The Future of Electronic Publications: A Perspective from the Ocean Drilling Program

Mary Chapman, Patrick H Edwards, and Jennie L Lamb

Introduction

The Ocean Drilling Program (ODP) was an international partnership of scientists and research institutions organized to explore the evolution and structure of the earth. ODP was the successor to the Deep Sea Drilling Project (DSDP), which operated from 1968 to 1983.[1] The National Science Foundation was the principal funding agency for ODP.

As the science operator for ODP, Texas A&M University (TAMU) was responsible for collecting core samples of rocks and sediments from beneath the world's oceans and ensuring that adequate facilities were maintained for analysis and preservation of the samples. To discharge that responsibility, ODP-TAMU operated the drilling vessel *JOIDES Resolution*. ODP concluded drilling operations on 10 September 2003. Its successor, the Integrated Ocean Drilling Program (IODP), is scheduled to begin collecting core samples in June 2004. TAMU has been chosen to be the science operator for IODP and will continue to produce the remaining publications for ODP.

Each cruise of the *JOIDES Resolution* was called a "leg" and lasted about 2 months. Every leg had mission-specific scientific

MARY CHAPMAN, *formerly an editor at the Ocean Drilling Program (ODP), is now a technology applications instructor at Jane Long Middle School in Bryan, Texas.* PATRICK H EDWARDS, *formerly a production specialist at the Integrated Ocean Drilling Program (IODP), is now a content developer at K2Share in College Station, Texas.* JENNIE L LAMB is *a graphic specialist at IODP at Texas A&M University in College Station, Texas. IODP is the successor program to ODP.*

goals chosen through a careful review process. The ship's complement consisted of scientists, engineers, technicians, and various crew members. During a typical cruise, scientists drilled holes deep into the seafloor to penetrate millions of years of the earth's geologic history. In the cores extracted from the holes, scientists retrieved sediment, rock samples, and geophysical data. This information revealed clues about the earth's origin, evolution, and present-day structure. All cores retrieved were split into halves. One half was consumed in scientific study on board. The other was photographed and then placed in cold storage. After the cruise ended, each half-core was transported to one of four repository locations for future research.

Each leg resulted in the production of four publications (*Scientific Prospectus, Preliminary Report, Initial Reports,* and *Scientific Results*). We discuss here the electronic publication process for

- *The Proceedings of the Ocean Drilling Program, Initial Reports* volume (published about a year after the cruise), which contains a thorough record of cruise objectives and summarizes the results.

- *The Proceedings of the Ocean Drilling Program, Scientific Results* volume (published about 4 years after the cruise), which documents postcruise research results.

Those volumes were distributed in print, on CD-ROM primarily in portable document format (PDF), and on the World Wide Web in both HTML and PDF format. There were six legs per year, on the average, and we produced 12 volumes per year.

Originally, *Initial Reports* and *Scientific Results* volumes were printed in their entirety, which frequently ran to 800 pages or more. For the accompanying CD-ROM, we directly exported the Adobe FrameMaker documents to PostScript and converted them to PDF with Adobe Acrobat Distiller. We generated thumbnails and bookmarks from specially tagged headings in the text. However, we made virtually no use of interactive features, such as hyperlinks within and between PDF files and embedded QuickTime movies. In fact, only the authors' e-mail addresses were linked.

In 1995, the JOIDES Science Committee (SciCom) and Executive Committee (ExCom) panels mandated that ODP substantially reduce the Publications Department budget and increase the availability of our publications. We responded to that challenge and changed from a print-oriented publication model to an electronic format. For a full description of the process and the challenges it entailed, see the *Science Editor* article by Klaus and Miller.[2] In the present article, we discuss specific software currently used to produce the final electronic publications for ODP and the direction that electronic publications may take in the future.

Selected Technology Used in ODP's Electronic Publishing Model

Text Preparation

We use Adobe FrameMaker 7 for page layout because it is well suited to long-form scientific publishing with large amounts of tabular data. We import all figures and plates by reference, copy tables into the FrameMaker chapter document, and use the "Equation Editor" to typeset equations that are too complex to be represented in

Electronic Publications continued

text. We also create cross-references to automatically insert page numbers where required in the text.[3]

When a chapter manuscript reaches "page proof" stage (the manuscript should be in its final corrected form), we add interactive PDF features from within FrameMaker. Using the features in the "Hypertext" dialog box, we add named destinations for all figures, tables, and plates and for any section headings referenced by name in the text. We then create hyperlinks to those destinations in the same file and in other FrameMaker files. We also use hypertext to link e-mail addresses and World Wide Web URLs. By the time we create our PDF files, we have taken advantage of FrameMaker's powerful PDF setup features, and little formatting is left to do.

After the preparation work in FrameMaker, we are ready to convert the FrameMaker files to HTML with Quadralay WebWorks Professional, a program designed for that purpose. In the HTML version of our manuscripts, each figure, table, or plate is hyperlinked to open in a separate window for easier viewing.

Figure Preparation

Once the figures are in-house, they are formatted in their native applications—which can include Adobe Illustrator, Canvas, and Kaleida Graph—to conform to ODP style and scientific illustration standards. After the preliminary editorial and author reviews of the material are completed, the figures are converted into either Illustrator EPS or Photoshop TIFF files so that they can be imported into the FrameMaker files.

Once the material is ready to be converted to HTML format for publication on the Web, we convert the figure EPS and TIFF files to JPEG or GIF files. Because a leg may have many figures, our senior graphic designer, Deborah Partain, has streamlined the conversion process by using AppleScript. She has developed two custom scripts: a "Cleanup" script, which deletes stray objects from EPS or TIFF files, and a "Convert" script, which determines
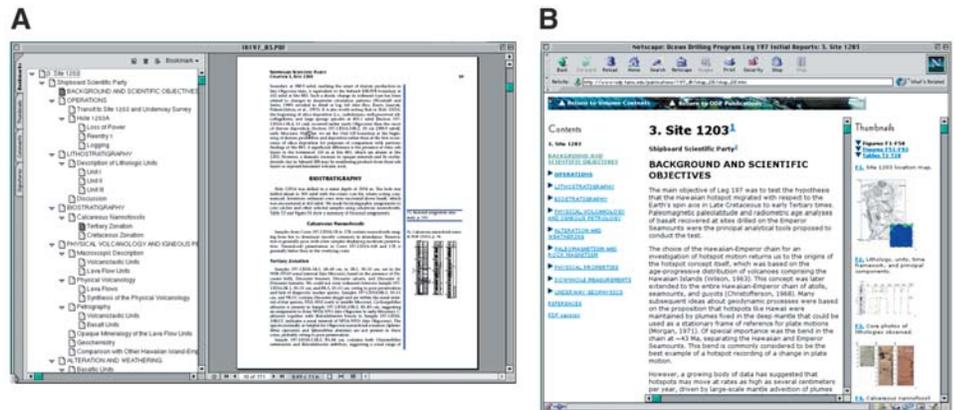

Figure 1. Examples of ODP electronic chapter layout in PDF format (A) and HTML format (B).
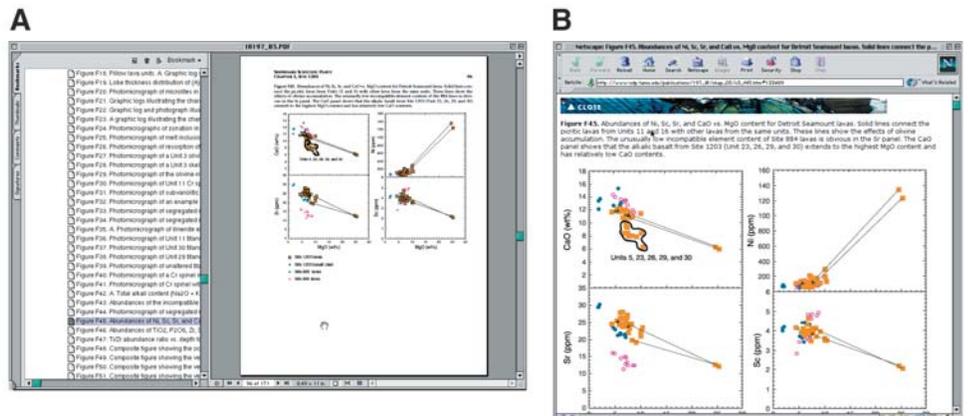

Figure 2. Examples of ODP electronic figure layout in PDF format (A) and HTML format (B).

Figure 3. Examples of ODP electronic table layout in PDF format (A) and HTML format (B).

the scaling percentage and format of the HTML figure graphics. If the file contains text of a point size less than or equal to 5 points, it is scaled 200%; otherwise, it is scaled 150%. (All figures are scaled larger than 100% for better screen readability.) If the file contains any raster information, it is converted into a JPEG file; otherwise, it becomes a GIF file.

Core-Description Preparation

The core-description portion of the *Initial Reports* volumes differs greatly from the volume text in substance and form. Whereas the chapter portion of the volume is heavy in text, tables, and graphics and provides some scientific interpretation of the data collected during the leg, the core-description files are predominantly graphic representations of only the leg's prime data and contain no interpretation. In fact, they are often referred to as field notes and therefore do not undergo as rigorous an editing cycle as other volume material.

The Graphic Design section of the Publications Department handles the entire production process of this portion of the volume. Several types of prime data may be included in a leg. The first type is visual core descriptions (VCDs), graphic representations of recovered core in its original state. There are two types of VCDs: barrel sheets and hard rocks. Barrel sheets (descriptions of sediment cores) are prepared electronically in AppleCore, a proprietary software application developed by Mike Ranger for the Macintosh platform. Hard rocks (descriptions of hard rock cores) are usually hand-drawn, scanned, and then traced in Illustrator; the text accompaniment is formatted in the same document.

"Smear slides" (descriptions of a type of sample taken from sediment cores) and "thin sections" (descriptions of samples extracted from hard rock cores) are tabular data and are prepared in Microsoft Excel. Other types of tabular data are occasionally included in the core-description files, such as alteration and vein logs.

After author review of the material, it is edited in the native applications. The files are then exported into different for-

Figure 4. Example of ODP electronic core-description layout in PDF format.

mats (barrel sheets are exported as PICT, hard rock descriptions as EPS, and Excel tables as ASCII) that can be imported into FrameMaker.

Our volumes tend to have anywhere from 20 to 150 VCDs per site. It is this magnitude of work that prompted the development of the automated techniques that we now use. Three custom AppleScripts[4] are run to automate the completion of the VCDs. The "Newlink Import" script imports and tags the text that will become the bookmarks in the VCD PDF file. It also sets up a named destination for each VCD. The "Message Openfile" script imports, tags, and hypertext-marks the text that will become the link to a separate, corresponding core image. The third and final script is the "VCD Import" script, which imports and scales the PICT or EPS graphic file and places it on the page.

After the VCD portion is completed, the tabular data are imported and formatted in the FrameMaker file. The FrameMaker file is then ready for conversion to PDF. We generate a PostScript file, which we modify in BBEdit by Bare Bones Software (our text-editing application of choice) to include our document information text, organize the bookmarks, and manage our article threads. Acrobat Distiller is used to convert the file. The modifications made in the PostScript file eliminates any "finishing" of the PDF file in Acrobat.
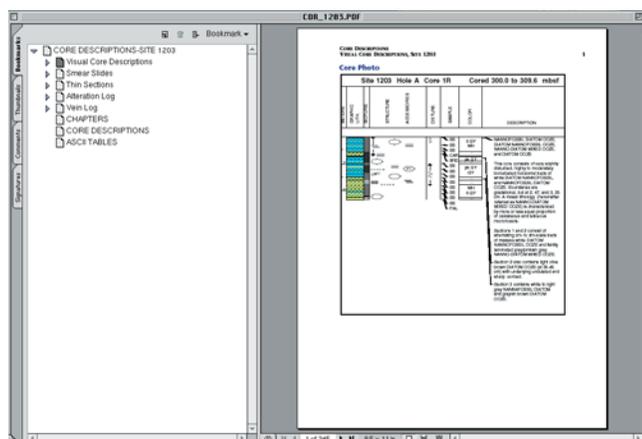
CD-ROM Preparation

For the CD-ROM, we print PostScript files from FrameMaker and distill the CD-ROM PDFs at 600 dpi. The CD-ROM may contain drilling-site maps, ASCII files, supplementary material, oversized figures and tables, and such other materials as QuickTime movies. We also create two indexes for each volume with Acrobat Catalog: the Volume Index, which enables the user to search for terms across all the PDFs on the CD-ROM, and the Compiled Index, which enables the user to search for terms across all the ODP proceedings volumes.

For the CD-ROM, we put navigation bookmarks in the PDF files. These link to three particular parts of the table of contents: chapters, core descriptions, and ASCII tables. We do not include these in the PDFs posted to our Web site, because we want the user to open each PDF file in a separate window and we want to refer the user to the online HTML "Table of Contents". We force the user to keep coming back to the one open HTML page in order to simplify navigation and make all our resources available from one window.

Future Directions

What Do Authors Want and Need?

The question to be answered as editors

Electronic Publications continued

move toward an increasingly electronic workflow and publishing process is—what will best facilitate the needs of authors, reviewers, and publications? To have some idea, we must first understand how scientists use electronic publications.

In a survey done at Stanford University, scientists in a variety of fields were queried as to their habits regarding the use of electronic media in their fields. Much of what was learned shows that the general user moves around the electronic environment of the publication and takes the needed pieces without necessarily having a firm grasp of all the uses and capabilities of the system. The study also showed that use changed from time to time, depending on need. Finally, and perhaps most important, it showed that no user was completely paperless.[5]

A survey by Tenopir and King, reported at the 2002 CSE annual meeting, showed that the number of personal subscriptions to scholarly journals has decreased every year since 1977. Scientists now procure scholarly articles in a variety of ways (library-provided subscriptions, department collections, electronic journals with links, and so on). Most scientists using electronic publications preferred PDF format for various reasons, the most important being cost-effectiveness (you can print the article from the Web site for pennies in comparison with a personal subscription to a scholarly journal). Material in PDF format is especially readable when printed out, as opposed to HTML-formatted documents. Scientists were interested in procuring not only recent publications but older ones as well. The American Astronomical Society (AAS) has digitized all scholarly articles back through 1849. It is interesting to note that 66% of the articles read by scientists using the AAS archive were at least a year old.

These findings dovetail nicely with those of an internal survey of electronic publications at ODP. In a questionnaire on the ODP Web site, users were asked questions regarding their use of the electronic version of the ODP proceedings and their preferences regarding an electronic

workflow. As expected, the lower cost and convenience of online digital publications were the driving factors behind their use. However, somewhat surprisingly, 52% of the respondents found HTML versions of the text useful despite their preference for PDF-formatted documents. We also questioned interest in digitizing of older volumes to serve as an archive of ODP data. Interest in this was exceptionally high (87% were interested in accessing nonelectronic ODP volumes). The survey asked what the respondents would like and be willing to do with respect to electronic submission workflow. Only 26% of those questioned had submitted a manuscript to a journal electronically; however, 74% said they would use an electronic submission workflow if it were offered. Some 91% of those surveyed were willing to provide their own PDF-formatted manuscripts, but most (65%) preferred to allow ODP to generate the file for them. Some (13%) had served as reviewers in an online or electronic review process (receiving electronic manuscripts and returning comments as embedded PDF files or e-mail documents). Most were willing to participate in online peer review (65%). Respondents said they would prefer to print and read on paper the PDF-formatted document (65%) and return their comments via e-mail (48%) instead of embedding comments in the PDF version of the manuscript (22%).

Those statistics are at best a rough estimation, but they give some idea of the direction that electronic publication may be headed. Participants seem willing and able to begin working in a more streamlined electronic workflow. Bearing in mind the software now being used, what new directions can publications take as they move to a more encompassing electronic workflow?

Streamlined Electronic Submission
The submission of manuscripts electronically through a Web-based submission process would allow authors and reviewers more flexibility and help in the timely return of manuscripts and reviews. A number of companies offer services to facilitate

online workflow, and in larger organizations the infrastructure necessary for such online submission may be created in house. These systems provide servers for storage of manuscripts, databases designed to track the progress of manuscripts through the review process, and electronic calendars and automated e-mail to alert authors and editors when a manuscript has missed deadlines. All those tools will facilitate timely publication of manuscripts.

A process implementing annotation features available in Acrobat during the review and production process by authors and the staff of a journal would allow notes to be passed back and forth with greater ease; however, use of the annotation feature is difficult and still somewhat daunting.

Structured FrameMaker and XML
FrameMaker 7 supports the creation of structured documents that can be exported to XML via WebWorks for publication on the Web. Using WebWorks, we can map paragraph and character tags to online styles, determine how the FrameMaker document is divided into separate files, and use Scalable Vector Graphic (SVG) images directly. Computer specifications and hardware and software requirements change, so conversion of future and existing publications to XML format can help to ensure the usefulness and accessibility of data in the future.

Scripting and Automation
AppleScript, FrameScript, and VisualBasic for Applications automation of more features would be useful in various ways. AppleScript can be used to streamline importation of figures and table template formats. That would expedite the process for graphic designers and production editors. VisualBasic can be used to create powerful macros that automate tasks in Microsoft applications, thereby freeing up chunks of valuable editorial time.

SVG for Graphics
Illustrator is used to export EPS files to GIF or JPEG, and Macromedia Fireworks

is used to create thumbnail versions. These are bitmaps and work well for our current Web implementation, but there is room for improvement. SVG is a creation of Adobe. Defined by the World Wide Web Consortium (W3C), it could eliminate creation of separate versions of figures for the Web. SVG is supported in Web browsers with plug-ins and allows addition of interactive features with JavaScript.

Archiving and Preservation

The life span of acid-free paper is about 100 years, which is good for printed publications, but doesn't address electronic efforts. The life span of a CD-ROM, depending on whom you ask, is estimated to be 50 years at best. And what guarantees are there of the future availability of HTML files? When an endeavor with a finite life span, such as ODP, ends, what happens to the data collected and published by that program when they turn out the lights for the very last time ever?

Currently, ODP Publications has mirror sites that will preserve our content, and we have a commitment from the TAMU Digital Library (*dl.tamu.edu*) to make our data available indefinitely. Also, our data will be available through the National Geophysical Data Center (NGDC).

Conclusion

The future of electronic publishing is changing constantly, leaving editors the daunting task of assessing the new technology available and how it will integrate into the workflow of their publications. The wide variety of software used to automate tasks combined with powerful page-layout programs moved ODP from a print to an electronic publications model fairly easily. What will come next, we hope, is a gradual movement toward a completely electronic workflow from initial submission of manuscript via either e-mail or a Web-based submission framework, through an electronic peer-review process, and finally to

publication of more integrated and accessible manuscripts on the Web. The key to this transition for both IODP and other publications lies in keeping abreast of new technology and the changing needs of editors and authors.

References

1. A guide to the Ocean Drilling Program. JOIDES J 1999;25.

2. Klaus AD, Miller AT. Cross-media publications and the editor: the Ocean Drilling Program experience. Sci Ed 2002;25:86-9.

3. Rumford JP, Benson KO, Gracia JA. A case study using FrameMaker for dual-purpose documents (print, PDF, and HTML). Minneapolis: Proc 1999 FrameUsers Conf; 1999. p 429-56.

4. Partain DL. Using FrameMaker to create a PDF catalog of images for CD-ROM/Web distribution. San Diego: Proc 2000 FrameUsers Conf; 2000. p 265-82.

5. Institute for the Future, 2002. Final synthesis report of the e-Journal User Study. ejust.stanford.edu/SR-786.ejustfinal.pdf. Accessed 3 Mar 2003.