files. The format is more generic than any file except a bare-bones American Standard Code for Information Interchange (ASCII) file. This is a positive aspect indicating that future format migration will be possible. The indication is hopeful, but there is no accompanying written guarantee.

Format migration is one of the necessary components an electronic publishing specialist must take into consideration, because unlike printed documents which have been able to depend on the reliability of printed language to span generations of readers[1], electronic files are rapidly outdated by the changes within software and hardware systems.

Use of HTML/XML as an electronic publication tool is not enough, though. It must be supplemented with a supporting library of Acrobat Portable Document Format (PDF) files (see below). When combined, these two modes of publication deliver what readers really need: speed for accessibility and browsing, non-linear presentation of information, search capabilities, and a traditional printed document.

The beauty of a PDF file lies in its ability to:

1. Maintain excellent graphics (this is, after all, a form of the PostScript language),
2. Cross multi-platform barriers *usually* painlessly,
3. Deliver embedded fonts and symbols (Greek and mathematical characters),
4. Export text and graphics, and
5. Print a document with a single print command (instead of the multiple print commands required to print the many sections of an HTML file and the ensuing nightmare when collating).

An Unfulfilled Promise

Because of these characteristics, Acrobat files could support a print-on-demand system for documents, to replace the current system wherein publishers print and distribute books and journals. However, the promise of print-on-demand is not yet fulfilled.

Print-on-demand is often touted as a solution to the high cost of printing traditional documents. In reality it simply shifts the cost for the printing of documents from the publisher to the consumer, or an organization such as a library.

A print-on-demand system has two facets:

1. The production of high-quality books/journals/articles for archival purposes, and
2. The production of simpler paper copies for individual use.

A product for archival use would be printed at 2400 dpi on archive-quality coated paper, AND assembled as a case-bound volume, thus meeting the standards of the U.S. Library of Congress and other archival organizations. The life span goal of such a document is 100 years. An individual user, on the other hand, typically has access to a 300 or 600 dpi laser printer that can print a document on one side of the sheet only, and requiring collation for stapling or insertion into a binder. The quality of such documents usually falls far short of archival specifications, though. In some areas local print and copier businesses offer this service by printing supplied electronic files, usually on a laser printer or copier. Some may even have WWW access.

Desktop printers available to most individual users do not yet reproduce images of the quality required by paleontologists, petrologists, and other geologists, but this situation is likely to change within a few years. Considering archival requirements, "photo-quality" printers are becoming available, but are still based on the application of ink (or toner) to a sheet of paper usually in dots or spray, and are thus not capable of reproducing the full white-to-black scale that is found in a photograph. These printers are not designed for speed or mass production. Then again, mass production is usually not an issue in a print-on-demand scenario.

Imminent Change in Viewing

To date, most of the discussion in the work place regarding viewing electronic documents has assumed their being read on computer monitors, or as printouts on paper. A third technology, combining characteristics of both paper copies and reading from a monitor is upon us in the form of the electronic book, or e-book (see the e-Book Network). These devices, approximating in size a single- or double-page book, offer a convenient way to carry about and read electronic documents that have been downloaded from a remote server. Files in both HTML and Acrobat PDF formats will be appropriate for this type of viewing. One type of e-book has the astonishing capacity of half a million pages in color. Thus it could hold a good-sized reference library, with a fraction of the wall space. Further in the future, but presently under development, is a technology where documents are viewed on a kind of electronic paper (see MicroMedia Laboratory, which is a part of the MIT Media Laboratory).

MIRROR SITES: ARCHIVE AND ACCESSIBILITY

Modern Medieval Monasteries

At this moment many electronic publications can be frustrating to access. Sites may be poorly designed (e.g., over-sized graphics or files that are too big to download) or an individual's access may be restricted due to narrow bandwidth or limited access.

To improve accessibility and enhance archive insurance, publishers[2] should make use of mirror sites. Reliable mirror sites are like medieval monasteries where monks held the surviving copies of important documents. Such world-wide mirror sites will help ensure that documents remain available and accessible. Multiple sites also help make those same documents more easily accessible as the Internet pipelines linking countries and continents do not always run at optimal power. A good scenario comprises of mirror sites on each continent; a better one, mirror sites in multiple countries; and, the best, multiple sites in individual countries at many different institutions. Users around the world would be able to log onto the nearest site ensuring optimal access.

The optimal location for a mirror site within an institution should be at a library or a museum. It is possible for an educational institution to close a geoscience department, but a library is usually part of the infrastructure of a university or college and likely to remain intact. This same relationship hold true for a museum. As has been the case historically, publishers remain in the business of preparing and producing documents, while librarians and museum curators continue to work as the archivists of our civilization.

Statistics come and go on the Internet, but an alarming number of web pages appear and just as quickly disappear. Although unsustantiated reports regularly appear in e-mail communications that the average web page lasts 24 hours or less than a week.

The fate of one mirror site is illustrated in this poignant e-mail: "Unfortunately ... our site is now gone forever as we have an unreliable electricity supply to the server and it's not worth sorting out ..." (Riedel 1999).

The Issue of Archive and Accessibility

The Library as Archive. Throughout history, the structures that maintain the continuity of civilizations have been attacked and often destroyed. In many instances, though documents from those times and before did survive, often because copies were stored in out-of- the-way archives.

Although not as devastating as a period of social unrest, another attack on libraries came about due to the use of acidic paper, mostly since the end of WWII. At the University of Toronto library it is estimated that 6 million books are in jeopardy from this threat (Patterson 1998). Electronic publications face a similar threat. Unlike a printed volume, which required a wide distribution and acid-free paper to secure its lifespan, electronic publications are affected by:

1. The transitory nature of file formats,
2. The rise and fall of the popularity and availability of CD-ROM media,
3. Operating systems that change several times a year, and
4. The uncertain lifespan of the World Wide Web.

Document survival is a key goal for all publications, including those of electronic format, and not surprisingly, it is librarians who are stepping into the breach to ensure this. Libraries, known as "super libraries", are making the commitment to keep documents on-line by keeping older systems functional, and by migrating documents from outdated formats to accessible formats. In 1998 Texas A&M University added an entire wing devoted to electronic publications.

Long-term Accessibility. The Association of Research Libraries has formed the Scholarly Publishing and Academic Resources Coalition (SPARC). This group is encouraging the formation of electronic clearing houses, which could make documents available to a wide audience at a reasonable cost (Machen 1998). This concept bypasses traditional methods of scholarly information dissemination, and is clearly a response to the fact that library site licenses for WWW versions of journals can cost as much as 25% more than paper versions. The SPARC is also collaborating with the American Chemical Society to produce an organic chemistry journal, one of the first efforts by libraries to compete with traditional scholarly publishing.

The WWW allows authors and publishers to think of new ways to develop publications. In the past, a document was defined by it's publication date. Although other books could be published, once off the press, a traditional book is complete. This is no longer the case. For instance, the ODP is an internationally funded program that exists within a

definitive time period, but the shelf-life of the prime data generated by the program should be indefinite. Papers based on ODP-research will continue to appear for many years after the end of the program. If the archive of information is gifted to an institution willing to support the migration of material to accessible platforms and to support the review process for submitted papers, publications could continue indefinitely, expanding further and further the breadth of the ODP collection.

RESOLUTION

The benefits of electronic publishing can overcome emotional resistance on the part of authors, if the continuing technological improvements take place to overcome resistance on the part of readers. Every document on the WWW has the opportunity to grow in value as its relationship to other documents widens. Archiving issues are being settled as librarians, the archivists of our civilization, take on the new challenge.

The world is a wide place, but the scholarly community is forging a complex new unity based in electrical impulses.

REFERENCES

Chodorow, Stanley 1997. "The Medieval Future of Intellectual Culture: Scholars and Librarians in the Age of the Electron." **University of Pennsylvania from OAH Newsletter, Organization of American Historians**, vol.25, no.4 (Nov. 1997). p.10-11.

Hammer, Oyvind 1998. Computer Simulation of the Evolution of Foraging Strategies: Application to the Ichnological Record. **Palaeontologia Electronica**, 1(2): 21 p., 98 KB. http://www-odp.tamu.edu/paleo/1998_2/hammer/issue2.htm

Harnad, Stevan 1998. "On-line journals and financial fire walls," **Nature**, Volume 395, September 1998, p. 127-128.

Kay, Alan 1992. (Apple Computer, Inc.), personal communication to Edward J. Valauskas, 7 December 1992.

Lyons, Patrick D. and Laurence Head 1998. QuickTime VR: A Powerful New Illustrative Tool for Micropaleontological Research. **Palaeontologia Electronica**, 1(2): 12 p., 219 KB.
http://www-odp.tamu.edu/paleo/1998_2/lyons/issue2.htm

Machen, Rod 1998. "Libraries attempt to compete with scholarly publishing." In **The Battalion**, Texas A&M University (August 6 1998.) p. 1.

Monks, Neale and Jeremy R. Young, 1998. Body Position and the FUncation Morphology of Cretaceous Heteromorph Ammonites. **Palaeontologia Electronica**, 1(1): 15 p., 1.8 MB.
http://www-odp.tamu.edu/paleo/1998_1/monks/issue1.htm

Patterson, R. Timothy 1998. Carlton University, personal communication to Jennifer Rumford, October 1998.

Riedel, William R. 1999. Scripps Institution of Oceanography, personal communication to Jennifer Rumford, January 4, 1999.

Tufte, Edward R. 1992. (Yale University), personal communication to Edward J. Valauskas, 13 November 1992.

Valauskas, Edward J. 1994. "Reading and computers - paper-based or digital text: what's best?, **Computers in LIbraries**, 14:1 (January 1994), 44-47.

Valauskas, Edward J. 1998. "**First Monday** and the Evolution of Electronic Journals." In the **Journal of Electronic Publishing**, http://www.press.umich.edu/jep/03-01/FirstMonday.html

LINKS TO MORE INFORMATION

Places to go to learn more about electronic publishing. See also, the Glossary in the Plain-Language Summary.

First Monday, Peer-Reviewed Journal on the Internet

Waiting for Thomas Kuhn: First Monday and the Evolution of Electronic Journals, Edward J. Valauskas

First Monday and the Evolution of Electronic Journals, Edward J. Valauskas

JEP, the Journal of Electronic Publishing

Why I Don't Read Electronic Journals: An Iconoclast Speaks Out, Peter Grenquist

Transcending the Limitations of the Printed Page, Judy Holoviak and Keith L. Seitter

eBook Network

MIT Micromedia Lab

Electronic Paper

Nature: Web Matters http://helix.nature.com/webmatters/ Nature (no direct link allowed)

Glossary

ASCII: American Standard Code for Information Interchange. This is the primary standard for coding text used on almost all computers and printers, but is very limited in the number and type of characters available for use.

Browser: The software that allows a reader to find locations on the WWW.

HTML: HyperText Markup Language. This is a description language (based on SGML) that allows different browsers to see the same document. Unfortunately, not all browsers are equal, not all platforms or computers are standardized, so no one sees exactly the same document. Only the content is the same.

Home page: The entry point for each web site or server.

Internet: A worldwide network of computers. Not a destination, but a means to

reach a destination.

PDF: Acrobat Portable Document Format.

SGML: Standard Generalized Markup Language. This is a much more sophisticated language (more than HTML) used to define the structure of a document. Any document has a set of common elements, titles, paragraphs, list, etc. These elements can be defined and labelled for automated indexing, footnotes, or cross-references.

URL: Uniform Resource Locators (the address of a given file on the WWW)

WWW: World Wide Web. From the World Wide Web Initiative, a cooperative organization originally developed in 1990 at CERN, the European Particle Physics Laboratory in Switzerland.

XML: Extensible Markup Language. This language is a subset of both SGML and HTML. It is one of the current Great Hopes of electronic publishing, as it promises to offer the best of both of the other two languages.

Figure 1. Use of movie clips **(A)** and animations **(B)** can enhance the reader's understanding of the document. [Link to actual article, **A** for movie, (Lyons and Head, 1998), **B** for animation (Monks and Young, 1998).

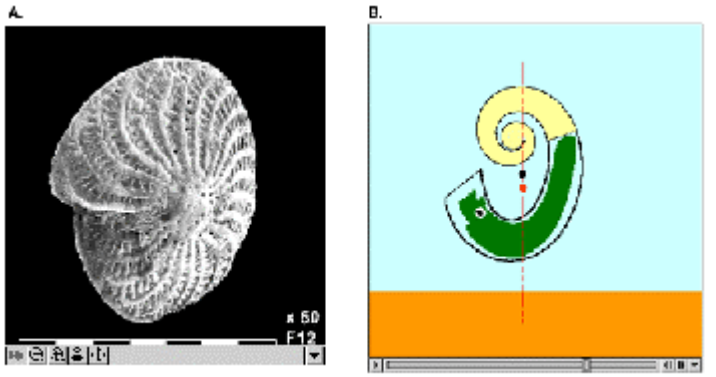

Figure 2. Linking references to their sources, both forward and backwards, enhances the reader's ability to use the document. [Link to actual article] (Monks and Young, 1998)

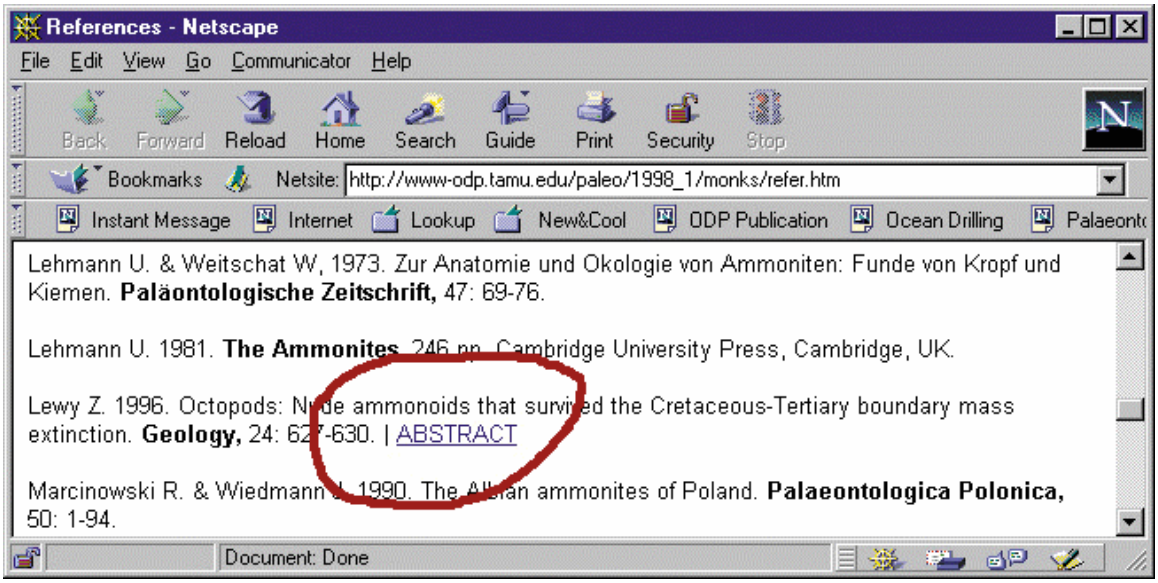

Figure 3. Sample of on-line discussion. [Link to actual article] (Hammer, 1998)