The International Journal of Digital Curation
Volume 7, Issue 2 | 2012

Developing Virtual CD-ROM Collections: The Voyager Company Publications

Geoffrey Brown,
Indiana University School of Informatics and Computing

Abstract

Over the past 20 years, many thousands of CD-ROM titles were published; many of these have lasting cultural significance, yet present a difficult challenge for libraries due to obsolescence of the supporting software and hardware, and the consequent decline in the technical knowledge required to support them. The current trend appears to be one of abandonment – for example, the Indiana University Libraries no longer maintain machines capable of accessing early CD-ROM titles.

In previous work, (Woods & Brown, 2010) we proposed an access model based upon networked "virtual collections" of CD-ROMs which can enable consortia of libraries to pool the technical expertise necessary to provide continued access to such materials for a geographically sparse base of patrons, who may have limited technical knowledge.

In this paper, we extend this idea to CD-ROMs designed to operate on "classic" Macintosh systems with an extensive case study – the catalog of the Voyager Company publications, which was the first major innovator in interactive CD-ROMs. The work described includes emulator extensions to support obsolete CD formats and to enable networked access to the virtual collection.

International Journal of Digital Curation (2012), *7*(2), 3–20. http://dx.doi.org/10.2218/ijdc.v7i2.226

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

Emulation has been widely discussed as a preservation strategy for digital artifacts, such as multimedia presentations, that are intimately tied to their original hardware/software platform for interpretation (McCray & Gallagher, 2001; Gilheany, 1998; Heminger & Robertson, 2000; Rothenberg, 1995; Rothenberg, 2000b; Rothenberg, 2000a). Emulation has been successfully tested to preserve individual artifacts, such as the BBC Doomsday book project, various multimedia art works (Mellor, 2003; Variable Media, 2004) and is widelyused for the preservation of console games (Guttenbrunner et al., 2010). At this point it is clear that emulation can be used to successfully access software on many obsolete platforms. The fundamental question addressed by this paper is how emulation technologies might be scaled to support convenient access to large collections of born-digital materials.

We have previously proposed a general model for preserving "virtual CD-ROM" collections and explored the use of emulation of Windows based platforms (Woods & Brown, 2010). In this paper, we extend this work to emulation of classic Macintoshes through a significant case study – the CD-ROMs published by the Voyager Company. Although the work required non-trivial modifications to an existing open source emulator, we successfully demonstrate that the fundamental model is both sound and practical for those CD-ROMs that depend upon classic Macintosh environments. This work also explores an important architectural alternative to our previous work. Previously, we provided custom compute environments by using artifact specific customization on a client-side "generic" emulation environment. In this work, we utilize pre-customized server-side emulation environments accessed from the client machine. Both approaches have clear advantages and it is gratifying to demonstrate that the latter also works well in practice.

The Voyager Company led by Bob (Robert) Stein is widely viewed as one of the first and most influential publishers of interactive media on CD-ROM (Virshup, 1996; Young, 2010a; Young, 2010b). Over the period 1989-1997 the company published approximately 75 CD-ROMs, as well as a large number of "extended" books – the latter are the natural ancestors for today's Kindle and other electronic books. Unfortunately, this pioneering work is largely inaccessible to contemporary users and scholars because of its dependence upon obsolete versions of the Macintosh operating system and related software. Indeed, it was a chance meeting with John Eakin, a Professor of English at Indiana University, whose scholarly interest in biography and in particular "The Complete Maus: A Survivor's Tale" [CD-29][1] was the catalyst for the work described in this paper. The Voyager version of Spiegelman's Pulitzer Prize winning work includes the original book augmented through hyperlinks with interviews, videos, as well as preliminary drawings in a delightful interactive CD-ROM. As Professor Eakin noted, while the Indiana University Libraries hold a copy of this work, it has an attached sticker announcing that "We [the Libraries] no longer have systems to support format" (Eakin, 2009).

The historical significance of the Voyager CD-ROMs extends beyond their specific content – many of them are important representatives of interactive documents built

[1] Labels of the form [CD-*i*] refer to the table in the Appendix.

upon the Macintosh application HyperCard, which allowed “non-programmers” to build interactive documents. Matthew Lasar (2012) argues that Hypercard is the “missing link” to the web, demonstrating important connections between the behavior of early browsers and HyperCard. HyperCard was also the underlying technology of Myst, one of the best selling computer games of the 1990s. Finally, Wiki inventor Ward Cunningham has spoken extensively about how HyperCard provided the initial inspiration for Wiki, which he first built as a Hypercard prototype (Cunningham, 2006).

The basic tools necessary to execute (most of) the Voyager CD-ROMs in a software emulation of the classic Macintosh have existed for several years in the form of two open source (and tightly interwoven) projects: SheepShaver[2] and BasiliskII[3]. SheepShaver emulates the later PowerPC based Macintoshes, while BasiliskII emulates the earlier 680xx based Macintoshes. Both are capable of running versions of the Mac operating system that can support the Voyager CD-ROMs. However, there is a significant gap between “capable” and “practical” for the casual user, which might represent an insurmountable barrier for the less technically sophisticated. For example, installing an emulation system capable of executing the Voyager CD-ROMs requires installing (and possibly compiling) the emulator, finding and installing suitable versions of the Mac operating system and system ROM, and finally installing the CD-ROM itself.

A goal of the work described in this paper is to enable easy access to the CD-ROMs for casual users and scholars by developing the technologies required to make network access to a collection of the CD-ROMs practical with minimal software installation on the end-user’s workstation. In particular, we discuss an approach that encapsulates all the necessary operating system and application software in a small, downloadable package, which may be executed on any machine with the necessary emulator installed, and in which the CD-ROMs themselves are accessed over the Internet. The techniques presented make it possible for libraries holding a collection of CD-ROMs to provide public access through modern workstations with no technical knowledge required by patrons. Furthermore, we show how existing distributed file-system software can enable the creation of consortia to pool the resources and technical knowledge necessary to support geographically distributed patrons while enabling strict access controls.

Clearly, this work raises serious legal questions. All of the Voyager CD-ROMs, as well as the Macintosh operating system (and system ROMs), are protected by copyright and it is certainly not legal to provide unconstrained access to these materials. We assume that the best case scenario will require clear controls limiting access to authorized patrons and the technologies described in this paper enable such access controls.

The preservation challenges presented by the Voyager CD-ROMs were well described by Jeff Martin (2010). He provides a clear, basic overview of the history of these publications, interesting observations based upon interviews with Voyager programmers about the underlying software, and the results of testing four of these CD-ROMs under OS 8.5.1 on a legacy Mac, as well as under the “classic mode” that

[2] The official SheepShaver home page: http://sheepshaver.cebix.net

[3] The official BasiliskII home page: http://basilisk.cebix.net.

 doi:10.2218/ijdc.v7i2.226

was part of the early OS X distributions. Most (three) of these examples performed poorly even under 8.5.1, suggesting that support for the Voyager CD-ROMs requires access to earlier versions of the Mac operating system. Martin did not explore emulation as a possible approach to preservation; his testing was limited to four CD-ROMs, compared to the 48 tested for this paper, and he provided no solutions to the 75% failure rate he encountered under OS 8.5.1. In contrast, the methodology described in this paper was successfully used on all 48 CD-ROMs tested (more than half of the Voyager publications), although, as we note, there are a handful of titles whose behavior is “fragile”.

The remainder of this paper is organized as follows. We begin by describing our vision for virtual CD-ROM collections in greater detail and discuss how existing technology can form a viable foundation for creating such collections. We provide a brief overview of the Voyager CD-ROMs with an emphasis on those tested for this work. We then discuss the key emulation and CD-ROM imaging technologies and conclude with a discussion of copyright, sustaining the solution and results.As mentioned above, existing emulation tools can support most of the Voyager CD-ROMs with the notable exception of those utilizing mixed-mode data/audio CD formats. However, these happen to be some of the most novel of the Voyager publications including their first: – a hyper-media presentation of Beethoven’s Ninth Symphony – as well as Pedro Meyer’s “I Photograph to Remember” [CD-2] which was the first CD-ROM with continuous sound and images ever produced (Meyer, 2001). As part of this work, we have extended both SheepShaver and BasiliskII to support these CD-ROMs and describe the necessary changes in this paper.

Virtual CD-ROM Collections

Although the Voyager CD-ROMs have substantial historical significance, they, and most other published CD-ROMs, are destined to have a dwindling user base whose expertise in the systems required to use them is in sharp decline. The physical machines required to execute them have already disappeared from most educational institutions and even the operating systems are increasingly hard to find; at Indiana University, which once had many hundreds of “classic macs”, only one person within our University IT Services had distribution disks of the corresponding operating system software. The physical copies of these CD-ROMs are disappearing from library shelves. In seeking examples for this paper we made extensive use of inter-library loan and we found that many cataloged copies of Voyager CD-ROMs are either missing or damaged.

The long-term probability for individual libraries providing physical access to the Voyager and other published CD-ROMs is nearly nil. The user base is dwindling, the existing hardware and software support disappearing, and the physical media degrading. While we believe these materials have substantial historical significance, their ultimate survival depends upon spreading the preservation burden across many institutions through a virtual collection that enables networked access for a sparsely distributed base of patrons using modern work-stations.

A virtual CD-ROM collection consists of two primary components: one or more servers that maintain bit-faithful images of the CD-ROMs and corresponding support software, and patron workstations with appropriate emulation software installed.

Libraries and educational institutions could collaborate in creating images of CD-ROMs in their collections, as well as customizing supporting software images for these CD-ROMs. In our previous work, we assumed a client-side emulator pre-configured to execute a generic Windows XP environment, and utilized a helper application to customize this environment for a particular CD-ROM (Woods & Brown, 2010). Where emulator environment size is substantial compared to CD-ROM size (three to four times for Windows XP), this represents a substantial space saving. In the present work, we experimented with custom "Mac OS" environments stored on the server and accessed in a non-destructive manner by the client emulator. For this work, the additional storage overhead is only about 20% and hence this simpler model is viable. The advantages to the simpler model are that it is more robust and provides tighter control over materials covered by copyright.

The enabling technology for this vision is a distributed file system. The basic idea is that emulation software is provided with remote access to CD-ROM images through the file system. For example, the emulator "mounts" a CD-ROM by opening the corresponding networked file. The actual bits corresponding to a CD-ROM are pulled to the emulator as needed. In contrast, a web-server based solution would require copying images before access. Consider that a CD-ROM may contain 650 MBytes; copying such a CD-ROM across a network before use could involve a substantial delay. Playing the audio portion of a CD-ROM requires a bandwidth of less than 1.25 Mbits/second. These bandwidth requirements can be met by most DSL connections. We expect a patron in a library would see no discernible performance penalty over local copies of the Voyager materials. Indeed, we store CD-ROM images on the university research file system in Indianapolis, yet work in Bloomington. We perceived few performance issues over an (optimistic) 3 Mbit home DSL connection and no issues over a much faster office connection. Thus, using a distributed file system to store CD-ROM images offers near instantaneous access, while a web-server approach requires users to endure substantial delay. Furthermore, as discussed below, a distributed file system more clearly meets the spirit of copyright restrictions because the CD-ROM images are streamed during access and are never copied to the end-user's work-station.There are many examples of network file-systems including NFS, Samba, WebDAV, and others. We use OpenAFS[4]; however, our work is not tied to this system. In addition to support on all the major operating systems, the major advantages of OpenAFS for virtual CD-ROM collections are:

- Federated authentication (in this case Kerberos),
- Fine-grained access control through access control lists,
- Local volumes,
- Unified name space.

Federated authentication through Kerberos means that any educational institution (currently many) with Kerberos-based user authentication could authenticate their users on behalf of the collection. As mentioned above, libraries must be able to control patron access to specific items. OpenAFS access control to individual files can be specified at both the user and machine level. For example, a library could limit access for their images to specific machines or to specific local users. However, OpenAFS

[4] OpenAFS: http://www.openafs.org

enables inter-library loan by providing a mechanism enabling temporary access to users at other institutions. By storing images on local volume stores, individual libraries can satisfy any mandate on where digital copies may be kept. Finally, a unified name space means that any user can refer to a file by a single global name. For example, the files we maintain are in the (OpenAFS) directory:

```
/afs/iu.edu/home/projects/sudoc/Voyager
```

Any user anywhere with the correct access permissions can read this directory using this name on any machine supporting OpenAFS. No other distributed file-system currently provides all of these capabilities.

Voyager CD-ROMs

The Voyager Company was notable both for its pioneering CD-ROM titles as well as for their broad range. The company produced a diverse range of content including music, movies, books, poetry, and art. Most of their titles are based upon a simple premise: migrate an existing work in another medium to CD-ROM while expanding the original content with audio, video, and background material. In the best of these "migrations", the unique (pre-web) ability of CD-ROM content to rapidly and randomly access audio and video provided a vastly enhanced experience. For example, the audio recordings developed by Robert Winter provide the ability to tie a text analyzing the underlying music or the score itself to specific sections of music. [CD-3] [CD-4] [CD-9]. Some of the most interesting and influential works were developed from scratch for CD-ROM. These include interactive "games" such as Laurie Anderson's "Puppet Motel" [CD-40][5] and the Resident's "Freak Show"[CD-30][6]. While some of the Voyager CD-ROMs have aged well, others, such as those that utilize video, have not. Most suffer from the limitations of the underlying programming technology, which makes them appear relatively primitive by today's standards. Nevertheless, the Voyager CD-ROMs represent an important and pioneering body of work.

The first Voyager CD-ROM, and possibly the first consumer CD-ROM, was an interactive version of Beethoven's Ninth Symphony designed by Robert Winter, a UCLA music professor, and originally released in 1989. An online demo is available (Winter, 2009). As with other Voyager interactive music CD-ROMs, this was a hybrid disk consisting of a single data track and multiple audio tracks. The interactive component was a HyperCard stack. HyperCard was an immensely influential product from Apple that allowed non-programmers to create interactive programs (Smackerel, 2005).[7]

Other hybrid CD-ROMs include "I Photograph to Remember" (IPTR) [CD-2] and "All My Humming-birds Have Alibis" (ALLMY) [CD-5]. Pedro Meyer, the author of IPTR, claims that this was the first CD-ROM with continuous sound and music ever

[5] An online demo is available at: http://blip.tv/file/462115/.

[6] Lynn Ginsburg, writing in WIRED, claimed Freak Show was "hailed as the best CD-ROM ever" (Ginsburg, 1995).

[7] HyperCard was also used to create the immensely popular adventure game Myst (by the Cyan Company), which also works well with the technologies described in this paper.

produced. (Meyer, 2001). Although the work is now available on the web, the CD-ROM provides an interesting opportunity to study the limits of CD-ROM technology. ALLMY is an experimental composition of three "imaginary ballets" composed by electronic music pioneer Morton Subotnick, based upon illustrated novels by the surrealist Max Ernst (Davison, 1996).

Another historically significant Voyager publication is "Who Built America" (Darien, 1998), which supplemented the original printed text with multimedia materials [CD-16] because of a censorship attempt allegedly made by Apple Computer because of its treatment of homosexuality, birth control, and abortion (Reynolds, 1995).

Space restrictions preclude a more in-depth discussion of the Voyager Company materials. We trust this brief discussion has helped to illustrate why preservation of such materials is valuable. As we show through the remainder of this paper, preservation of a large collection of CD-ROMs does not require large-scale technological development.

Emulation

The primary tool used for the research discussed in this paper was SheepShaver, an open source and multi-platform Macintosh emulator. SheepShaver models the later PowerPC-based classic Macintoshes executing System 7.5.2 through 9.0.4. A related program, BasiliskII, models the earlier 600xx based Macs running all operating systems up to 8.1. In this section, we describe these emulation platforms and extensions we made to support mixed-mode CD-ROMs, as well as network support for shared emulation environments.

Since the Voyager CD-ROMs were primarily published 1991-1997, we expected that any of the 7.x operating systems would be capable of supporting the Voyager CD-ROMs and previous work by Martin had provided clear indication that later operating systems are probably incompatible with some of the Voyager CD-ROMs. In testing the Voyager CD-ROMs, we initially used Basilisk II executing System 7.6 (the final, and hence most refined version of the 7.x series); however, we encountered some performance issues (notably executing "A Hard Days Night" [CD-12]) and switched to SheepShaver, which has some significant performance enhancements, executing System 7.6.[8] With the addition of code to support mixed-mode CD-ROMs, all of the Voyager CD-ROMs tested appear to execute correctly in SheepShaver.[9]

Testing consisted of installing and executing each CD-ROM in our emulation environment. We did not attempt to completely execute every aspect of every CD-ROM, but rather to determine with reasonable confidence that they behaved correctly. On average we spent 10-20 minutes with each title exploring their various features. We did not attempt to systematically test each title and we did not compare results

[8] System 7.5.5 is available for free download from Apple and hence would be a good choice for implementing a virtual library using the techniques discussed in this paper.

[9] A possible exception is Blam! [CD-10], which was designed to both take charge of the user's computer and to be annoying. It is hard to be sure if the behavior of this title under emulation is annoying by intent or by virtue of imperfect emulation. Furthermore, a few of the experimental collections ([CD-37], [CD-45]) can best be described as "fragile."

 doi:10.2218/ijdc.v7i2.226

with running the titles on original Macintosh hardware. Our primary goal was to determine whether emulation was viable for a broad selection of titles rather than detailed analysis of specific titles.

Many of the Voyager CD-ROMs (and CD-ROMs in general) require some installation before they can be used. We believe this installation process represents a serious barrier to casual users, hence we have investigated ways to create custom configurations for each unique CD-ROM that an end-user could download and execute. An emulator configuration consists of three components: a binary image of one of the Mac system ROMs, a disk drive image containing an installation of a Mac operating system, and a "preference" file defining the configuration of the emulated machine. The preference file defines the locations of the ROM image, the disk drive image, and any CD-ROM images, as well as other hardware parameters. For each of the Voyager CD-ROMs we created a separate preference file. After starting the emulator with a generic installation of System 7.6, we performed any necessary CD-ROM installation. The resulting hard disk image and preference file together form a configuration that could be cloned for end-users. With a distributed file-system supporting a global name space, there is no need to copy CD-ROM images to an end-user's machine. For example, a preference file linking to one of the Voyager CD-ROMs (e.g. "All My Hummingbirds...") would contain an entry connecting a virtual CD-ROM drive to the corresponding image (allmy.cue).

```
Cdrom  /afs/iu.edu/home/projects/sudoc/Voyager/allmy.cue
```

Cloning the hard drive image requires copying it to the end-user machine (roughly 75Mbytes), or, using a technology we describe later, copying a small "shadow copy" (a few Kbytes) that records only the changes to a networked reference copy. Before describing that technology, it's important to understand the architecture of these emulators and the steps required to extend them.

SheepShaver and BasiliskII are examples of "para-virtualized" emulators,[10] which depend upon modifications to the underlying operating system for their operation. The most complicated aspect of system emulation is accurately modeling the I/O hardware, such as the video monitor, network interface, and storage devices. Para-virtualization provides a significant simplification by modifying those portions of system software that access I/O hardware to allow that functionality to be provided by the host system without the need to model the target system hardware. In the case of SheepShaver and BasiliskII, para-virtualization is achieved by "patching" the ROMs that provide low-level hardware access.

[10] The distinction between emulation and virtualization can be quite fuzzy. Even systems supporting "true" virtualization must emulating I/O devices and often partially emulate the processor itself.

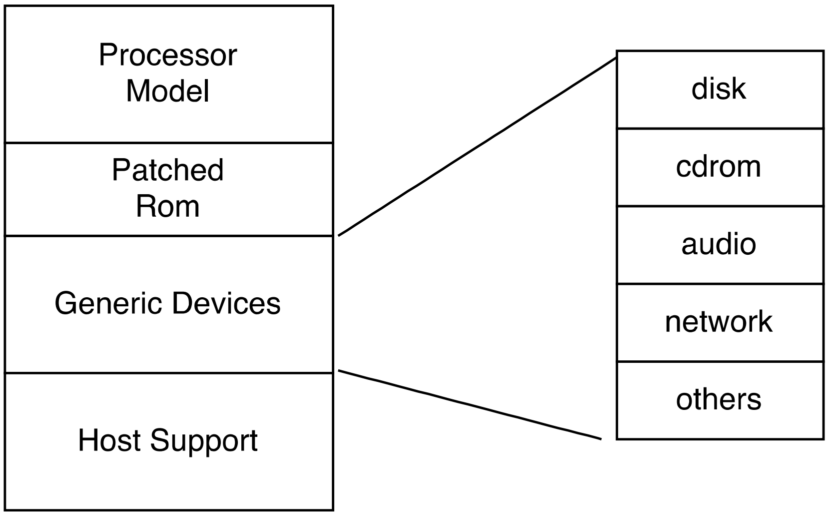

Figure 1. SheepShaver and BasiliskII Architecture.

SheepShaver and BasiliskII are organized approximately as illustrated in Figure 1. A processor model executes native Mac instructions from the application, the operating system, or the patched ROM. Any system call that accesses a device supported by the ROM is diverted to host code implementing generic versions of these devices. For example, SheepShaver provides a single CD-ROM driver. Any of the system calls specified in Apple "Technical Note DV22" (Apple Computer, 1999), which defines the CD-ROM driver calls, is serviced by SheepShaver's generic CD-ROM driver. This, in turn, is supported by host-specific code, which may access host devices (for example, the host CD-ROM hardware) or simulate the equivalent hardware.

Images of hybrid CD-ROMs (indeed audio CD-ROMs) are not supported by most emulation or virtualization environments because the audio tracks are generally handled directly by a CD-ROM, which plays audio under the control of the host processor. In order to support images of hybrid CD-ROMs, we created code that provides a simulated CD-ROM drive for these images. This simulation includes access to the host audio device. Modifications were required to both the emulator subsystem as well as to the virtual CD drive. Although this code is now part of the standard distribution for SheepShaver and BasiliskII, hybrid CD-ROM images are currently supported only on Unix (Linux) and OS X. Adding support for Windows hosts will be considerably easier and can directly utilize the primary module we created.

As mentioned, both emulators simulate the processor (CPU). The simplest, although slowest way to do this is through an interpreter that decodes and evaluates each target machine instruction as it is executed. This approach tends to suffer from poor performance. Both emulators accelerate processor simulation by on-the-fly translation of target machine code to host machine code (a so-call JIT). Thus, performance of both platforms is good on most code. SheepShaver also provides acceleration for QuickDraw (the classic Mac's graphics code), which we believe is the reason it did not have problems with "A Hard Day's Night."

SheepShaver and BasiliskII both use files containing binary images to represent system hard-drives. These hard drive images need to be customized for many of the CD-ROMs to ensure the end-user does not have to navigate software installation procedures. Further, the hard drive images should be cloned for each user so they modify only a local copy. The use of such hard drive images imposes a significant start-up time in copying them to the end-user's workstation. To eliminate this source of delay, we integrated libvhd from the Xen[11] distribution with SheepShaver and BasiliskII. This modification enables the use of hard drive images in the Virtual Hard Disk (VHD) format used by Microsoft Virtual PC, Xen, and VirtualBox. (Microsoft, 2006). VHD images can be layered with a base image and layers providing "differences." Using such layered hard drive images, the base image may remain on the distributed file system while the difference image is copied to the end-user machine. This difference image is initially only a few kilobytes. As the end-user uses a particular image, any disk writes are made only to the local difference image and the base image remains unchanged. Thus, copying a usable emulator configuration to a patron's workstation involves moving only a few kilobytes, consisting of a configuration file and a virtual hard drive difference image from the collection server. Both the CD-ROM image and base virtual hard drive remain on the server and any required data is copied "on-the-fly" as the patron executes the emulator.

In general, we found SheepShaver and BasiliskII to be moderately stable platforms; however, both suffer from unexpected crashes. A close examination of the code base betrays its history as an open source project, which has resulted in a sprawling code base supporting many host platforms in a large variety of configurations. However, the basic architecture is sound and, with a modest investment, they could become viable platforms for preservation of Mac CD-ROMs.

CD-ROM Images

Most CD-ROMs can be saved as bit-faithful images that can be used in place of the physical media. An important exception, which does not apply to the Voyager CD-ROMs, are copy protected CD-ROMs implemented with violations to CD standards. A CD-ROM image consists of a binary file containing the "user data" of the original CD-ROM along with a "table of contents" file replicating the critical metadata from the CD-ROM. In an emulation environment, with an appropriate "device model", this image can be used in a manner that is indistinguishable from the original CD-ROM.

The Voyager CD-ROMs are based upon the Audio CD standard (the "Red Book") extended by the "YellowBook."[12] CD-ROMs are recorded as a sequence of fixed size (2352 byte) sectors, which may contain either audio data or user data. In the former case, a sector contains 1/75 second digitally encoded sound; in the latter case, a sector contains 2048 bytes of data protected by additional error correcting bits. CD-ROMs encode metadata in parallel with the sector data through a mechanism called "subchannels." These metadata include information about the organization of the disk into tracks – equivalent to a song in an audio CD – as well as basic cataloging information about the disk. By convention, sectors are numbered according to their

[11] Xen hypervisor: http://www.xen.org/products/xen_source.html

[12] A good overview is provided by the ECMA-130 standard that parallels the Yellow Book (ECMA, 1996).

temporal position from the beginning of the CD-ROM in units of minutes, seconds, and 1/75 second, even in pure data CD-ROMs where time has no real meaning.

The Voyager CD-ROMs fall into two (physical) categories. Most consist of a single "track" containing an Apple (HFS) file system and in practice behave as read-only disk drives. A small number are mixed-mode CD-ROMs (later called Enhanced CD) with multiple tracks – a data track containing an HFS file system followed by one or more audio tracks. The data-only CD-ROMs present no problems for SheepShaver or BasiliskII in either their physical or image forms. The mixed-mode CD-ROMs do not work on most platforms in their physical form and, prior to our work, were not supported as images.

Imaging either type of CD-ROM is fully supported by existing open source tools. In a Linux or OS X environment, data-only CD-ROMs can be imaged by the Unix program dd. For example, the following command will create a binary image named "cd.bin" of a CD-ROM in the drive /dev/cdrom.

```
dd if=/dev/cdrom of=cd.bin
```

In SheepShaver or Basilisk, this file may be "mounted" by including the following in the preference file read at start up.

```
cdrom cd.bin
```

Similarly, most PC virtualization tools, such as VMWare and VirtualBox, support such binary files directly, although the operating systems running on these tools cannot generally read HFS file systems.

Existing open source tools do support creating images of the Voyager mixed-mode CD-ROMs that can be executed in SheepShaver using the extensions we developed. To create hybrid images we utilized the Linux tool cdrdao. For example, to create table of contents and data files allmy.toc, allmy.bin from a physical CD-ROM, execute:

```
cdrdao read-cd --read-raw --device 1,0,0 --datafile allmy.bin allmy.toc
```

It appears that none of the most widely used PC emulators (VMWare, Virtual Box, Xen, Qemu) support hybrid CD-ROMs. In addition to the Voyager materials, there were numerous hybrid audio discs published over a ten year period.

A Note on Copyright

Throughout this paper we have noted that copyright law may be the single greatest impediment to the preservation of born-digital materials. For the work we have described, there are three entities covered by copyright – the Macintosh ROM used by the emulator, the Macintosh Operating System, and the CD-ROMs themselves. Apple has made versions of the Mac operating system (7.5) publicly available, although the conditions of use are not clear. No version of the Macintosh ROM is freely available. At least these two system components have a single entity with which a library or archive would need to negotiate for creating a virtual CD-ROM collection. A similar situation exists for Windows based emulation, although the existence of commercial emulators suggests that, for the operating system and BIOS, clear use protocols exist. The problem with CD-ROMs is more complicated and the existing laws are unclear, yet it is unlikely that a library or archive could reasonably negotiate with the individual copyright holders. The following describes the laws in the United States, as we understand them.

Under existing law 17 USC § 108, libraries may make three copies of a work for preservation purposes, provided they are not distributed in digital form outside the premises of the library or archives, and this right is exercised to replace a work that is damaged, deteriorating, lost, stolen, or in an obsolete format. (Hirtle, 2003; Section 108 Study Group[13]). In addition, it must be determined that an unused copy cannot be obtained at a fair price.

In practical terms, current copyright law limits the immediate application of the techniques described in this paper to libraries and archives who wish to provide their patrons with electronic access to materials which they own. In the long term, our vision requires networked access to collections of such digital materials. The Section 108 Study Group (Section 108 Study Group, 2008) of the United States Copyright Office has recommended that:

> "The prohibition on off-site lending of digital replacement copies should be modified so that if the library's or archives' original copy of a work is in a physical digital medium that can lawfully be lent off-site, then it may also lend for off-site use any replacement copy reproduced in the same or equivalent physical digital medium, with technological protection measures equivalent to those applied to the original (if any)."

The Digital Millennium Copyright Act (DMCA) appears to complicate the situation where breaking copy protection is a necessary step in creating replacement copies; however, for libraries and archives a preservation exemption has been created for programs or video games distributed in obsolete formats requiring original media or hardware as a condition of access (Copyright Office, Library of Congress, 2006).

A comprehensive international survey of the copyright issues relating to digital preservation was createdby the Library of Congress and others (Library of Congress, 2008).

Preserving the Emulator

A key criticism of emulation is that there are "turtles all the way down". Using an emulator begs the question: "who preserves the emulator?" With emulators such as SheepShaver, which are entirely dependent upon a declining band of enthusiasts, this is a particular problem. One strategy that has been suggested is to build a "universal virtual computer" which can be used as a platform on which to build virtual machines (Lorie & Diessen, 2005). This approach seems hopelessly naïve, as it is a mistake to view a computer as simply the underlying hardware. Rather it is the software infrastructure, including operating systems, compilers and other tools, that are the real

[13] Section 108 Study Group:http://www.section108.gov/about.html

barrier to preservation. Perhaps it would be better to view a standard PC running Ubuntu Linux circa 200x as a reference platform. The hardware platform is well understood and supported by multiple emulators, both open source and commercial. The software platform is available in source form and can be completely bootstrapped with a basic C compiler. Compiling SheepShaver for this platform is relatively easy using the tools available on the platform. Thus, a "preserved" emulator might consist of two layers: our reference platform executing on one of the many x86 emulators and SheepShaver executing on our reference platform. The current code base of SheepShaver supports multiple platforms and is increasingly hard to maintain as these platforms evolve. Forking the code base and supporting a single, stable platform appears to be a more viable approach for an institution wishing to apply this technology to their collections.

Discussion

The major conclusion of this work is that CD-ROMs published for the Classic Macintosh can probably be preserved for access by future generations using existing technology and a networked archive could be created using the additional techniques described in this paper. The specific combination of emulator and operating system, SheepShaver executing System 7.6, proved to be a very good platform for accessing the Voyager Company CD-ROMs. We successfully tested 48 of the approximately 75 published Voyagers CD-ROMs, including instances from the all the years of publication. The full set of tested CD-ROMs is provided in the Appendix. This table also provides a summary of the software provided on each CD-ROM, for example, HyperCard or MacroMind Director, and hence some indication of the technologies required. Most required some form of installation ranging from copying fonts to the system folder to running a provided installer. None required externally provided software in System 7.6; although, when executed under later versions of the Mac operating system they may. Note that a handful of titles – most notably those that were compilations of experimental multimedia art – crash easily. In contrast, the more commercial titles are quite robust.

We believe that many of the Classic Macintosh CD-ROMs produced by other publishers can also be preserved using the techniques we have described. Although we did test a small number of other titles, including the wildly popular game "Myst", our confidence comes from the relatively tight control Apple exerted over its platform during the Classic period.

The SheepShaver platform performs well and is generally easy to use; however, it can be more fragile than is desirable. For example, hard crashes are more common than we recall from our past use of classic Macintoshes. We believe this could be improved with some engineering effort. As mentioned previously, existing emulation tools can support most of the Voyager CD-ROMs with the notable exception of those utilizing mixed-mode data/audio CD formats; we have described emulator extensions to support these.

As we have shown, all of the pieces exist to enable a virtual collection of the Voyager Company CD-ROMs – from distributed file systems to enable sharing across institutions, to the required emulation tools, to any required software. We have successfully extended SheepShaver to support hybrid disk images and to support virtual hard drives, which will allow end-users to download very small CD-ROM specific configurations. On the OS X environment, these configurations open through mouse clicks, which makes their use especially simple; this functionality could be added to the Windows and Linux environments. The code we have added is now part of the standard SheepShaver source distribution; there remains a modest task of extending this to the Windows platform. This approach also represents a significant architectural alternative to the technique we previously used to deliver customized emulation environments. Our approach does ultimately depend upon preserving the underlying emulator; the KEEP project[14] has as its aim the preservation of such environments.

One of the anonymous referees for this paper asked why we don't use Remote Desktop or VNC to offload the client's workstation completely. The short answer is these technologies do not work well with multimedia. For example, VNC doesn't support audio and no remote access technology we have tested works well with video.

While the overall message of this paper can be viewed as positive – no super-human technical effort is required to preserve an important slice of computer history – the clock is ticking. The software required is disappearing rapidly and the set of people with sufficient technical knowledge to bring the existing emulators up to "production quality" is declining.[15] Finally, there are significant legal challenges to building an archive of commercial software.

Acknowledgements

This material is based upon work supported by the National Science Foundation under grant No. IIS-1016967. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

Apple Computer. (1999). *CD-ROM driver calls: Technical report DV22.* Retrieved from http://developer.apple.com/legacy/mac/library/#technotes/dv/dv_22.html

Copyright Office, Library of Congress. (2006). Exemption to prohibition on circumvention of copyright protection schemes for access control technologies. *Federal Register 71*(227).

Cunningham, W. (2006). An evening with wiki inventor Ward Cunningham in conversation with John Gage. http://www.computerhistory.org/collections/accession/102654181

Darien, A. (1998). CD-ROM review: Who built America. *The Journal of Multi Media History, 1*(1). Retrieved from http://www.albany.edu/jmmh/vol1no1/wba_review1.html

[14] Keeping Emulation Environments Portable: http://www.keep-project.eu

[15] During the period in which we performed the work described in this document, Apple removed or made it difficult to access many of the technical documents describing the Classic Mac OS.

Davison, S. (1996). All my hummingbirds have alibis (a review). *Notes, Second Series 53*(2). Retrieved from http://www.jstor.org/stable/900152

Eakin, J.P. (2009). *Reading comics: Art Spiegleman on CD-ROM.* Provided by the author.

ECMA. (1996). *Standard ECMA-130 data interchange on read-only 120mm optical data disks (CD-ROM)*. Retrieved from http://www.ecma-international.org/publications/standards/Ecma-130.htm

Gilheany, S. (1998). Preserving digital information forever and a call for emulators. *Digital Libraries Asia 98 - The Digital Era: Implications, Challenges, and Issues.*

Ginsburg, L. (1995). Twin peaks meets SimCity. *WIRED*. Retrieved from http://www.wired.com/wired/archive/3.09/residents.html

Guttenbrunner, M., Becker, C., & Rauber, A. (2010). Keeping the game alive: Evaluating strategies for the preservation of console video games. *The International Journal of Digital Curation, 5*(1), 64–90.

Heminger, A.R. & Robertson, S. (2000). The digital rosetta stone: A model for maintaining long-term access to static digital documents. *Communications of AIS, 3(1es), 2*.

Hirtle, P.B. (2003). *Digital preservation and copyright*. Retrieved from http://fairuse.stanford.edu/commentary_and_analysis/2003_11_hirtle.html

Lasar, M. (2012). 25 years of HyperCard: The missing link to the Web. *Arstechnica*. Retrieved from http://arstechnica.com/apple/2012/05/25-years-of-hypercard-the-missing-link-to-the-web/

Library of Congress. (2008). International study on the impact of copyright law on digital preservation. Retrieved from http://www.digitalpreservation.gov/documents/digital_preservation_final_report2008.pdf.

Lorie, R. & van Diessen, R. (2005). *UVC: A universal virtual computer for long-term preservation of digital information* (IBM Research Report RJ 10338).

Martin, J. (2010). *Voyager company CD-ROMs: Production history and preservation challenges of commercial interactive media.* Retrieved from http://www.eai.org/resourceguide/preservation/computer/pdf-docs/voyager_casestudy.pdf

McCray, A.T. & Gallagher, M.E. (2001). Principles for digital library development. *Communications of the ACM, 44*(5), 48–54.

Mellor, P. (2003). CaMiLEON: Emulation and BBC doomsday. *RLG DigiNews, 7*(2).

Meyer, P. (2001). *"...Some background thoughts"*. Retrieved from http://www.zonezero.com/exposiciones/fotografos/fotografio/work.html

Microsoft. (2006). *Virtual hard disk image format specification.* Retrieved from http://technet.microsoft.com/en-us/virtualserver/bb676673.aspx

Reynolds, J.F. (1995). *Who built America controversy*. Retrieved from http://h-net.msu.edu/cgi-bin/logbrowse.pl?trx=vx&list=h-mmedia&month=9503&week=d&msg=H8QFstE81ZsTSIZHxLF5XA&user=&pw=

Rothenberg, J. (1995). Ensuring the longevity of digital information. *Scientific American, 272*(1), 42–47.

Rothenberg, J. (2000a). *An experiment in using emulation to preserve digital publications.* Technical report, Koninklijke Bibliotheek.

Rothenberg, J. (2000b). *Using emulation to preserve digital documents.* Technical report, Koninklijke Bibliotheek.

Section 108 Study Group. (2008). *The Section 108 Study Group Report.* Technical report.

Smackerel. (2005). *When multimedia was black and white.* Retrieved from http://www.smackerel.net/black_white.html

Variable Media. (2004). Seeing double: Emulation theory and practice. Retrieved from http://www.variablemedia.net/e/seeingdouble/home.html

Virshup, A. (1996). The teachings of Bob Stein. *WIRED*. Retrieved from http://www.wired.com/wired/archive/4.07/stein_pr.html

Winter, R. (2009). The CD companion to Beethoven's Ninth Symphony: Flash demo. Retrieved from http://www.futureofthebook.org/blog/archives/2009/11/published_by_the_voyager_compa.html

Woods, K. & Brown, G. (2010). Assisted emulation for legacy executables. *The International Journal of Digital Curation, 5*(1), 160–171.

Young, D. (2010a). *Inventing interactive: Early hypercard creativity.* Retrieved from http://www.inventinginteractive.com/2010/04/15/early-hypercard-creativity/

Young, D. (2010b). I*nventing interactive: Voyager (1989-2000).* Retrieved from http://www.inventinginteractive.com/2010/02/22/voyager-1989-2000/

Appendix

Table 1. Summary of the results from testing 48 Voyager CD-ROMS.

Key: **H** uses Hypercard, **P** includes custom program, **F** includes custom fonts, **D** MacroMind Director, **M** Mixed-mode CD-ROM.

Title	Year	H	P	F	D	M
1. Baseball's Greatest Hits	1991	✓	-	✓	-	-
2. I Photograph to Remember	1991	-	-	-	✓	✓
3. Ludwig van Beethoven Symphony No. 9 (multi-language edition)	1991	✓	-	✓	-	✓
4. Mozart String Quartet in C Major	1991	✓	-	✓	-	✓
5. All My Hummingbirds Have Alibis	1992		-	✓	✓	✓
6. Bach and Before (So I've Heard Vol. 1)	1992	✓	-	✓	-	✓
7. Classical Ideal (So I've Heard Vol. 2)	1992	✓	-	✓	-	✓
8. Poetry in Motion	1992	✓	-	-	-	-
9. Richard Strauss Three Tone Poems	1992	✓	-	✓	-	✓
10. Blam![16]	1993	✓	-	-	✓	-
11. Children's Songbook	1993		-	-	✓	-
12. A Hard Day's Night[17]	1993	✓	-	✓	-	-
13. Hikaruhana (Shining Flower)	1993		-	-	✓	-
14. Planetary Taxi	1993	✓	-	✓	-	-
15. Take Five[18]	1993		-	✓	✓	-
16. Who Built America	1993	✓	-	✓	-	-
17. American Poetry	1994	✓	-	-	-	-
18. Antonín Dvořák Symphony No. 9	1994	✓	-	✓	-	✓
19. Beethoven and Beyond (So I've Heard Vol. 3)	1994	✓	-	✓	-	✓
20. Comic Book Confidential[19]	1994	✓	-	✓	-	-
21. Criterion Goes to the Movies	1994	✓	-	✓	-	-

[16] "Hangs" for long periods, may be intended behavior.

[17] Doesn't perform adequately in BailiskII.

[18] Seems a bit fragile – doesn't restart properly without a reboot.

[19] Movie crashes unless version of hypercard on CD-ROM is used.

Title	Year	H	P	F	D	M
22. Dazzeloids	1994	-	✓	-	-	-
23. Defending Human Attributes in the Age of Machines	1994	✓	-	✓	-	-
24. Ephemeral Films	1994	-	✓	-	-	-
25. Exotic Japan	1994	-	-	-	✓	-
26. The First Emperor of China	1994	-	✓	-	-	-
27. For All Mankind	1994	-	✓	-	-	-
28. Macbeth	1994	✓	-	-	-	-
29. The Complete Maus: a Survivor's Tale	1994	✓	-	✓	-	-
30. The Residents Freakshow	1994	-	-	-	✓	-
31. The Society of Mind	1994	✓	-	✓	-	-
32. Stephen Jay Gould on Evolution	1994	✓	-	✓	-	-
33. Amnesty Interactive	1995	-	✓	-	-	-
34. Day After Trinity	1995	-	✓	-	-	-
35. Live From Death Row	1995	-	✓	-	-	-
36. Morton Subotnick's Making Music[20]	1995	-	✓	-	-	-
37. New Voices New Visions[21]	1995	✓	-	✓	-	-
38. Our Secret Century (Volumes 1-4)[22]	1995	-	✓	-	-	-
39. Poetry in Motion II	1995	-	✓	-	-	-
40. Puppet Motel	1995	-	-	✓	-	-
41. Starry Night	1995	-	-	✓	-	-
42. Theatre of the Imagination	1995	-	-	✓	-	-
43. Truths & Fictions: A Journey from Documentary to Digital Photography	1995	-	-	✓	-	-
44. With Open Eyes	1995	-	-	✓	-	-
45. New Voices New Visions 1995[23]	1996	✓	-	-	✓	-
46. Sacred and Secular: The Aerial Photography of Marilyn Bridges	1996	-	-	-	-	-
47. Witness to the Future	1996	-	-	✓	-	-
48. Fun With Architecture 1997	1997	-	-	✓	-	-

[20] MIDI interface is not supported in emulator, but not required.

[21] Collection of multimedia art from a competition – some of these crash easily.

[22] Uses Oracle media objects.

[23] Second collection of multimedia art from competition – some of these crash easily.