An overview of cross-platform document technology

Darren Foltinek

ABSTRACT

Recent advances in electronic document technology now allow documents that are produced on different computer platforms using a wide variety of software to be translated into a common format for printing and distribution. This paper gives an overview of this technology and describes how it was used in the production of the CREWES research report. Benefits include maximum image quality, flexible means of distribution, and the ability to archive documents.

INTRODUCTION

Writing a technical paper usually involves assembling text, equations and figures into a document. In the past, the text was usually written on a word processor, with spaces left for figures and equations. Equations were created using either a specialized mathematical language or more modern visual equation layout programs. Figures were assembled by hand from paper plots with glued-on annotation and then pasted into the document. In the field of geophysics, these figures are often highly detailed seismic sections with overlaid interpretations, full color cross sections, or time slices. The highly graphical nature of exploration geophysics has, until now, made it difficult or impossible to produce a technical paper without using scissors and glue to paste in paper plots.

There are several major problems that arise from the use of the glue and scissors (G&S) technique of producing papers. Every time a paper copy is reproduced the quality of the reproduction suffers, since most photocopy machines cannot create high quality reproductions of gray shades or colors. This means that a master copy of the paper must be carefully stored so that first generation copies of it may be made. It is also difficult to reuse figures that are made using the G&S technique. In the course of doing research, the same figures are often used to produce one or more papers, slides for a talk, or are then included in a thesis. Traditionally, each time a figure is reused, it must either be created from scratch, a time consuming and tedious process, or the

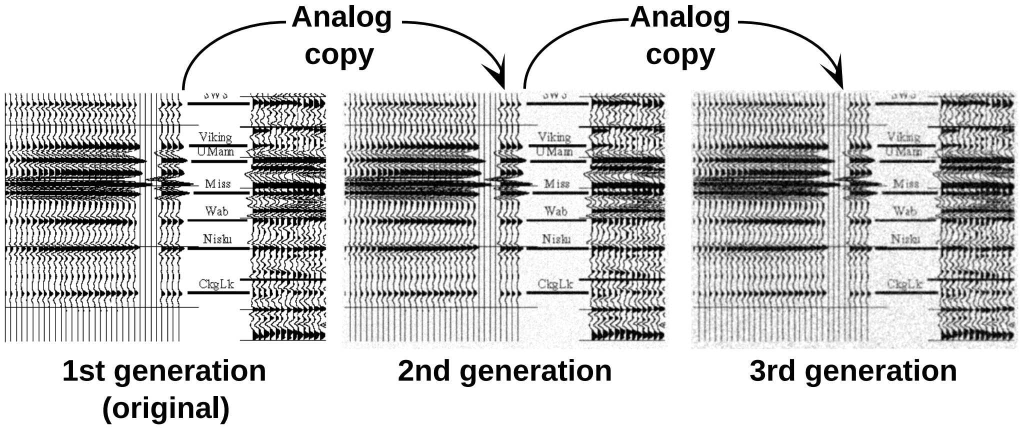


Figure 1: Typical image degradation in successive analog copies. Note increased noise and decreased contrast and sharpness.

figure must be removed from the master copy of the document. Distribution of a paper report is relatively slow and expensive, since it must be photocopied and sent by mail or courier. The final major problem with the G&S technique is that at the end of the process there exists only one master copy, which must be carefully archived so that it can be reproduced when needed.

By moving to a fully digital document flow, all of these problems are addressed. The reproduction quality problem is solved since a fully digital document may be be printed anew each time it is needed. This creates a first generation paper copy, with the resulting maximum image quality. The problem of figure reuse is also solved, since digital figures may be easily copied and made into slides, inserted into another paper, or used in a thesis. Since each figure is digital, there is no loss of quality when it is copied. The distribution of a document is made much easier when it is fully electronic, since it can then be quickly transmitted over the Internet or by modem almost anywhere in the world. Finally, an electronic report can be archived very efficiently and safely.

DIGITAL DOCUMENT FLOW

Figure 2 shows the flow that a document takes as it is created and then goes through the translation to Postscript and PDF for printing and electronic distribution. The gray boxes (Postscript and PDF file) highlight the key points in the flow where the document is translated into platform independent formats. Not shown is the process of creating a figure, which often involves moving information from platform to platform, as workstation display screens are captured and then annotated.

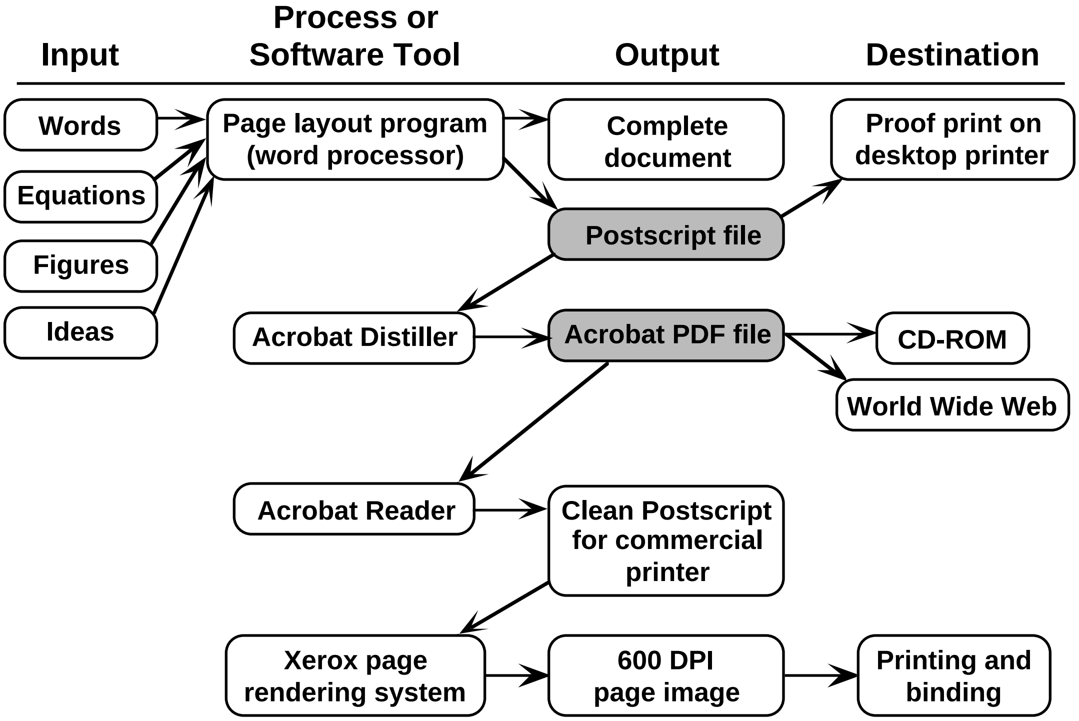


Figure 2: Fully digital document flow

Document Figures

A figure in a geophysical paper is often composed of an image (seismic section, time slice, map etc.) and some overlaid annotation.

The image usually comes from interpretation, processing or modelling software. In this case, the image may be either captured or translated so that it can be used in a drawing package to produce the final figure. If the image can be displayed in adequate detail on a workstation screen, the easiest way get that image is to “capture the screen”, which means to make a digital copy of the screen display. The resulting image is then edited to remove window borders, menu displays, or other unnecessary information. If the entire image cannot be displayed with adequate detail on a single screen, two options exist: translation or image merging. For example, a complete seismic section cannot usually be displayed with adequate detail on a single workstation screen. For this application, the section may be captured in single-screen sized pieces and then merged together using image editing software (Adobe Photoshop or Corel PhotoPaint, for example).

Translation software may be used to produce an image from the output of a package (usually CGM or Postscript) into a format suitable for import into a drawing package. If the image is only available on paper it can be scanned using a flat-bed scanner. The desired result of any image acquisition technique is an image of satisfactory detail which can be imported into a drawing package and annotated. An important item to note is that the detail level (resolution) should be the minimum necessary to show the required information. This keeps file sizes small and manageable.

Text and Layout

“Text and Layout” refers to the work usually done in a word processing package: text entry and formatting, image import and placement, and page design. The main concerns in text and layout are image format compatibility, pagination control, consistent formatting and font compatibility.

The image format that works best varies for each word processor or layout package. In general though, the format that is used internally by the operating system’s cut and paste functions has been found to work fairly well. For MS-Windows machines, this is the Window Meta File (WMF) format, and for Macintosh computers, it is the PICT format. Another fairly universal format is Encapsulated Postscript (EPS). Unfortunately, problems still exist, and all too often some part of a figure will be lost or damaged when it is inserted into a document. There are no hard and fast rules for solving these problems, and usually experience and trial and error are needed to fix these problematic figures.

These three graphic file formats are also somewhat compatible across different computer platforms. Most good layout packages on the three major platforms (MS-Windows, Macintosh and Unix) support EPS files, and there are several drawing packages on the Windows and Mac platforms which support both PICT and WMF files. We have found that the most reliable format for moving graphics between MS-Windows and Macintosh machines is the EPS format.

To ensure a consistent appearance of each chapter in the CREWES research report, template documents were made for each of the word processing packages used. These templates defined a standard set of paragraph styles (title, headers, footers, figure captions, etc.). Since most word processors now support the use of styles, it became easy for authors to create chapters with a consistent appearance. Properly designed styles address most of the pagination, formatting and font concerns.

Integration: Postscript and Acrobat

The CREWES research report contains approximately 50 chapters written by over 30 authors. These chapters are integrated into a single document that is printed from digital files and distributed on paper and in electronic format. What makes this possible are two related software technologies from Abode Systems Incorporated: Postscript and Acrobat.

Postscript is a page description language, first introduced in 1985, that has become the defacto standard language for driving printers. (Adobe 1996, Postscript Overview) Today, most applications on most computer platforms, and certainly all document layout programs, can output Postscript. One of the most important features of Postscript is that it is a device independent page (or image) description language. This means that a Postscript file will print at the maximum resolution of the output device, which can vary anywhere from common 300 dot-per-inch (DPI) laser printers to the 2400 DPI resolution of a professional typesetting machine. Postscript is the common format which provides the means to gather together all the output from a wide variety of layout programs on different platforms.

In 1994, Adobe introduced the Portable Document Format (PDF). PDF is based on the Postscript language. The imaging engine is the same, but the language has been simplified by removing the programming constructs (loops for example). PDF also improved upon Postscript by improving device and resolution independence, reducing file sizes, and adding the ability to include hypertext links (Adobe, 1996, Acrobat FAQ). An important feature of PDF is that it was introduced together with Acrobat Reader - a program that can display and print the contents of PDF files. Acrobat Reader is available for all major computer platforms and operating systems. (Adobe 1996, Acrobat Overview) The end result allows users of all of today's major computer systems to view and print PDF files.

Acrobat PDF files can be generated from Postscript files. This process is the key intermediate step which completes the cross-platform document flow (see fig. 2). Postscript files from a variety of packages on different computer platforms are translated into PDF files which can be displayed and printed by Acrobat reader on all major computer platforms. PDF files retain all of the formatting details and graphic quality of the original document. This means that paper copies printed from a PDF file are first generation prints, with the same quality as the original.

When a Postscript (PS) file is translated to PDF any problems with the file will show up. Typical problems include missing or low quality images, incorrect pagination, or formatting errors. Since the PS -> PDF translation is the first step in the printing flow, it gives us the opportunity to catch any problematic chapters early on, so that they can be corrected on time.

PRINTING

The CREWES research report is printed and bound by a commercial printing company. As discussed in the introduction, the maximum quality of the paper copy is obtained by printing a first generation original for each copy, rather than analog duplication of a master copy. This is especially important for preserving the quality of grayscale images.

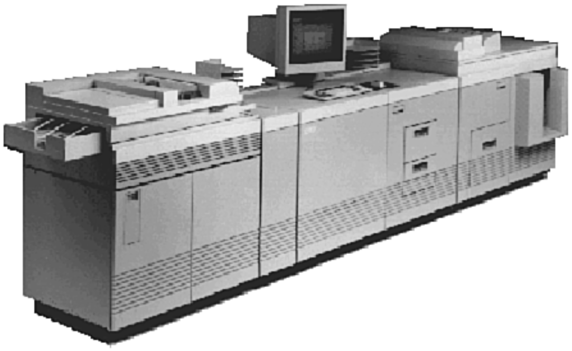


Figure 3: Xerox DocuTech 135 printer: 600 DPI and 135 pages per minute

The CREWES research report is printed on a Xerox Docutech printer (see fig. 3), which is a very high throughput 600 DPI laser printer (Xerox, 1996). To print the report, we deliver a collection of Postscript files and hardcopy to the printer. Each Postscript file (usually one for each chapter) is processed (rendered) to produce the 600 DPI image of each page. The hardcopy is just used as a reference, to ensure that each chapter is rendered correctly. When all 500 pages have been successfully rendered, the page images are copied to the Docutech and printed.

Even though Postscript is, in theory, a device independent page description language, there can still exist incompatibilities which will cause the printing of Postscript document to fail on a particular printer. For this reason, we have consulted with the commercial printer to know what is the best variant of Postscript to use so that their Docutech page rendering system will be able to process each page. Having PDF versions of each chapter enables us to recreate the Postscript in a variant which is best for the Docutech.

ELECTRONIC DISTRIBUTION

The CREWES research report is currently distributed to our sponsors in three ways; the bound paper copy, a CD-ROM containing the PDF files, and via our Web site on the Internet.

CD-ROM

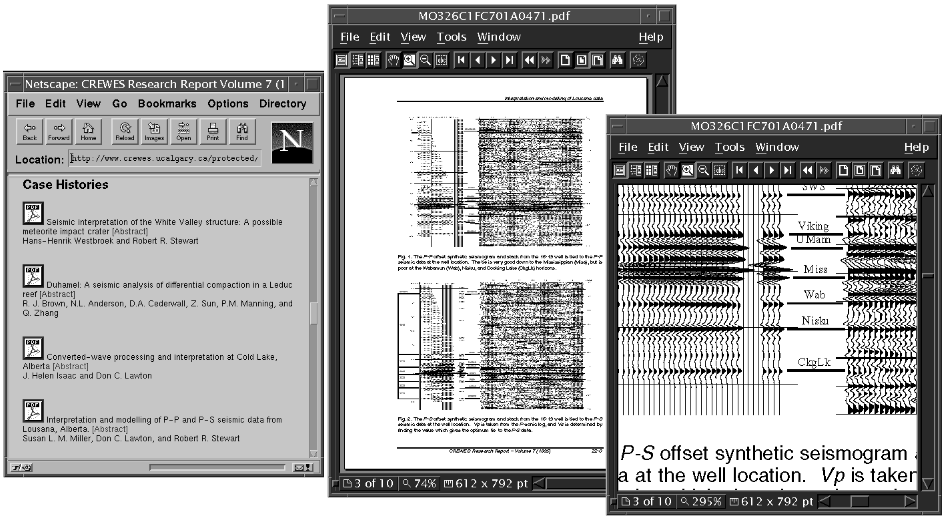
The CD-ROM version of the CREWES research report is a very convenient way to distribute and view the report. It contains all the PDF files and Acrobat Reader software for most platforms (Bland, 1995). An additional advantage of distributing the report on CD-ROM is that the figures are in color, whereas the report is printed in black and white. The CD is in multi-platform format, which means that it can be used on MS-Windows, Macintosh, and Unix platforms, and includes the latest versions of Acrobat reader software.

World Wide Web

The World Wide Web (WWW, or simply Web) has revolutionized the dissemination of information. More and more of the world's knowledge base is becoming instantly available at the click of a mouse, and CREWES has had a Web site since May 1994 (Foltinek, 1994). A fully digital CREWES research report in PDF format on our Web site enables our sponsors (with Web access) instantly access the research report from their desktop. Of course, to ensure confidentiality, we have password protected these sponsors-only portions of our Web site.

The PDF format is multi-platform by design and the files are relatively small compared to both Postscript and the original document files. The Acrobat Reader software can be integrated into most Web browser software, making it very easy for people who are browsing the Web to view PDF files. These features make PDF the best format to date for distributing complex documentation over the Web.

In order to benefit from the material that CREWES has on its Web site, including the research reports, paper reprints, theses, software updates and news, a sponsoring company must allow their employees access to the Web. More and more companies are doing this, now that the professional benefits of Web access are being realized. Also needed is the Acrobat Reader software for the users computer platform, which is free from Adobe Systems. Note that we have included the latest version of Acrobat Reader, for most computer platforms, on the CREWES CD-ROM.


Netscape: CREWES Research Report Volume 7 (1
File Edit View Go Bookmarks Options Directory
Back Forward Home Reload Images Open Print Find
MO326C1FC701A0471.pdf
File Edit View Tools Window Help
SWS
Viking
UMann
Miss
Wab
Nisku
CkgLk
3 of 10 74% 612 x 792 pt
3 of 10 295% 612 x 792 pt

Location: http://www.crewes.ucalgary.ca/protected/

Case Histories

Seismic interpretation of the White Valley structure: A possible meteorite impact crater [Abstract]
Hans-Henrik Westbroek and Robert R. Stewart

Duhamel: A seismic analysis of differential compaction in a Leduc reef [Abstract]
R. J. Brown, N.L. Anderson, D.A. Cederwall, Z. Sun, P.M. Manning, and Q. Zhang

Converted-wave processing and interpretation at Cold Lake, Alberta [Abstract]
J. Helen Isaac and Don C. Lawton

Interpretation and modelling of P-P and P-S seismic data from Lousana, Alberta. [Abstract]
Susan L. M. Miller, Don C. Lawton, and Robert R. Stewart

P-S offset synthetic seismogram
a at the well location. *Vp* is taker

1995 CREWES Research Report, on the Web, viewed in Netscape

Chapter 22 PDF file viewed in Acrobat Reader

Zoom in of detail in lower figure, viewed in Acrobat Reader

Figure 4: The table of contents of the 1995 CREWES report, downloaded from the Web (left). A single chapter is displayed in Acrobat Reader (center). The window on the right shows the very high image quality that is maintained in the PDF format.

Speed and ease are two major benefits of distributing the CREWES report over the Web. If a sponsor has Web access from the office, they can, literally within minutes, fetch a chapter of the latest CREWES report from our Web server and read it at their desk. This ease of access enables our results to be widely read within a sponsoring company, and ultimately improves the communication between CREWES and its sponsors.

DIGITAL DOCUMENT COSTS

There are some additional costs incurred when producing a fully digital report. The first is the additional user training and support requirements. The authors of our report are the CREWES students, faculty and staff, who’s skills in computer aided drawing and document layout vary greatly. All authors must be trained in the use of these software tools, and then technical support must be provided when problems are encountered or assistance is needed. We feel that these are valuable skills for any technical professional to possess, and so find this investment in training worthwhile. This time investment is also paid back to us in the increased quality and ease of preparation of presentations and papers once a student has learned how to use the tools of digital document production.

A significant cost is the additional computer hardware and software needed to work with high quality images and fully digital documents. For the 36 people that make up CREWES, we have a total of 12 personal computers (mainly Macintosh) that are used for document production. This is in addition to the Unix workstations that perform processing and visualization tasks. (Bland and Foltinek, 1996)

Printing a digital document is not significantly more expensive than printing an analog one. The same Xerox DocuTech printer is used, but in a digital flow, each page image is created from Postscript, whereas in an analog flow, each page image is created by scanned from paper.

The costs of distributing an electronic document over the Internet are very low. Distribution costs themselves (Internet connection fees and bandwidth) are carried by the University, leaving only the maintenance of the Web site as direct costs to CREWES.

CONCLUSION

By moving to a fully digital flow for the production of the CREWES research report, we have maximized the printed quality, improved the ease and efficiency of distribution, and efficiently archived the results of our research. We feel that these benefits are very good value for the extra costs incurred. The end result is better communication with our sponsors by improved access to our research results.

REFERENCES

Adobe Systems Incorporated, 1996, Adobe PostScript Overview, http://www.adobe.com/ prodindex/ postscript/ overview.html

Adobe Systems Incorporated, 1996, Adobe Acrobat FAQ, http://www.adobe.com/ acrobat/ acrofaq.html

Adobe Systems Incorporated, 1996, Adobe Acrobat Overview, http://www.adobe.com/ acrobat/ overview.html

Foltinek, Darren S., Stewart, Robert R., and Lawton, Don C., Electronic Documents on the World Wide Web, CREWES Research Report, Volume 6, 1994), ch. 20

Bland, Henry C., and Foltinek, Darren S., An introduction to the CREWES CD, CREWES Research Report, Volume 7, 1995, ch.46

Bland, Henry C., and Foltinek, Darren S., CREWES computer systems, CREWES Research Report, Volume 8, 1996, ch 10

Xerox Corporation, 1996: The Docutech 135, http://www.xerox.com/xps/products/dt135/index.htm