EXPERIENCES WITH HETEROGENEOUS PLATFORM CD-ROM'S

Sandy Ressler

National Institute of Standards and Technology
Gaithersburg, MD. 20899
Bldg. 220, Rm. A127
ressler@cme.nist.gov

Abstract

This paper describes some experiences creating a CD-ROM with functional software for both the Macintosh and IBM PC compatible computing platforms. Developing a CD-ROM which functions on multiple platforms requires attention to a number of areas. Even though an international standard for volume and file structure, ISO 9660, exists for files to be placed onto CD-ROMs, a number of other factors such as name space and end-of-line conventions can become quite troublesome.

The particular CD-ROM described in this document is intended to be of use by individuals involved in the development of formal product data standards. The CD-ROM contains software for perusing the documents, and a number of other items of interest to the product data community. These "browsers" function on either a Macintosh or PC compatible computer. Creating a multi-platform CD-ROM with software that functions on both a Macintosh and a PC is apparently an unusual practice, however one that will hopefully become more common.

Background

As with all information processing standards, one of the end results of a standards development effort will be paper documents. These documents must eventually be read, comprehended and used as reference material. The information in the standards must function both as educational material and as requirements for actual implementations. (The evolving series of standards for product data exchange known over time both as PDES and STEP is the domain covered by this work.) Reading over a thousand pages of dense technical material consisting primarily of data structure definitions is not a pleasant task. An on-line form of the document that can be searched and browsed would clearly be useful. An initial proof of concept, of standards browsing, was implemented on a Macintosh computer using HyperCard [1, 8]. Figure 1 illustrates the opening card for that prototype. (HyperStandard is a generic term describing the application of hypertext and multimedia technologies to standards in general.)
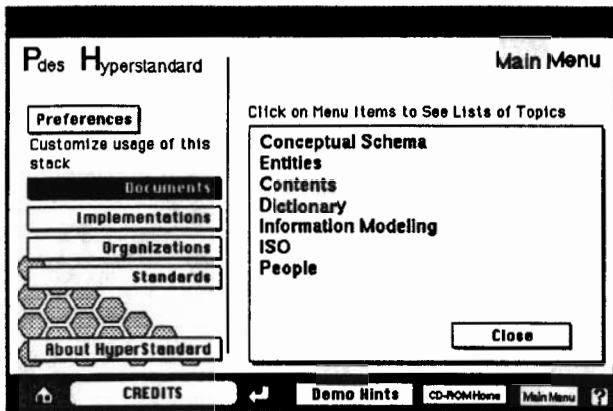
Figure 1 *Opening Card of the Original HyperStandard proof-of-concept*

Data Preparation

The document content which may be browsed on either a PC or Macintosh originated as documents in LaTeX [5, 6], which are drafts of the proposed STEP standard. Two completely different data preparation sequences were followed, one for the Macintosh to work with HyperCard and another for the PC to use Quantum Access [7], a PC based CD-ROM software package. The CD-ROM was produced according to the ISO 9660 standard for volume and file structuring [2] and files on the CD-ROM may be read from any machine whose device driver supports that standard. Figure 2 illustrates the main data preparation steps used to create the HyperCard-based browsers illustrated in Figure 3. (DeTeX is a SNOBOL program to convert LaTeX files into Pablum files. Pablum is a very simple markup created for this project)
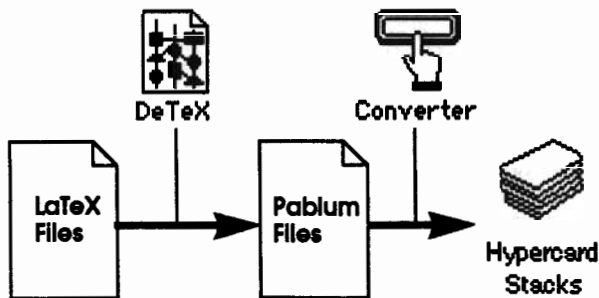
Figure 2 LaTeX to HyperCard Data Preparation path
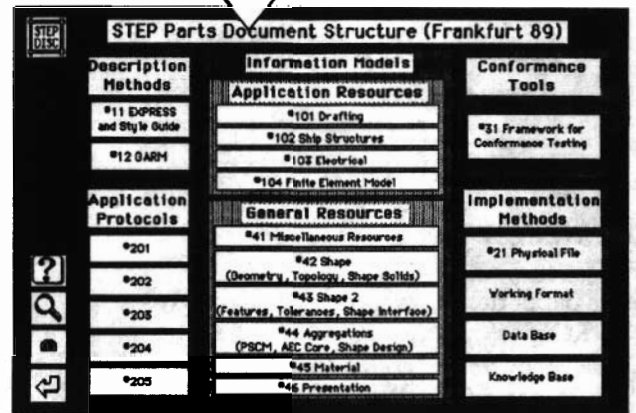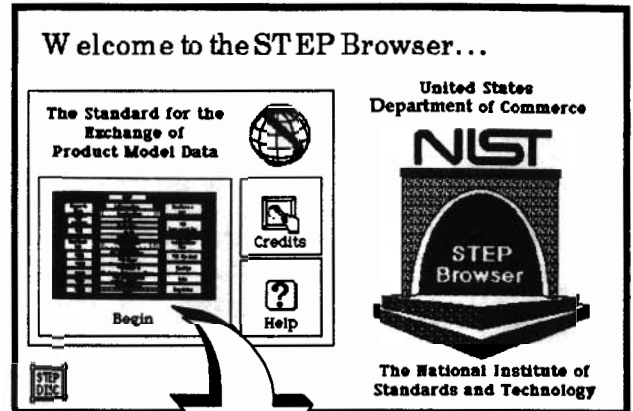
Figure 3   The HyperCard browser opening and initial cards

The PC compatible software also used the same LaTeX documents as the starting point but immediately diverged from the Macintosh data preparation path. The Quantum Access software had a well defined set of data input requirements. It mandated creation of a "flow" file which represented a table of contents kind of entry into the document browsing. In addition one or more source files containing the document contents also were marked up with QAML, a proprietary markup language used by the Quantum Access software. Figure 4 illustrates the major steps for the PC data preparation that was used to produce the application illustrated in Figure 5.
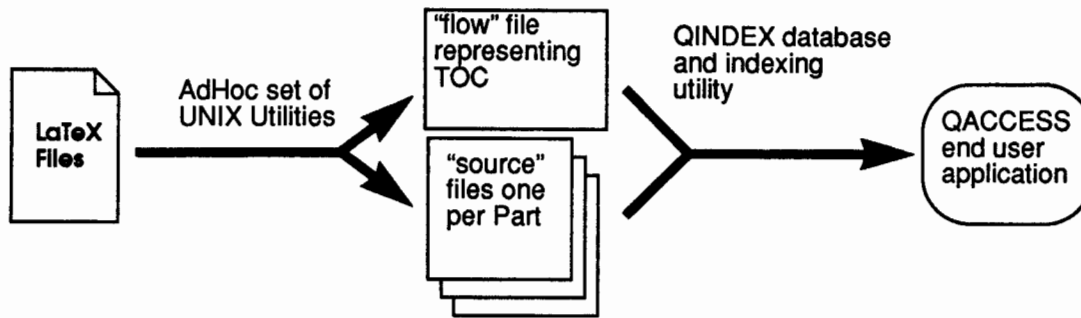
Figure 4 LaTeX to Quantum Access Data Preparation path


```
┌─F1──────────F2──────────F3──────────F4──────────F5──────────F6──────┐
│  HELP    │   FILE   │  SEARCH  │ OUTLINE  │   READ   │   NOTE   │
├──────────┴──────────┴──────────┴──────────┴──────────┴──────────┤
│ STEP/PDES (IPIM)                                                  │
│       STEP Document Parts                                         │
│           Part 41: Miscellaneous Resources                       │
│           Part 42: Shape Geometry Topology                       │
│           Part 43: Shape Interface                               │
│           Part 44: Product Structure Configuration Management    │
│           Part 45: Material                                      │
│           Part 46: Presentation                                  │
│           Part 101: Drafting                                     │
│           Part 102: Shape Structure                              │
│           Part 103: Electrical Applications                      │
│           Part 104: Finite Element Model                         │
│       IPIM (Tokyo Version)                                       │
│           Introduction                                           │
│           General Resources                                      │
│           Type Definitions                                       │
│           Miscellaneous Resources                                │
├──────────────────────────────────────────────┬─────────────────┤
│ Q in File Menu to EXIT, Down/Up arrow keys to move bar. │v2.0 Copyright 1990│
│ Headings displayed in bold text are entity names.       │Quantum Access,Inc.│
│                                                         │Houston, TX 77027  │
└─────────────────────────────────────────────────────────────────┘
```

Initial Table of Contents Screen


```
┌─Winter 1990 NIST Product Data Resource CD-ROM Initial Prototype─┐
├─F1──────────F2──────────F3──────────F4──────────F5──────────F6──┤
│  HELP    │   FILE   │  SEARCH  │ OUTLINE  │   READ   │   NOTE   │
├──────────┴──────────┴──────────┴──────────┴──────────┴──────────┤
│ (d + 2 + k) & = & N & = & (d + 1) + md + (d + 1)                │
│ So the knot sequence is (0, 1, \ldots m, (m+1)) with multiplicities│
│ (d+1, d, \ldots d, d+1).                                         │
│ The setup for surfaces is the same except that m is always 0 as there│
│ is only one patch.                                               │
│                                                                  │
│ EXPRESS Definition                                               │
│ ENTITY b_spline_curve                                            │
│ SUBTYPE OF (bounded_curve);                                      │
│ degree : INTEGER;                                                │
│ upper_index_on_control_points : INTEGER;                         │
│ control_points : ARRAY [0:upper_index_on_control_points]         │
│ OF cartesian_point;                                              │
│ uniform : OPTIONAL uniform_type;                                 │
│ upper_index_on_knots : INTEGER;                                  │
│ knot_multiplicities : OPTIONAL ARRAY [1:upper_index_on_knots]    │
│ OF INTEGER;                                                      │
├──────────────────────────────────────────────┬─────────────────┤
│ Use Down/Up arrow keys to move through the text.│v2.0 Copyright 1990│
│                                                 │Quantum Access,Inc.│
│ B-SPLINE CURVE                                  │Houston, TX 77027  │
└─────────────────────────────────────────────────────────────────┘
```

Display with Individual Entity Definition

Figure 5 Browser using Quantum Leap on a PC

Problem Areas

The following sections describe the major problem areas encountered in creating the heterogeneous platform CD-ROM. These are: name space requirements, device driver differences, end-of-line conventions and platform specific anomalies.

File Name Restrictions

ISO 9660 specifies the allowable file structure, including allowable file names, which may be "mastered" onto a conformant CD-ROM. Files on an ISO 9660 CD-ROM can only have eight upper case alpha-numeric characters, followed by a dot (period) followed by three more alpha-numeric characters. File names must also end with a semicolon and version number. Furthermore, directories may not have an extension and files must have the dot separator.

For example the name HELLO.DIR is illegal as a *directory* name because it contains a file extension, the .DIR. The name HELLO is illegal as a *file* name because it does not contain the required dot for an extension. A file called HELLO will be renamed to HELLO.;1 after the premastering process. Typically it is the premastering process which adds the semicolon version number to the file name. However some CD-ROM device drivers ignore this information (see section on device driver differences below.) Strictly speaking ISO 9660 specifies three levels. Virtually all disks follow the level 1 restrictions. A few use level 2 and level 3 CD-ROMs generally don't exist.

File names on the Macintosh can be up to 31 characters long with almost any character being quite acceptable, including spaces. Since the major portion of development took place on a Macintosh there existed all sorts of files with long names and quite a few spaces. Because of the naming requirements two year old HyperCard stacks had to be altered to be ISO 9660 conformant [9].

One nasty side effect resulting from the semicolon version number name change requirement is that programs which expect to find files of a particular name don't work. This happened with OzTeX, a public domain TeX document processor included on the CD-ROM which expects to find a file called config.sys. The file name was converted, by the premastering software, to config.sys;1. Hard coded names can not simply be changed (... source code may not be available or name a... mecha-

nisms don't exist. It would not surprise me if other hardcoded naming anomalies exist.

CD-ROM Device Driver Differences

Another area for differences among "standard" CD-ROM's is the device drivers which control the CD-ROM readers for each computing platform. The main difference is in how they interpret the file name version information. On the PC the Microsoft CD-ROM extensions ignore the semicolon so applications never know it's there. On the Macintosh, the Foreign File Access for ISO 9660 CD-ROM's simply treat the semicolon version number as part of the file name and applications must deal with them. The CD-ROM has been successfully mounted on a Sun workstation and since its CD-ROM driver ignore the semicolon version number. Although the ISO 9660 standard was used and adhered to, the interpretation of the standard by different device drivers may differ. This will affect the manner in which applications run.

Problems with End-of-Line Characters

End Of Line (EOL) characters are different among different machines and one must deal with them appropriately. I discovered the problem while demonstrating the CD to someone while using a PC. The first thing one does after placing the CD-ROM into the player is to read the README file. I entered the DOS command type to display the file and the stream of characters over wrote each other on the bottom line of the screen.

Since I was gathering all of the files onto a Macintosh, the file transfer process, usually FTP (file transfer protocol), translated the ASCII files into the form of ASCII usually used on the Mac and therefore substituted all CR,LF with CR. The result was that most of the ASCII files on the CD-ROM have EOL characters of only a CR. The files are usable when read into almost any word processor. While the ISO 9660 CD-ROM is readable on many UNIX platforms, it should also be noted that the EOL convention for UNIX is a single LF.

This lack of EOL standards is problematic even in the simple case of placing ASCII files on a CD-ROM. Each platform must have an application which interprets the EOL form of the file on the CD-ROM as a valid file. From a user point of view this may be quite cumbersome. Users must read the ASCII files in almost any type of word processor to view many of the files. No simple solution exists other than one which would put an EOL interpreter into the device driver itself. Otherwise one must either provide

duplicate files with proper EOL's or an application to view the files for each platform.

Macintosh Issues

Another problem area for the Macintosh is how files are represented. The iconic interface of the Macintosh, its look and feel, is one of Macintosh's great strengths. Files on the Macintosh have a number of odd components. The files are structured into two "forks," the data and resource forks. The icon for a file is stored in a resource fork of the application program. The position of an icon in a window is stored with the directory "folder" information. In the premastering process, that information does not fit into the ISO 9660 standard so all icons disappear. Actually they don't go away but simply turn into the generic document and application icons. Double clicking an application or document on the CD-ROM functions as normal because the creator and application ID's are maintained. However this behavior and look is disconcerting and quite ugly as can be seen in Figure 6.

Apple is working on a potential solution to this problem, with advice from Meridian Data. The Macintosh uses a utility called "Foreign File Access" to read files from other types of file systems. The solution is to modify the ISO 9660 foreign file access utility, to extent the use of the XAR (Extended attribute blocks) specified in the ISO 9660 standard for system use. This provides the Macintosh with data storage for the icon, icon positioning and name length to make the CD-ROM appear as a normal Macintosh disk. These *system use* fields of the XAR's are reserved for just such an application and would be an elegant solution to the problem.
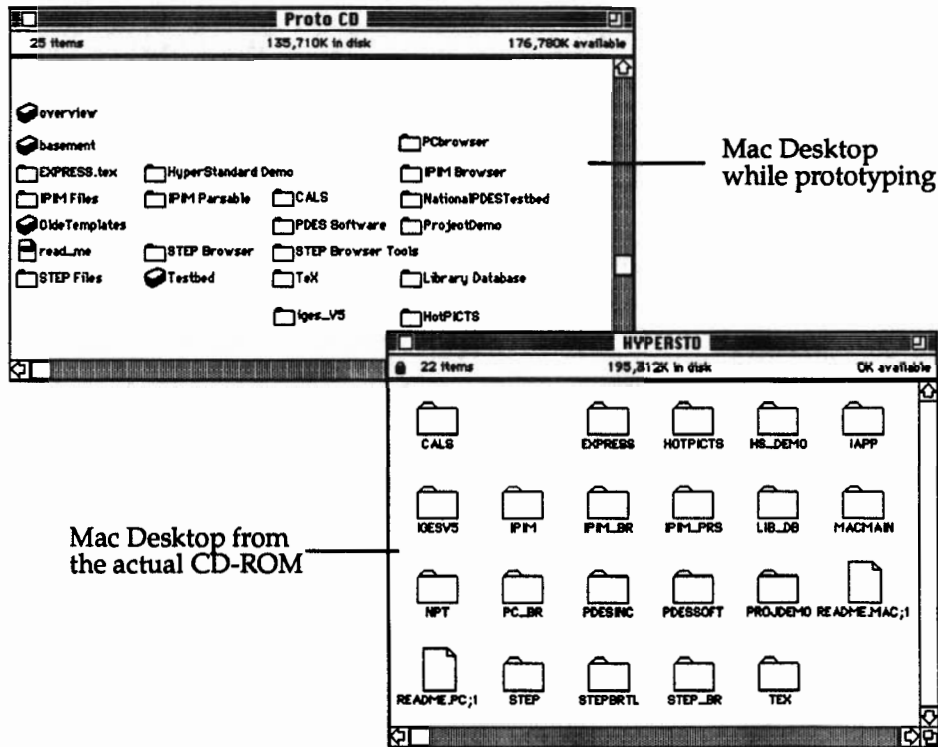

Figure 6 Before and after premastering filename conversion on the Macintosh

Conclusions and Lessons Learned

Multi-platform CD-ROMs are possible given uniform implementations of ISO 9660 and great attention to detail. Providing truly useful software without duplicating the data is difficult. Whatever the application software, the lack of EOL standards will force the inclusion of ASCII file viewers which can interpret the ASCII file correctly for the host platform. Applications which have binary compatible data files will have an advantage in this arena.

Another area where improvement is needed is in the document processing procedures. As LaTeX documents were the source of the content, many implicit heuristics were used to derive structure from these documents. It is clear that the use of SGML (Standard Generalized Markup Language) [3] which would make the document structure explicit, would greatly improve the document conversion process.

Production of this CD-ROM is another example of the "chicken and the egg" problem. Most of the technical experts involved in product data standards do not have CD-ROM players because useful information does not exist on CD-ROM's. It is hoped that this CD-ROM will encourage this technical community to start examining the use of CD-ROMs both as a medium for publishing information and as an aid in technology transfer.

Acknowledgments

References

1. Apple Computer Inc., Macintosh HyperCard User's Guide, 1987.

2. International Organization for Standardization, Information Processing - Volume and file structure of CD-ROM for information interchange, ISO 9660, 1988.

3. International Organization for Standardization, Information processing - Text and office systems - Standard Generalized Markup Language (SGML), ISO 8879, 1986.

4. Knuth, D.E. The TeXbook, Addison-Wesley, Reading, Mass., 1984.

5. Lambert, S. and Ropiequet, S. eds., CD ROM The New Papyrus, Microsoft Press, Redmond, WA,1986.

6. Lamport, Leslie, LaTeX: A Document Preparation System, Addison-Wesley, Reading, Mass., 1986.

7. Quantum Access Inc., Quantum Leap CD-ROM Development System User's Manual, Quantum Access Inc., Houston Texas, 1990.

8. Ressler, Sanford, The HyperStandard: Applying Hypertext Technology to Standards Development, Dissemination and Implementation, NISTIR 88-3859, National Institute of Standards and Technology, Gaithersburg, MD, Sept. 1989.

9. Ressler, Sanford, An Overview of the Product Data HyperStandard CD-ROM Prototype, NISTIR 4470, National Institute of Standards and Technology, Gaithersburg, MD, Nov. 1990.