Stanford Exploration Project, Report 84, May 9, 2001, pages 1–387

CD-ROM versus The Web

Jon Claerbout and Mathias Schwab[1]

ABSTRACT

SEP has learned how to deliver reproducible research on CD-ROM. Publication via the Internet's World Wide Web may in the future offer a more flexible and efficient alternative.

INTRODUCTION

SEP has been producing research reports on CD-ROM since 1991 (**?**). Recently, the world wide web (WWW) presented itself as an alternative medium for distributing SEP's research (**?**Karrenbach et al., 1994) In this paper we reassess our efforts to distribute our reproducible research by CD-ROM and evaluate the opportunities the world wide web offers.

Purpose

The purpose of our CD-ROM publication effort is to provide reproducible research and technology transfer to

1. the authors
2. the authors' colleagues
3. the next generation of students
4. the sponsors
5. the general public and posterity

It may seem strange to put the author's own name at the top of the list to whom we wish to provide the reproducible research, but it often seems that the greatest beneficiary of preparing the work in a reproducible form is the original author! Author's lose reproducibility of their work by the passage of time exacerbated by their departure from Stanford.

The CD-reproducibility effort has two parts: The first part is defining and learning the standards for reproducible research. The second part is assembling and building the CD-ROM. These efforts are independent. With the system we have in place, the researcher does

[1]**email:** not available

almost none of the CD-ROM work beyond the work required to prepare reproducible research and a paper document. The researcher only copies his work to a staging area.

Space and "read-only" limitations may also create a few extra minor chores. The CD-ROM being "read-only" requires (which is good practice anyway) that researchers have in place a rule for removing all "intermediate" files, namely, those files that are created by programs from other files. This clean up is required for the CD-ROM to enable push-button reproducibility. It also makes revisiting an old directory less onerous for humans. The required existence of a makefile with a clean rule for intermediates and a burn rule for targets defines the test for reproducible research.

ACCEPTANCE

We expected widespread adoption of CD-ROM in the world of PC, Mac, and UNIX. We were wrong for UNIX.

- Although UNIX system managers generally have a CD-ROM reader connected to their file servers, few UNIX users have one connected to their workstations.
- UNIX vendors have not provided the needed software support and as a consequence
  - Most UNIX users do not have authority to mount a CD on their UNIX system.
  - Important UNIX vendors do not supply Rock Ridge, a device driver that can make the CD-ROM file system look like a UNIX file system.
  - Well known file system concepts for facilitating read-only file systems have not been adopted by UNIX thereby forcing our users to undergo a tedious step of "building a link tree".

We have had problems with our CD-ROMs not working on many systems. Our CD-ROMs have source code, documents, a few data sets, fonts, word processing software, and binary programs that we hoped would work on popular computers as well as all workstations manufactured by our sponsoring companies. The problems that we have had include

- Fonts not available on the displaying computer
- Bugs and differences in the C-shell
- Bugs and differences in Fortran-C linkage
- Solaris versus pre-Solaris
- Ghostscript versus X (still unexplained)
- Increasing complexity in makefiles in large projects
- Path problems associated with moving directories full of code
- Environment variables when accessing more than one CD-ROM at a time.

ECONOMICS

The economics of CD-ROM usage have improved in the sense that CD-ROM readers have dropped in price from about a thousand dollars to a few hundred dollars. The economics of CD-ROM production have not improved for us. We still consider a CD-ROM report to be a $2000 event, about the same as our paper report production event. We typically spend a couple hundred dollars on one-offs. We have some licensed software ($ 900 per year) for making UNIX CD-ROMs. The cost of mastering is about a thousand dollars. We make 250 disks for about $350 dollars, and have printing costs for the instruction folder. Additionally a skilled computer person assembles and tests the CD-ROM for about a week.

We might be able to cut costs by not mastering the CD-ROMs but instead making many singles. We could get by with about 45 singles (one for each sponsor and one for each researcher). I would hope that 50 singles would cost a total of about $2000. These figures assume we do not buy equipment but continue to use outside suppliers. Lower single incremental costs would apply if we purchased equipment at about $2-3000, and UNIX software estimated at $5-10,000.

PROSPECTS FOR WWW REPLACING CD-ROM

The World Wide Web[2] (WWW) coupled with bargain prices for hard disks offers promise of an alternate approach. We recently considered purchase of a large CD-ROM jukebox. On comparing its price to that of hard disk, we found them comparable, with the hard disk offering a much clearer future upgrade path.

The web offers a marvelous way to browse a file system. If you have not seen a demonstration of it, you should do so soon. Some people will know it under the name of popular browser software such as Mosaic and Netscape. (This is not the place for me to recount the marvels of the web.) By comparison, CD-ROM is a near failure as a browsable medium for SEP reports and theses.

For $3500 we could purchase a 9Gb disk and half fill it with our 10 previous CD-ROMs in read-only format. Many machine dependent binaries and fonts could be thrown away, meaning the cost for keeping this info on the web would be about $1500 for the lifetime of the disk. Material could be partitioned into public areas and restricted areas. We should do this regardless of our CD-ROM plans.

People would not be able to directly do reproducible work on our archival read-only disk file system. They would need to make copies and work from the copies. Naturally, the difficult chore of building a complete environment, as done on our CD-ROM needs to be done if reproducibility is to be achieved. We have not demonstrated a prototype for such local reproducibility but it seems possible. For people at remote sites to be able to build the required environment is even more problematic but not an unrealistic goal.

[2] Our web address is http://sepwww.stanford.edu

My goal is for people who have graduated and left Stanford for a couple of years to be able to rebuild the illustrations in their thesis. It turns out that by the time you reach that goal you have approximately met the goal of providing reproducibility for colleagues and sponsors and the general public. The University long-range position is that the research results should be reported in such a way that the work is reproducible by others. SEP is the first place where we have interpreted this quite literally.

We have had sponsors request web serving in preference to the CD-ROM. We do not know how many sponsors have web access but we are certain that some do not. In a few years we expect that most will have web service and we can supply the remaining with magnetic tape.

We have found potential colleagues in reproducible research in the Statistics Department at Stanford University (http://playfair.stanford.edu:80/ wavelab/). We see future prospects of interaction over the web (http://java.sun.com/) but is it too early to predict its impact on reproducible research.

CONCLUSION

We plan to continue making CD-ROM theses and reports until we learn how to deliver reproducible research on the Web. Until then we, nevertheless, publish research on SEP's Web site in a non-reproducible form, such as postscript files of completed theses or figures and papers in individual student pages.

REFERENCES

Karrenbach, M., Claerbout, J., and Berlioux, A., 1994, SEP goes World Wide Web: SEP–**82**, 265–268.