RESCUING THE LEGACY PROJECT: A CASE STUDY IN DIGITAL PRESERVATION AND TECHNICAL OBSOLESCENCE

A Thesis
Presented to
The Academic Faculty

By

Leah M. Mickens

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Literature, Communication, and Culture

Georgia Institute of Technology
May 2009

RESCUING THE LEGACY PROJECT: A CASE STUDY IN DIGITAL PRESERVATION AND TECHNICAL OBSOLESCENCE

Approved by:

Dr. Kenneth Knoespel, Advisor
School of Literature, Communication, and Culture
Georgia Institute of Technology

Dr. Rebecca Burnett
School of Literature, Communication, and Culture
Georgia Institute of Technology

Dr. Douglas Harrell
School of Literature, Communication, and Culture
Georgia Institute of Technology

Dr. TyAnna Herrington
School of Literature, Communication, and Culture
Georgia Institute of Technology

Date Approved: April 2, 2009

ACKNOWLEDGEMENTS

I wish to thank the faculty members who served on my committee (Dr. Kenneth Knoespel, Dr. Rebecca Burnett, Dr. Douglas Harrell, and Dr. TyAnna Herrington) and provided insight throughout the research and writing process. I also want to thank Ms. Ruth Einstein for her time and patience in answering questions about the Legacy Project that was installed at the William Breman Jewish Heritage Museum. Mr. Wayne Coleman's responses to questions concerning an installation at the Birmingham Civil Rights Institute were also appreciated. Finally, thanks is extended to my parents and brother for their continued support of all of my endeavors.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AAC	Advanced Audio Coding
ACPI	Advanced Configuration and Power Interface
ADC	Analog to Digital Converter
AIFF	Audio Interchange File Format
ASCII	American Standard Code for Information Interchange
BBC	British Broadcast Corporation
BCD	Binary Coded Decimal
BCRI	Birmingham Civil Rights Institute
BMP	Windows Bitmap
BPI	Bits Per Inch
BWI	Broadcast Wave Format
CAD	Computer-aided Designs
CAMiLEON	Creative Archiving at Michigan and Leeds; Emulating the Old on the New
CD	Compact Disc
CD-DA	Compact Disc-Digital Auto
CD-R	Compact Disc-Readable
CD-ROM	Compact Disc Read-Only Memory
CD-RW	Compact Disc ReWritable
CMYK	Cyan-Magenta-Yellow-Black
CPU	Computer Processing Unit
DAC	Digital to Analog Converter
DAT	Digital Audio Tape

DDS	Digital Data Storage Manufacturers Group
DIMMS	Dual In-Line Memory Module
DLT	Digital Linear Tape
DTD	Document Type Definition
DVD	Digital Versatile Disc
DVD-R	Digital Versatile Disc-Recordable
DVD-RAM	Digital Versatile Disc-Random Access Memory
DVD-RW	Digital Versatile Disc-ReWritable
DWG	Drawing Format
DXF	Drawing Exchange Format
EBCDIC	Extended Binary Coded Decimal Interchange Code
GB	Gigabit
GHz	Gigahertz
GIF	Graphics Image File
HD DVD	High Definition Digital Versatile Disc
HSV	Hue-Saturation-Value
HTML	Hypertext Markup Language
IDE	Integrated Drive Electronics Disk
IBM	International Business Machines
IT	Information Technology
JFIF	JPEG File Interchange Format
kHz	Kilohertz
LTO	Linear Tape-open
LZW	Lempel-Ziv-Welch
Mac OS X	Mac Operating System X

MANS	Media Art Notation System
MB	Megabyte
MHz	Megahertz
MOS	Metal-oxide Semiconductor
MPEG	Motion Picture Experts Group
MPEG-2	Motion Picture Experts Group-Two
MP1	MPEG-1 Audio Layer I
MP2	MPEG-1 Audio Layer II
MP3	Motion Picture Audio Layer III
MP4	Advanced Audio Coding
MS-DOS	Microsoft Disk Operating System
PC	Personal Computer
PDF	Portable Document Format
PERC	Dell PowerEdge
PNG	Portable Network Graphics
QIC	Quarter-Inch Cartridge Drive Standard Incorporation
RA	Real Audio
RAID	Redundant Array of Inexpensive/Independent Disks
RGB	Red-Green-Blue
RPM	Revolutions Per Minute
SCSI	Small Computer System Interface Port
SDRAM	Synchronous Dynamic Random Access Memory
SQL	Structured Query Language
SVG	Scalable Vector Graphics
TIFF	Tagged File Format

USB	Universal Serial Bus
UVC	Universal Virtual Computer
VHS	Video Home System
WAV	Waveform Audio File
WORM	Write Once Ready Many
XML	eXtensible Markup Language
XSLT	eXtensible Stylesheet Language Transformations

SUMMARY

The ability to maintain continuous access to digital documents and artifacts is one of the most significant problems facing the archival, manuscript repository, and record management communities in the twenty-first century. This problem with access is particularly troublesome in the case of complex digital installments, which resist simple migration and emulation strategies. The Legacy Project, which was produced by the William Breman Jewish Heritage Museum in Atlanta, was created in the early 2000s as a means of telling the stories of Holocaust survivors who settled in metropolitan Atlanta. Legacy was an interactive multimedia kiosk that enabled museum visitors to read accounts, watch digital video, and examine photographs about these survivors. However, several years after Legacy was completed, it became inoperable, due to technological obsolescence. By using Legacy as a case study, I examine how institutions can preserve access to complex digital artifacts and how they can rescue digital information that is in danger of being lost.

CHAPTER 1

INTRODUCTION

Since the 1940s, museums, state parks, and other cultural and historical sites have employed interactive audio-visual assets to enhance the learning experience. By incorporating new presentation technology into their exhibits, institutions hoped to make content more accessible to audiences that were accustomed to relating to the world through the mass media channels like radio, film, and records (Mackintosh 2000). In the 1980s, museums began to augment their exhibits with digital installations. Whereas prior electronic forms limited interactivity to pushing a button to hear simple narrative expositions, digital media enabled users to independently explore rich webs of content and draw their own conclusions. This was particularly true of hypertext multimedia presentations, which provided visitors greater access to a cultural site's collections than traditional site-based exhibits (Lands 2003, 2).

Digital exhibits provide exciting new affordances, but they require maintenance and upkeep like any other type of artifact. Unfortunately, museums often assume that electronic objects are intrinsically more durable than "old fashion" analog ones. This cavalier attitude towards digital objects can have disastrous implications for long-term accessibility. Yet, many museums do not consider the level of maintenance that is needed to keep digital objects accessible for more than a few years. Tim Au Yeung, manager of Digital Object Repository Technology at the University of Calgary, notes the following:

> For some the digital object is the result of a project that has been completed or it may be a temporary surrogate for access purposes. For others, since the digital object only represents a small part of the overall organization, it is not given the considerations as other larger parts of the organization are given. In either case,

> digital objects may often be considered more peripheral to the organization and thus in greater danger of being lost as a result. (Yeung 2004b, 7)

In reality, digital assets require more work to keep them functional than traditional physically instantiated artifacts. Information recorded in analog media like paper, stone, or clay persists over time, until an outside force intervenes (e.g., natural disaster, pests, human error). In contrast, digital information quickly becomes inaccessible unless the software and the hardware needed to access it is kept in working order (Besser 2000, 156). Although the library science community has been studying the long-term preservation of digital documents, most of their research has focused on simple files (e.g., emails, word processing files, bitmap images), not the complex digital artifacts found in most museums (Yeung 2004a).

The library science community has struggled to address long-term digital preservation because it continues to use an outmoded paradigm of archival theory. Modern archival theory arose in the mid-nineteenth century as a way to keep track of the paperwork created in the wake of the steadily expanding bureaucracy. It was influenced by a number of other disciplines, especially diplomatics (the study of the archival properties of handwritten documents), library science, history, and management and organizational history (Gilliand-Swetland 2000,7). Unfortunately, the standards in these fields are often ill-suited to the new world of digital artifacts.

I experienced the challenges of digital preservation during my archival internship at the Breman Museum in 2008. It was here that I discovered the Legacy Project. The Legacy Project was a multimedia, site-based digital exhibition about Holocaust survivors who settled in the metropolitan Atlanta area. Although it was created in 2003, it was already inoperable by the year 2005. I became curious as to how this unique resource had fallen into disrepair and how it might be rendered operable.

This thesis provides two important contributions to the discourse on digital preservation. First, it uses the Breman‘s Legacy Project to illustrate the defects in traditional archival theory as applied to complex digital objects. Second, it proposes a new archival methodology that can be used for long-term digital preservation. Using the Legacy as a case study, I examine how institutions can preserve access to new digital artifacts and rescue old digital information that is in danger of being lost.

This thesis has seven core sections:

- Chapter 2 defines technical terms from the library science and archival fields that are used throughout the document.
- Chapter 3 provides a literature review of digital preservation strategies. The review reveals the existence of a substantial body of literature about digital preservation. However, most of the information emphasizes record management. These techniques are inadequate for addressing problems specific to complex digital objects.
- Chapter 4 explains how digital information is rendered in a computing environment. Chapter 5 describes how digital storage devices operate. This background information prepares readers for the case study presented in later chapters.
- Chapter 6 examines the current state of research in the field of long-term digital preservation. The four most promising methods of long-term digital preservation (hardware preservation, migration, emulation, and XML) are described and the

pros and cons of each technique are explained. This section enables readers to understand why it is so difficult for the library science community to come to a consensus on how to preserve endangered digital files.

- Chapter 7 provides an in-depth description of the Legacy Project, including its history, its technical specifications, and an explanation of what went wrong. I also compare the Legacy Project to the Birmingham Civil Rights Institute's Richard Arrington, Jr. Resource Center. Although both projects were produced by the same company and had similar technological specifications, the latter did not fall victim to technological obsolescence.
- Chapter 8 discusses some of the issues raised by the Legacy Project, namely the use of reinterpretation as a form of digital preservation, the role of digital installations in the museum setting, and the question of how to preserve web sites.
- I conclude my thesis in Chapter 9 with a summary of the changes that need to be made in the archival field to ensure the long-term maintenance of complex digital files.

CHAPTER 2

TERMS, CONCEPTS, AND DEFINITIONS

Before delving into the details of how digital information is to be preserved, it is important to understand what the terms “digital,” “information,” and “preservation” mean. Although these terms are commonly used in the popular press in a variety of contexts, they are technical terms with precise definitions. Similarly, it is also crucial for the reader to become acquainted with the definitions of some related terms – *analog, electronic, data, record, manuscript, document, archive, library, and conservation* – that are used in this work.

Analog systems are those in which the values are in a permanent state of flux. For example, the minute hand on an analog clock is in continuous motion. Consequently, Hunter describes analog reality as “a wave, a flowing continuum of choices.” In comparison, digital systems use discrete values in which data is classed into distinct, predefined categories. In a computing environment, digital data is rendered into binary code. Digital representations are either discrete (text, numbers) or continuous (sound, images, video) (Hunter 2000, 15).

The words *digital* and *electronic* are often used interchangeably to describe information that is not physically instantiated, but these terms refer to very different concepts. When information is in an electronic form, it is stored using the movement of electrons. Thus, it is possible for information to be electronic without being digital, while digital information can be represented using non-electronic forms. For example, VHS tapes are electronic, but not digital. Similarly, the Braille writing system encodes text and numbers using a non-electric digital system (Hunter 2000, 3). For the purposes of this paper, the term *digital* refers exclusively to information that originates and exists within a computing environment.

Similarly, the terms *data* and *information* have different meanings, but are often used interchangeably. *Data* refers to raw intellectual content. *Information* is *data* that has been organized in a specific way to ascribe meaning to the user. A *document* is a string of structured information that provides context and narrative (Hunter 2004, 4). Thus, the contents of a digital document include the intellectual content, combined with the relevant metadata that describes the document (Borghoff 2005, 10). Finally, the terms *records* and *manuscripts* refer to specific types of documents. A *record i*s a government document. A *manuscript* is a non-governmental document. As both records and manuscripts are considered documents with enduring historical or cultural value, they are stored for posterity in government archives and manuscript repositories respectively (Hunter 2000, 4).

It is important to note the difference between digital libraries and digital archives. A digital archive is responsible for safeguarding the long-term access of society's cultural output that exists in digital forms, whereas a digital library provides access to digital information, but may not be responsible for maintaining long-term access to said information (Hunter 2000, 4). However, given the ephemeral nature of digital information, digital librarians may become digital archivists to some degree to ensure that their digital assets remain accessible over long periods of time.

Finally, there is a difference between *preservation* and *conservation*. Deegan and Tanner define preservation as:

> the continuous process of creating and maintaining the best environment possible for the storage and/or use of an artifact (sic) to prevent damage or degradation and to enable it to live as long a life time as possible (Deegan and Tanner 2006, 3).

Preservation is a broad term that encompasses a number of activities, including the acquisition of new collections, appraisal, description, and arranging (Hunter 2000, 2). *Conservation* is a subcategory of preservation. The primary objective of conservation is to restore damaged artifacts to a state in which they are functional and capable of being reproduced. Conservation also involves reversing previous restoration attempts that were either unsuccessful or harmful. Consequently, conservation activities should always be reversible and should leave the artifact's physical, chemical, and historical properties unchanged, should the procedure need to be undone at a later date (Hunter 2003, 170).

CHAPTER 3

LITERATURE REVIEW

Literature concerning long-term digital preservation principally examines the issue from a records management perspective. Records management is a subset of library science that focuses on protecting the information capital of a particular agency or organization. These assets consist of paper-based records, databases, emails, web site content, and information stored on computers and storage devices (ARMA International 2009). The assumptions implicit in the records management perspective include the notion that preservation decisions are made at the end of its life cycle, and that records can be stored indefinitely with little to no damage to the physical or intellectual integrity of the records, neither of which are true in the context of digital files (O'Shea 1996). It is not surprising that most digital preservation literature would adopt a records management perspective, given that the primary types of digital objects generated by organizations and individuals resemble traditional paper-based records (e.g., e-mails, word processing files, images) and can often be printed out and treated as such. However, the literature seldom provides information about preserving digital objects that cannot be migrated to paper, such as web sites, video games, and digital art. Consequently, few resources are available for archives and museums that are seeking assistance for maintaining the more complex digital objects in their custody.

The Problem with Digital Obsolescence and Media Decay

A recurring theme in the literature is that technical obsolescence is an ever present threat to the continued existence of digital files. Digital information – whether "digitally born" or digitized – increases access to content, but being in a digital form markedly decreases the lifespan of information (Rothenburg 1999, 1). Digital files require a specific combination of hardware, software, and peripheral device to make the information intelligible to humans. If one of these components is lacking, the information remains unreadable (Rothenberg 1999, 2). For this reason, the Commission on Preservation and Access characterizes the lifespan of digital information as "nasty, brutish, and short" due to "rapid changes in the means of recording information, in the formats for storage, and in the technologies for use (Task Force 1996, 2)." Similarly, Gagnier and his colleagues warn that if cultural institutions do not act quickly, technical obsolescence will render digital objects extinct (Gagnier et al. 2008). The problem is not that digital information is innately impaired, but that because of rapid innovations in hardware, software, and standards, a file's bitstreams may not be readable for very long after its creation (Deegan and Tanner, 2006, 6).

Besser characterizes the current status quo as a "Tower of Babel," given the number of mutually incompatible permutations of hardware, software, and data formats (Besser 2000, 157). This "Tower of Babel" environment is aggravated by the fact that digital data formats are becoming increasingly intermixed. With analog information, certain types of information are associated with certain formats: text was written on paper, audio recorded on vinyl, images painted on canvas or developed on photographic paper, and moving pictures were shot on celluloid film. In the modern digital environment, it is not unusual for a single digital file to contain a mixture of text, audio, images, and moving pictures. Within that file, there may be more than a dozen standards (e.g., ASCII, JPEG, MPEG-2) used to render each type of information. If one of these formats becomes obsolete, the context of the entire file may be lost (Lesk). Chapter 4 will discuss in-depth the various ways in which digital information can be rendered.

The decay of storage media is also a problem for the long-term maintenance of digital information. The life span of storage media for digital information (e.g., magnetic tape, magnetic tapes, optical disks) is measured in terms of single or double digit years, not centuries as is the case with paper and microfilm (Deegan and Tanner 2006, 15). Storage media degradation is not considered to be as much of a threat to the longevity of digital information as technological obsolescence, because files can be copied to new storage mediums if the old devices show signs of wear (Besser 2000, 155). However, many storage devices for digital information do not show obvious physical signs of decay until it is too late to save the information. This is particularly true for files that are not accessed often (Jones and Beagrie 2001, 19). (Chapter 5 discusses in-depth the physical, technological, and archival properties of storage devices.)

Competing Methods of Digital Preservation

While there is agreement in the library science profession that the long-term preservation of digital files is problematic, there is no consensus about how this situation is to be rectified. One of the biggest sources of contention surrounds how digital files should be preserved. Migration and emulation are generally viewed as the two most promising methods of long-term digital preservation. Migration is the periodic transfer of files from an obsolete format to one that is compatible with modern computers. Emulation refers to a type of software program that replicates the operating system of obsolete computers on contemporary hardware (Besser 2000, 160).

Strengths and weaknesses are associated with each method, but many authors agree that migration is the best, if not the only practical method of digital preservation. According to the Dutch Digital Preservation Testbed, “organizations are beginning to turn to migration either as an interim solution, or as a total solution” (Digital Preservation Testbed 2001, 16). Since migration has a long history in the field of information technology, this method requires a small learning curve for the majority of archivists and information technology (IT) specialists (Borghoff et al. 2005, 14). Migration’s popularity also stems from the perception that it is the cheapest, most straightforward way to extend the accessibility of digital files, even when the process unintentionally corrupts file bit streams (Deegan and Tanner 2006, 19).

Despite the widespread view that emulation is not a feasible option for long-term preservation, this approach has many defenders. For example, Oltmans and Kol reject the notion that emulation is too expensive to be practical and argue that it may be cheaper than migration. Based on previous research into the Universal Virtual Computer (UVC), Oltmans and Kol estimate that emulator development would cost about $150,000 and implementation approximately $20,000. Although they admit that it is unknown how much it might cost to access a file using an emulator, Oltmans and Kol propose that a single emulator could be shared by multiple repositories, which would defer expenses. In comparison, migration requires that every digital file in the repository be converted to a new format on a regular basis; as more digital files accumulate, the costs of migration rise, whereas most of the expense for an emulator is in the initial development (Oltmans and Kol).

The most notable proponent of emulation is Jeff Rothenberg, who has created a substantial body of work on this topic. Rothenberg criticizes the assumption of many institutions that they must accept migration as a default (Rothenberg 1999a, 13). He points out that migration often alters digital documents in ways that affect their authenticity, readability, and formatting (Rothenberg 1999, 14). Rothenberg also notes that paradigm shifts in computer programming can render efforts at migration difficult, if not impossible (Rothenberg 1999b, 13). Responding to claims that emulation is too experimental to be of practical use, Rothenberg has conducted a number of experiments to demonstrate the reliability of emulators as a long-term digital preservation method. More recently, Rothenberg was part of a team of researchers that used emulation to successfully recover a previously inaccessible piece of media art from the 1980s called *The Erl King* (Rothenberg 2007, 9). Despite Rothenberg's belief in the inherent superiority of emulation over migration, he acknowledges that more research needs to be conducted to explore fully the potential of emulation as a viable long-term digital preservation method (Rothenberg 2000, 69).

Some authors take a more nuanced view of emulation. Wheatley states that while migration is best for simple digital files, emulation would be best for more complex digital objects like digital art, video games, and dynamic images (Wheatley 2001). Granger contends that emulation can be used to archive digital information as a short-term solution, but is doubtful that emulators can be used as a long-term preservation method. Granger also disagrees with Rothenberg's assumption that emulation will one day act as a "one size fits all" method for all types of digital files (Granger 2000).

However, within the last ten years, there have been a number of experiments in which digital files that were once considered inaccessible have been retrieved using emulator technology. The CAMiLEON (Creative Archiving at Michigan & Leeds: Emulating the Old on the New) Project, which is a joint endeavor between the University of Michigan and the University of Leeds (UK), successfully emulated the obsolete computer software needed to open the disks of the 1986 Domesday Project. This effort was a digital survey of the United Kingdom that was created by the British Broadcasting Company (BBC) in 1986 to celebrate the 900th centenary of the famous medieval survey. Millions of people, including children, were recruited to produce stories, maps, and photographs about their towns, schools, and regions. The result was an interactive, multimedia, computer-based record of British life in the 1980s. However, soon after the Project was completed, the hardware (the Acorn Microcomputer) became obsolete, making it impossible to read the video discs that contained the survey. The CAMiLEON Project successfully recovered the data from the disks through an emulator that mimicked the original interface of the Acorn Microcomputer (Darlington, Finney, and Pierce).

Another successful experiment that uses emulators as a method of long-term digital preservation is the Dioscuri emulator that was jointly developed by the Koninklijke Bibliotheek and the National Archief of the Netherlands. Dioscuri was developed with the explicit intention of creating an emulator that would be used within an "operational digital archiving environment" (van der Hoeven 2007, 23).

> [Dioscuri is] a modular emulator…[that] successfully runs 16-bit operating systems like MS-DOS and applications such as WordPerfect 5.1, DrawPerfect 1.1 and Norton Commander. Furthermore, it is capable of running many nostalgic DOS-games and a simple Linux kernel.

Most important, Dioscuri is downloadable as open source software for institutions and laypersons that need to create digital environments to read files written in the applications that the emulator supports (van der Hoeven 2007, 24). The development of Dioscuri and the other successful experiments mentioned above suggest that emulation technology may soon come into its own as a commonly used method of long-term digital preservation.

Although migration and emulation are the digital preservation methods most cited in the literature, mark-up languages, particularly eXtensible Markup Language (XML), are also being investigated. The primary advantage of XML is that it is independent of any specific platform (Digital Preservation Testbed 2002, 25). Consequently, a file rendered in XML would not have to be migrated each time a new version of software or hardware was developed (van Nispen et al. 2005). Because XML uses a plain text, Unicode base, it is capable of rendering almost any written character in any language. XML documents can also be converted into other (text-based) formats using XSLT (Barnes 2006, 8). However, it is unclear whether XML will remain in use long enough to be a truly universal standard or if more complex digital objects can be preserved using this method (Barnes 2006, 9).

Regardless of the type of preservation method being discussed, most articles focus on the preservation of digital records (e.g., emails, word processing files, spreadsheets), rather than more complex digital objects. This occurs because the bulk of the digital files contained in libraries, archives, and museums are not multimedia in nature as are more complex digital objects. However, the question of what is the most effective way of preserving digital objects that combine text, audio, images, and moving images will become a more pressing concern as institutions begin accumulating multimedia files. Dynamic websites, media art, and multimedia CD-ROM references are a few of the complex digital objects whose preservation needs must be addressed by the archival and computer science communities. The technology of digital preservation is discussed in greater detail in Chapter 5.

Literature about Digital Preservation in the Museum Context

Few resources are specifically related to preservation in the museum context. When digital preservation in museums is discussed, it usually focuses on the preservation of digital art. Digital art is a broad term that can apply to moving images, multimedia, computer generated art, and hypertext. Because these works are dependent on specific pieces of hardware to provide context, digital art has more in common with ephemeral forms of art like performance art, rather than traditional physically instantiated forms of artwork, such as painting, sculpture, or illuminated manuscripts (Besser 2001). Many pieces of digital art possess interactive, participatory, dynamic, and customizable qualities that require the audience to engage in physical contact with the installation. This contact transfers dirt, grime, and oil to the artwork, which decreases its lifespan. However, unless curators and archivists actually use the piece and risk incurring further damage, it may be impossible to determine the artifact's scope, a quality that is not always apparent from simply looking at a piece.

Rinehart proposes a partial solution to the question of how to preserve digital art with his Media Art Notation System (MANS). MANS is a three-tiered system. The first tier consists of a Score (the name that Rinehart assigns to a particular instance of the notation system) that contains high-level metadata and a nominal amount of XML tags.

The second tier contains more detailed descriptions about the sub-components of the work in question. The structure of the piece is expressed through a detailed XML schema and descriptors that have actual text, images, audio, or moving images from the work in question. As the most complex tier, the third level incorporates "technical metadata, choices that model every behavior of the Work, very granular description and structural markup to the level of individual Resources, and inline bitstreams or linked source files that comprise the work itself." The rationale for the level of detail in the third tier is that future curators could recreate digital art from the metadata, files, and markup contained in the Score (Rinehart 12).

Articles and best practices guides about digitization projects are common, but not those related to digital installations or artifacts. Digitization refers to the process of converting analog information into a series of zeros and ones to make the information intelligible to a computer (Hughes 2004, 4). Therefore, a digitization project is one that makes digital surrogates of analog information and/or artifacts available to the general public via a web site or a storage device, such as a DVD or CD-ROM (Hughes 2004, 6). There are numerous advantages to enabling digital access to a museum or library's collections. Digital information can be accessed without the need of a human intermediary, such as a reference librarian. Full-text search options and sophisticated search engine algorithms make finding digital information easier than pouring over a card catalog (Conway 2000, 6). However, digitization projects are very expensive and the costs are increasing. A viable business model has yet to be developed for institutions to profit from digitization projects (Conway 2000, 7). The costs of these projects are so high, that it behooves institutions to invest in digital preservation, lest the entire investment is lost by technical obsolescence or hardware failure (Jones and Beagrie 2001, 22). These best practice guides assume that each digital object is saved in a single discrete file, each with its own preservation needs. However, complex digital objects are comprised of dozens or hundreds of separate files. Hence, these materials may not be helpful for archivists seeking to preserve multimedia digital works.

Conclusion

Although a substantial body of literature on digital preservation is available, there is a dearth of material about the preservation needs of complex digital objects. Most of the existing literature is silent about practical steps that archivists and librarians can use to save the complex digital assets in their custody. Consequently, archivists, librarians, and museum curators embarking on their own digital preservation or digitalization projects have little information about how to avoid the problems that have plagued many previous endeavors or how to rescue threatened artifacts. More case studies need to be conducted to provide empirical evidence about which long-term digital preservation methods are appropriate for differing types of digital objects. In addition, more theoretical works are needed to address how digital files fit into the larger scheme of archival theory.

CHAPTER 4

THE SCIENCE OF DIGITAL INFORMATION

To understand the problems inherent in long-term digital preservation and why these technologies pose such a challenge to traditional archival theory, it is necessary to provide a brief explanation about how digital information works. Unlike information recorded in analog mediums, digital information requires two sets of external decoders to render it understandable to human beings: a viewing environment to project the data (software) and a hardware system to run the application. Digital information is usually placed on external storage devices that can only be used with a specific type of computer and/or operating system. Thus, even if the data for a particular digital file is intact on a storage device, the likelihood of finding a working computer and peripheral system that can read the device is unlikely (Besser 2000, 157).

In comparison, physically instantiated objects, such as books, stone obelisks, and murals, are immune from technical obsolescence. While ways of presenting analog data may change, such as the transition from the scroll to the codex, the information recorded on them persists until an external cause, such as a natural disaster, pests, or human error destroys the medium. Analog documents written in languages that are believed to be indecipherable, as in the case of the Rosetta Stone, may one day be rendered readable by continued linguistic research. To understand the difference between analog and digital artifacts, consider the difference between a twelfth century Bible and a word processing document from 1984 that is stored on a 5.5 inch-floppy disk. Assuming the Bible is still in reasonably good condition, it will be as readable to a modern scholar as it would have been to his or her medieval counterpart. However, the contents of a 5.5-inch floppy disk that is formatted to be read on an Apple IIe computer will not be readable, unless the discoverer has access to a working version of said computer (including a printer). Given the millions of files that are currently stored on old computers and storage devices, Dioscuri is (Besser 2000, 156).

Because the nature of digital information and analog information differs, archivists seeking to maintain long-term access to their digital assets must understand the specific preservation needs of these files. Archivists and museum curators cannot assume that digital exhibitions have the same properties as their analog counterparts or that digital assets are somehow immune to the ravages of time that afflict physically instantiated artifacts. This section explains how different types of information are represented in a digital environment. The first part discusses the nature of files, the generalized term that refers to all digital objects. Part two explains how numbers are derived from binary code. The third part describes how different coding schemes are used to represent text. The fourth part details how digital images work. The fifth part explains how digital audio is produced and how analog sound can be converted into a digital format. Each part concludes with brief descriptions of the most common formats used to render the information type under discussion.

Files

All digital documents, regardless of type or format, are known as files. Files are classified by the type of information being stored (e.g., text files, images file, audio files). Digital file formats are either proprietary or non-proprietary. Proprietary formats are associated with a particular company and/or operating system. Unless users explicitly tell their computer to do otherwise, most files when saved will be in the proprietary format of the application being used. The licensing agreements of most proprietary formats limit how users may utilize files saved in a particular format. However, in many instances the most popular proprietary formats can be read on competing applications. Non-proprietary formats, also known as neutral formats, enable the exchange of files between different applications and operating systems. However, when saved, many non-proprietary formats are converted into proprietary formats so the computer can better process them (Saffady 2002, 53).

Numbers

All computers process data as combinations of 0s and 1s, known as binary code. This occurs because engineers found it easier to build electronic devices that can assume two stable states, an on switch (1) and an off (0) switch. A singular 0 and 1 is a called a bit, which is short for binary digit. Eight bits comprise a byte (Deitel and Deitel 2007, 688). Binary code uses base 2, and numbers are rendered with different combinations of 0s and 1s in eight bit combinations. This is similar to the more commonly used base 10 system, which employs ten discrete symbols (i.e., numbers) and placeholders to represent numbers of increasing magnitude. The following chart illustrates how the value of bits increases in binary code:

128 =	64 =	32 =	16 =	8 =	4 =	2 =	1 =
2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0

Table 1: Example of bit values in the binary code system
(Hunter 2000, 17).

Hence, a computer renders a number by assigning it an eight-bit binary code. For example, “10011011” is equivalent to 161 in base 10. However, “10011011” could also be used to represent text, images, or sound, depending on how the computer is programmed to interpret the code (Hunter 2000, 18).

Digital Text

Since it is difficult for humans to read raw binary code, data is rendered for human use in the form of letters, numbers, and special symbols, known as characters (Deitel and Deitel 2007, 688). As noted earlier, each character is represented by a fixed series of 8-bits. Thus, a character is a single byte that can be rendered as a number, letter, or any other symbol encountered on the typical QWERTY keyboard (Saffady 2002, 54).

Coding schemes determine how each bit sequence should be interpreted. The most common of these is American Standard Code for Information Interchange (ASCII). When it was first introduced, ASCII consisted of 128 7-bit characters. An enlarged superset of ASCII, known as the ASCII extended character set, which contained 256 7-bit characters, was later introduced. This newer coding scheme includes characters specific to European languages, special mathematical symbols, and geometric shapes (Saffady 2002, 54-55).

Introduced in 1964 by IBM, extended binary coded decimal interchange code (EBCDIC), was the first coding scheme to use 8-bit sequences for character representation. Unlike ASCII, EBCDIC uses an 8-bit encoding system that is known as packed decimal representation. Packed decimal representation enables two characters to be combined in a single byte and is used more for data representation than text rendering.

EBCDIC is primarily used for text files generated or read by IBM and plug-compatible mainframe computers (Saffady 2002, 55).

A newer coding scheme that is gaining in popularity is Unicode. Unicode, which is comprised of over 94,000 characters, was developed to remedy the deficiencies in ASCII, namely the latter's inability to read non-Latin writing systems. Unicode contains codes for every language in the world, including Braille, Cyrillic, Greek, Arabic, Hebrew, Chinese, Korean, and Japanese. It also includes characters for diacritics, arrows, symbols, and dingbats (Unicode, Inc.). Since Unicode is currently supported by the latest versions of every major operating system, including Windows, Mac OS/X, and UNIX, there is a high likelihood that it will supersede ASCII in the near future (Saffady 2002, 55).

Although variants of ASCII and Unicode are the most commonly used standards, many archives may have older digital and electronic records that contain other encoding schemes. For example, binary coded decimal (BCD) code was recorded on seven-track tapes and utilized by IBM mainframes prior to the development of EBCDIC. BCD utilized only 6-bits, with an extra bit occasionally added to check for errors. However, it is unlikely that information recorded in BCD is accessible to modern archivists, as both seven-track players and the necessary software needed to read the tapes have not been manufactured for more than 45 years (Saffady 2002, 55-56).

ASCII Text Files

The simplest type of text files consist of ASCII text files. These files use ASCII characters to represent alphanumeric information, but are unable to support extensive formatting or the insertion of non-text elements. Because ASCII text files are not proprietary, such documents are easily accessed across platforms. Although word processing programs give users the option of saving their documents as ASCII text files, the formatting may be lost when the files are opened in other applications. Consequently, important details of a document like carriage returns, indentations, and spaces between words may need to be reinserted if an ASCII file migrates from one word processing program to another (Saffady 2002, 59).

Word Processing Files

Because raw information stored in coding schemes is unformatted, an application must be created that renders text files that provide structure and context for human users. One of the most common ways of formatting these files is to use a word processing program. These programs take the characters in a text file and add an additional set of embedded characters to provide additional formatting, such as page breaks, fonts, bold, italics, underlining, superscripts, and subscripts. Although the text characters in a word processing program are generated using ASCII, the specific embedded characters that produce document formatting are specific to each word processing program. For example, documents created in Wordstar, which was the most popular word processing program of the 1980s and early 1990s (see Figure 1), cannot be accessed using Microsoft Word (Saffady 2002, 56).

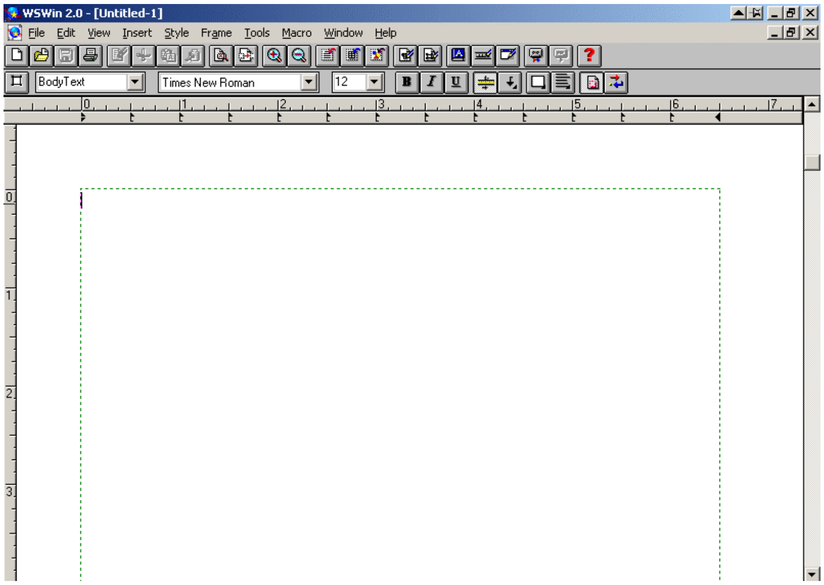


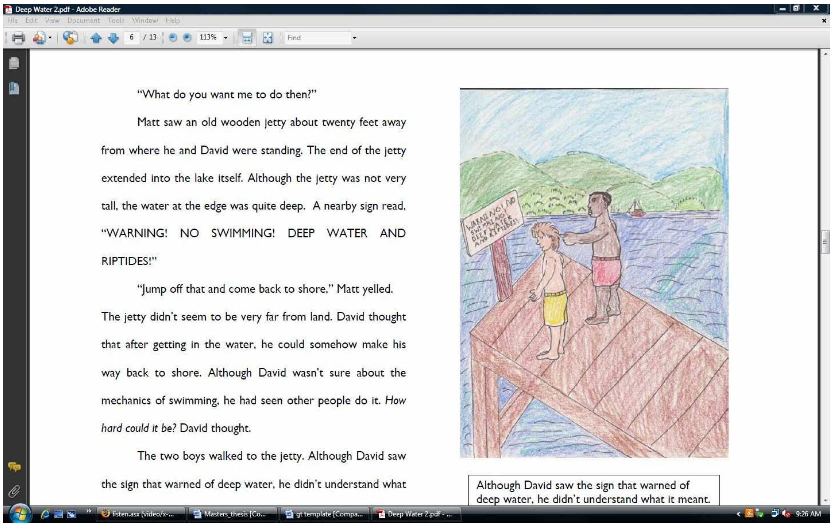
Figure 1: Screenshot of the now defunct WordStar word processing program

Further complicating matters is the fact that different versions of the same word processing program may consist of incompatible formats. Future versions of a particular product may not be able to recognize the formatting of previous versions or edit the contents. Similarly, users with older versions of a particular product may be unable to open documents created with more recent editions. Compatibility may also be a problem if a document created in one version of a word processing program (e.g., Word 2007 for Windows) is unreadable in another version (e.g., Word 2007 for Mac OS/X) (Saffady 2002, 57).

Preserving text files is difficult, because word processing programs are all proprietary formats. Although Microsoft Word is currently the de facto format for text files, it will most likely be replaced by another format at some point in the future. This situation will render all text files written in Word unreadable by the new format. For example, Microsoft could choose to change Word's format, forcing customers to upgrade to a version that is not compatible with pre-existing text files. This recently occurred with the debut of Microsoft's 2007 Office suite, where the file format for word processing documents was changed from doc to docx. Although files written in the previous doc format are backward compatible with Office 2007, the formatting is not always retained when the older format is read by the newer version. The terms of Word's licensing agreement could change, which could impact the degree to which users can access or edit documents produced with the program (Barnes 2006, 5).

Formats for Compound Documents

While current word processing programs are capable of supporting tables, images, and graphs, text file formats are technically only for the rendering of alphanumeric characters. Hence, Adobe's portable document format (PDF) is a better choice for supporting documents that possess complex formatting characteristics. As Figure 2 illustrates, PDF is capable of supporting images, formatted text, headers, and margins. Documents produced using PDF are easy to view, whether on a computer screen or printed out to paper. PDF files can be accessed from a wide variety of sources, including web sites, storage media, and e-mail attachments (Saffady 2002, 58). Although many repositories currently use PDF as a format for long-term digital preservation, Barnes notes that this approach is problematic. Like Microsoft Word, PDF is a proprietary format, meaning that the Adobe Company could choose to charge for the use of the Adobe Acrobat viewer or mandate that the source code of future versions be kept a trade secret (Barnes 2006, 7).



"What do you want me to do then?"

Matt saw an old wooden jetty about twenty feet away from where he and David were standing. The end of the jetty extended into the lake itself. Although the jetty was not very tall, the water at the edge was quite deep. A nearby sign read, "WARNING! NO SWIMMING! DEEP WATER AND RIPTIDES!"

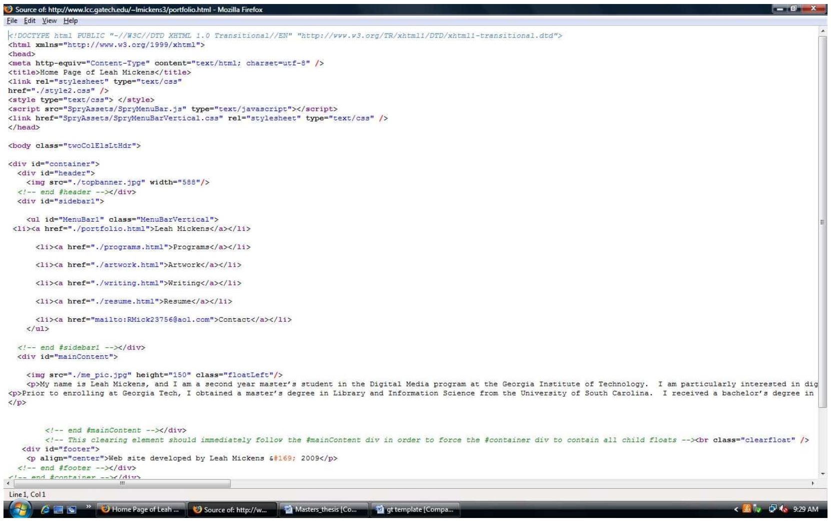
"Jump off that and come back to shore," Matt yelled. The jetty didn't seem to be very far from land. David thought that after getting in the water, he could somehow make his way back to shore. Although David wasn't sure about the mechanics of swimming, he had seen other people do it. *How hard could it be?* David thought.

The two boys walked to the jetty. Although David saw the sign that warned of deep water, he didn't understand what

Although David saw the sign that warned of deep water, he didn't understand what it meant.

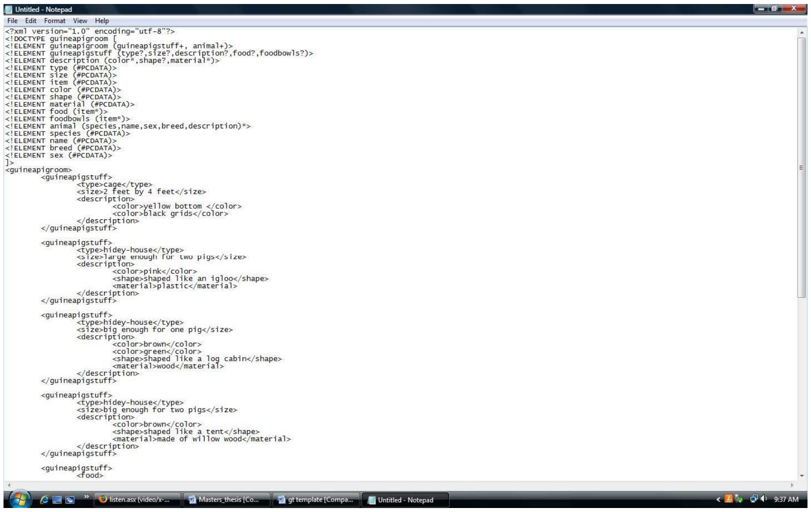
Figure 2: Screenshot of a PDF file created by the author

Markup languages are also capable of supporting compound documents, and unlike word processing programs or PDF, are not proprietary. The term markup refers to the markup or tags that provide the computer with information about how text and graphics should appear onscreen. These languages usually render text for web browsers. The most commonly used ones are hypertext markup language (HTML) and eXtensible markup language (XML), which can be seen in Figures 3 and 4 respectively. One advantage that markup languages possess from a preservation standpoint is that the tags separate content from form, as shown in Figure 4. Since HTML files are stored in plain text form, they can be read on a variety of platforms. However, newer versions of HTML are not viewable on older browsers. The possibilities of XML as a long-term preservation strategy are explored in greater detail in Chapter 6 (Saffady 2002, 58).



```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<title>Home Page of Leah Mickens</title>
<link rel="stylesheet" type="text/css"
href="./style2.css" />
<style type="text/css"> </style>
<script src="SpryAssets/SpryMenuBar.js" type="text/javascript"></script>
<link href="SpryAssets/SpryMenuBarVertical.css" rel="stylesheet" type="text/css" />
</head>

<body class="twoColElsLtHdr">

<div id="container">
  <div id="header">
    <img src="./topbanner.jpg" width="588"/>
  <!-- end #header --></div>
  <div id="sidebar1">

    <ul id="MenuBar1" class="MenuBarVertical">
 <li><a href="./portfolio.html">Leah Mickens</a></li>

      <li><a href="./programs.html">Programs</a></li>

      <li><a href="./artwork.html">Artwork</a></li>

      <li><a href="./writing.html">Writing</a></li>

      <li><a href="./resume.html">Resume</a></li>

      <li><a href="mailto:RMick23756@aol.com">Contact</a></li>
    </ul>

  <!-- end #sidebar1 --></div>
  <div id="mainContent">

    <img src="./me_pic.jpg" height="150" class="floatLeft"/>
    <p>My name is Leah Mickens, and I am a second year master's student in the Digital Media program at the Georgia Institute of Technology.  I am particularly interested in dig
<p>Prior to enrolling at Georgia Tech, I obtained a master's degree in Library and Information Science from the University of South Carolina.  I received a bachelor's degree in
</p>


       <!-- end #mainContent --></div>
       <!-- This clearing element should immediately follow the #mainContent div in order to force the #container div to contain all child floats --><br class="clearfloat" />
   <div id="footer">
    <p align="center">Web site developed by Leah Mickens &#169; 2009</p>
  <!-- end #footer --></div>
<!-- end #container --></div>
```

Figure 3: Screenshot of the HTML markup for the author's home page



```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE guineapigroom [
<!ELEMENT guineapigroom (guineapigstuff+, animal+)>
<!ELEMENT guineapigstuff (type?,size?,description?,food?,foodbowls?)>
<!ELEMENT description (color*,shape?,material*)>
<!ELEMENT type (#PCDATA)>
<!ELEMENT size (#PCDATA)>
<!ELEMENT item (#PCDATA)>
<!ELEMENT color (#PCDATA)>
<!ELEMENT shape (#PCDATA)>
<!ELEMENT material (#PCDATA)>
<!ELEMENT food (item*)>
<!ELEMENT foodbowls (item*)>
<!ELEMENT animal (species,name,sex,breed,description)*>
<!ELEMENT species (#PCDATA)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT breed (#PCDATA)>
<!ELEMENT sex (#PCDATA)>
]>
<guineapigroom>
        <guineapigstuff>
                <type>cage</type>
                <size>2 feet by 4 feet</size>
                <description>
                        <color>yellow bottom </color>
                        <color>black grids</color>
                </description>
        </guineapigstuff>

        <guineapigstuff>
                <type>hidey-house</type>
                <size>large enough for two pigs</size>
                <description>
                        <color>pink</color>
                        <shape>shaped like an igloo</shape>
                        <material>plastic</material>
                </description>
        </guineapigstuff>

        <guineapigstuff>
                <type>hidey-house</type>
                <size>big enough for one pig</size>
                <description>
                        <color>brown</color>
                        <color>green</color>
                        <shape>shaped like a log cabin</shape>
                        <material>wood</material>
                </description>
        </guineapigstuff>

        <guineapigstuff>
                <type>hidey-house</type>
                <size>big enough for two pigs</size>
                <description>
                        <color>brown</color>
                        <shape>shaped like a tent</shape>
                        <material>made of willow wood</material>
                </description>
        </guineapigstuff>

        <guineapigstuff>
                <food>
```

Figure 4: Screenshot of XML markup

Digital Images

With the rise of digital photography and digital manipulation software during the late 1990s and early 2000s, archivists and librarians are finding that digital imagery is becoming a fast-growing segment of their holdings. The term "digital imagery" can refer to a wide variety of documents, including photographs, computer-based artwork, slides, geographical information systems, graphs, tables, computer-aided designs, and scanned versions of paper-based images. Although some digital images can be migrated to paper (see Chapter 6), others can only be viewed in a digital environment using specific applications (Saffady 2002, 63).

Raster Graphics

Two methods of capturing digital images are raster graphics and vector graphics. Raster graphics, also known as bitmap images, are comprised of invisible grids that contain small units called pixels (Brown 2003, 4). The five most common color settings for raster images are 1-bit monochrome, 4-bit grayscale, 8-bit gray scale, 16-bit high color, and 24-bit true color. The pixels in monochrome images are either black or white. Grayscale images use pixels that express varying degrees of gray between pure white and pure black. The number of shades of gray varies depending on the color settings. For example, a 4-bit grayscale image contains 16 shades of gray, whereas an 8-bit grayscale image has 256 shades of gray (Brown 2003, 5). Unlike grayscale images, the pixels in color images are comprised of three numbers that represent the hue (i.e., the color), saturation (color intensity), and grayscale (how dark or light the color is) (Hunter 2000, 19). The true color setting displays 16,777,216 colors, which is roughly equivalent to the number of distinctive colors discernible by the human eye (Brown 2003, 5).

The colors for raster images derive from the mixture of colors in what is known as the color space. A color space is the system that determines which values correspond to which colors. The most common color space for digital images is RGB, which defines colors in terms of combinations of red, green, and blue components. Each component consists of one byte that corresponds to a value between 0 and 255. For example, the color yellow can be expressed in the RGB color space as 246 (Red), 232 (Green), and 43 (Blue). This is possible because RGB is based on the principles of mixing light, rather than pigment (Brown 2003, 5). When creating digital images for print mediums, the CMYK (Cyan, Magenta, Yellow, Black) color space is used. Another common color space for "digitally born" images is HSV (Hue, Saturation, Value). Innumerable digital image manipulation applications allow users to switch between these color spaces (Brown 2003, 6).

Many formats exist for raster files. The format that is most used for long-term record-keeping is the tagged file format (TIFF). TIFF was the result of the joint efforts of the Microsoft Corporation and the Aldus Corporation, which is now owned by Adobe. TIFF is a cross-platform and is widely supported by a number of software applications. All TIFF images contain a header that includes the title, contents, size and other relevant metadata (Saffady 2002, 63). TIFF is a "file wrapper," meaning that it encompasses all of the information needed to support a variety of image elements, such as vector images, raster images, monochrome, grayscale, RGB, and CMYK. Unlike other file formats, TIFF, "holds original data in the original order and format…it can contain a perfect version of the capture, with full resolution and color, pixel-by-pixel." Despite these advantages, some TIFF images, depending on the software from which they originate, may not be readable by some applications. This occurs because some image manipulation programs add proprietary information to TIFF files that cannot be read by other applications (Vitale 2007, 33).

The graphics image file (GIF) format is widely used for images that are viewed in an online environment. GIF was developed by CompuServe and was designed to allow images to be quickly downloaded by web browsers. This is accomplished by initially displaying a low resolution copy of the image that increases in quality as the web browser receives more information about the image from the server (Saffady 2002, 63). The image depth of GIF images ranges from 1-bit monochrome to 8-bit color (236 colors). Images in the GIF format are always compressed to ensure that the file size remains relatively small. The compression technology that is employed is known as LZW (Lempel-Ziv-Welch, the surnames of the developers). The June 1984 issue of *Computer* magazine included an article that first described LZW. Although the Unisys Corporation held the patent on LZW, this detail was not mentioned in the article. This omission led many developers to conclude that LZW and the GIF format were free to use. Complicating matters was the fact that CompuServe did not know that it had used a proprietary method of compression when it was developing the GIF format. Once GIF became a web standard in the late 1980s, Unisys demanded a $0.10 royalty payment per image for the use of LZW. In response, many web developers boycotted the GIF format in protest (Battilana 2004). The question surrounding royalties from the use of the GIF format ended in June 2003, when the United States patent for LZW expired (Brown, 2003, 8).

The Portable Network Graphics format was developed in 1996 by the PNG Development Group to create a royalty-free alternative to the GIF format. PNG provides greater color depth than GIF (48-bit color versus 8-bit color). Like GIF images, PNG images are always compressed. PNG uses the Deflate lossless compression logarithm, which is patent-free (Brown 2003, 9). Both GIF and PNG are single image formats, meaning that each file can only contain one file. While this attribute is not important when these formats are used on web pages, it can be problematic when scanning digital images from documents with many pages (Saffady 64, 2002).

In 1990, the Joint Photographic Expert Group created a file interface form (JFIF) that can be used with the JPEG image compression algorithm. Hence, when the JPEG standard is referenced, what is actually being discussed is the JFIF standard (Brown 8, 2003). JPEG uses lossy compression, which is defined as “an irreversible way of reducing the size of data by approximating it from the original bitmap image.” This is problematic from a preservation perspective; once information from a raster image is lost, it cannot be retrieved. Consequently, images saved in this format are not appropriate for long-term digital preservation (Vitale 2007, 35).

Windows Bitmap (BMP) is a format developed by the Microsoft Corporation to be the default format for raster images produced on the Windows operating system. Consequently, different versions BMP are released with each new edition of Windows. The color depth of BMP ranges from 1-bit monochrome to 32-bit color. BMP images originating in Windows 3.0 and higher offer optional lossless compression. Although BMP is free to use and well documented, it has declined in popularity with the introduction of more sophisticated graphics formats (Brown 11, 2003).

Vector Graphics

Vector-based images represent images as a series of interlocking geometric shapes, such as points, lines, circles, and polygons. While raster images are best for what the Technical Advisory Service for Images calls “continuous tone photo-realistic images,” vector images are best for rendering straight lines, curves, and geometric shapes (Technical Advisory Service for Images 2005, 1). Vector images are also widely used in word processing programs to generate fonts, drawing tools, tables, and graphs (Technical Advisory Service for Images 2005, 2). Depending on the application and format, vector images can be used to create 3-D images. The most common usage of vector images is for computer-aided designs (CAD), 3-D animation, maps, and graphs. Unlike raster images, vector images can be scaled without suffering from a subsequent loss of picture quality (Brown 2003, 7). Since vector images are generated using mathematically-based computer code, a change in the size of the image affects the variables in the shape's algorithm, not the actual image, as is the case with raster images (Technical Advisory Service for Images 2005, 2). Although vector images cannot be compressed, the size of such a file is proportional to the complexity of the image in question. Generally speaking, vector image files are much smaller than an equivalent raster image (Brown 2003, 7).

Unlike raster images, vector image formats are often proprietary. However, many CAD programs are capable of supporting several other formats (Saffady 2002, 64). The most common vector image format is the drawing format (DWG) developed by Autodesk. This is the native format for images created with AutoCAD, the most popular CAD program. Like Microsoft's BMP format, the DWG format changes with each subsequent version of AutoCAD. DWG can accommodate 24-bit color and 3-D graphic modeling. Due to the popularity of DWG, many other CAD packages support the format. However, the appearance of the image may vary from program to program (Brown 2003, 11).

Autodesk developed the Drawing Exchange Format (DXF) to provide users with a means of facilitating data exchange with various CAD programs. Not only is DXF compatible with many other CAD applications, but all DXF files can be converted into

DWG files and vice versa (Saffady 2002 64). DXF files can be represented using 7-bit ASCII characters or binary. DXF images have a color depth of 8-bit color and can support 3-D modeling. Although DXF files can be read by applications other than AutoCAD, the complexity of this format means that the appearance of DXF images that are read in other programs may vary considerably. Since the format specifications change with each version of AutoCAD, a DXF file written in an earlier version may not be readable on a newer application. Finally, Brown warns that "some applications may read a DXF file whist skipping unsupported features. This can lead to the loss of information in ways that may not be obvious to the user" (2003, 12).

Scalable Vector Graphics (SVG) is a non-proprietary XML-based format for 2-D images. Although SVG was designed for vector graphics, it can also be used for raster images in some instances. SVG is an extremely flexible format; not only does it have a color depth of 24-bit color, it can also support the creation of dynamic and interactive graphics. Because of its XML base, SVG graphics can be manipulated using XML editing tools or a proprietary CAD program. Consequently, SVG is emerging as a widely used standard for web-based vector graphics (Brown 2003, 13).

Digital Audio

In the non-digital world, sound is created when air vibrates at a given frequency (i.e., the number of vibrations per second). These frequencies travel on waves that, upon entering the ear canal, are converted into meaning by the brain. The waves of sound captured using analog equipment are rendered with voltage variations (Wilson et al. 2006, 48). In comparison, the waveforms of digital sound are represented as "a series of samples taken at specific time intervals, whose values are given as binary numbers. The sample rate is the number of times per second that the analogue [sic] signal is measured." Before a computer can manipulate analog sound, it must be converted into a digital format in a process known as sampling. During this procedure, an Analog to Digital Converter (ADC) chip calculates the voltage of each sampling interval and assigns it a numeric value based on the level of sound present on each wave. This value is translated into a digital form that is readable by the computer. To play the sound, a Digital to Analog Converter (DAC) chip takes the binary numbers of the digitized information and converts them into an output voltage. The voltages are then transmitted as analog sound waves through a speaker (Wilson et al. 2006, 49).

The quality of digital sound is judged on the following five attributes: bit depth, sampling rate, compression method (codec), the number of channels, and the format. Bit depth, also known as resolution, is the quantity of bits used to record each sample. As is the case with images, the bit depth of digital sound ranges from one to 24-bits. A higher bit depth directly corresponds to an increased level of audio quality, a wider volume range, and a larger file size (Wilson et al. 2006, 50). The sampling rate is the rate at which the amplitude of the sound wave is calculated. This amount is measured in kilohertz (kHz), or thousands of samples per second (Colorado Digitization Program Digital Audio Working Group 2006, 7). A codec is the technique, usually a software application, which compresses and decompresses an audio file. Codices are often used to deliver audio on the Internet to decrease download times. Unless the right codec is available, the audio sample in question will not play (Wilson et al. 2006, 50). Many formats are capable of recording digital audio. The best formats for long-term preservation are those that are non-proprietary, uncompressed, cross-compatibility, and at a low risk of obsolescence. Some of the most widely used formats are described in the next section (Colorado Digitization Program Digital Audio Working Group 2006, 14).

The waveform audio file (WAV) format is a proprietary format that was developed by the Microsoft Corporation and IBM. It is the standard for audio files that use the Windows operating system. Although WAV is a very popular format for sharing audio files on the Internet, it has a maximum capacity of 4 GB (Wilson et al. 2006, 53). Because of the ubiquity of the WAV format, all PCs and web browsers come with WAV players pre-installed. As WAV files are uncompressed, they must be downloaded before the contents can be played (Saffady 2002, 69). There is also a professional version of WAV called broadcast WAV (BWF) that contains metadata in the file header. Consequently, the BWF format is becoming the format of choice for the archiving of WAV files, even though the metadata header is not readable by all media players (Colorado Digitization Program Digital Audio Working Group 2006, 20-21).

Due to the growing ubiquity of portable digital music players, the Motion Picture Experts Group (MPEG) Audio Layer III (MP3) format is the most popular application for Internet audio upload and download. Because MP3 files are often accessed from an online environment, files stored in this format are highly compressed to keep the file size relatively small, while enabling a high degree of audio quality. However, because of changes that compression creates in files, MP3 is unsuitable as a long-term preservation format (Colorado Digitization Program Digital Audio Working Group 2006, 21).

MPEG also supports three other audio formats. Advanced Audio Coding (AAC), also known as MP4A, is for encoding CD-quality digital audio files. Because the audio quality of AAC/M4A files is higher than that of comparable MP3 files, MPEG anticipates that the former will one day replace the latter. The AAC/MP4A format is used for digital music sold on Apple, Inc.'s iTunes store (Wilson et al. 2006, 51). The MP1 and MP2 formats are for digital audio files that do not require the same level of quality found in MP3 and AAC/MP4A (Saffady 2002, 69).

The Real Audio (RA) format refers to a group of proprietary formats developed by RealNetworks (Wilson et al. 2006, 52). It is primarily used for streaming audio, which refers to audio that is broadcast from a Web browser in real time. However, RA files can also be downloaded and opened at a later date. Because Internet servers are subject to bandwidth restrictions, the audio quality of streaming RA is acceptable, but not of archival quality (Saffady 2002, 69).

The Audio Interchange File Format (AIFF) was developed by Apple, Inc. for its Mackintosh computers and is equivalent to Microsoft's WAV format. Although AIFF is mainly used for interchange, it can also be used for storage. Because AIFF is an uncompressed format, high quality audio files are large (Wilson et al. 2006, 51).

Conclusion

Because file formats become obsolete in the same manner as computer hardware and software, it is important for archivists to be knowledgeable about the properties of various types of digital information. Archivists should not assume that all digital files possess the same archival properties; the preservation needs of an ASCII text file are markedly different than those required of a Flash web site, for example. Ignorance of how to properly care for different types of digital information lies at the root of multifold inaccessible files. Consequently, many institutions do not maintain their digital assets in a way that ensures the continued existence of these files, because archival theory has not kept pace with technological advancement. The archival community must understand that digital files do not have the same characteristics as physically instantiated artifacts and should not be treated as such, even when said files are contained on a tangible storage device (Besser 2000, 156).

CHAPTER 5

STORAGE MEDIA

Storage media obsolescence poses a major threat to the long-term accessibility of digital information. Digital information must be stored somewhere, be it on an internal hard drive or an external storage device, such as a flash drive (Bergeron 2002, 40). The computer memory technology industry emphasizes the development of products that are smaller, faster, and possess a large storage capacity, not devices that are notable for their longevity, either in the market or in daily use (Bergeron 2002, 11). Unlike analog information formats like the book, whose configuration has remained constant for almost 2,000 years, improvements in computer memory technology have been occurring so quickly within the past 30 years that many archives and museums contain an overabundance of obsolete storage devices with no obvious way of accessing the information contained within them. As with digital file formats, archivists need to be knowledgeable about storage devices so they can accurately assess the preservation needs of the digital files in their care (Saffady 2002, 19). This is particularly true in the context of multimedia projects like the Legacy Project, in which the backup data sources for the project are stored on dozens or possibly hundreds of storage devices.

Storage media for digital information can be grouped into three categories: magnetic media, optical media, and flash memory. *Magnetic storage devices* "use the principles of electromagnetism to record and change electrical signals," whereas *optical media* uses "concentrated light in the form of lasers to alter reflectance on the surface of a disk" (Hunter 2000, 21). *Flash memory devices* use advanced semiconductor technology to store large amounts of data in small apparatuses that range from the size of a postage stamp to a pack of gum (Brown 2008, 5). All storage devices, regardless of type, keep the information contained on them readable by maintaining the proper sequence of zeros and ones. Data loss occurs when the bit signals are unable to sustain the order necessary to keep the information readable by the computer (Hunter 2000, 21).

This chapter provides an in-depth discussion of the archival properties and preservation needs of storage media for digital information. The first part focuses on magnetic media. The second part discusses optical disks. The third part describes flash storage devices. Each part begins with a general overview of the media in question. Descriptions are provided of the different formats for each storage medium. Each portion concludes with a description of the archival problems associated with the specific type of storage device. The conclusion of this chapter describes the general preservation problems common to all storage devices used in the computing environment.

Overview of Magnetic Storage Devices

Magnetic storage devices have the longest history of the storage media mentioned and are the most frequently used. Saffady defines a magnet as “a piece of metal that is capable of creating a magnetic field that will attract or repel other metals.” Although all matter is affected by magnetism to some degree, most materials react too weakly for it to be harnessed for any functional purposes. However, the magnetic recording devices utilized by computers are characterized by the presence of powerful magnetic properties, even without the presence of a robust external magnetic field. Magnetic storage devices record information when the external magnetic field, known as the read/write head, rearranges the atoms in the domains (microscopic data pits) in such a way as to represent zeros and ones (Saffady 2002, 20).

Magnetic media can be classified as either tapes or disks. Disks are primarily used for computer applications, whereas *tapes* are used for data storage and audio and video recording (Saffady 2002, 21). Regardless of type, all magnetic storage devices consist of three key components: the recording surface, substrate, and binder. The *recording surface* is a top layer that contains a material that becomes magnetized when exposed to an external magnetic field, known as a read/write head. Magnetic storage devices are able to retain information because the recording surface maintains its magnetic properties (and thus the information in question) even when the field is removed (Hunter 2000, 22-23). *Substrate* refers to the supporting material on which the recording surface sits. The presence of this additional backing is necessary, because the recording surface is too thin to be used as a standalone layer (Van Bogart 1995, 3). The substrate layer is also used to transmit the information. Substrate can be made of a variety of materials, including aluminum, polyester, ceramic, or glass. (Hunter 2000, 22). The *binder* serves several purposes. Its primary function is to attach the recording surface to the substrate. The binder also acts as a lubricant between the recording head and the recording surface. Figure 5 illustrates a cross section of the different layers present in a strip of magnetic tape (Van Bogart 1995, 2-3).

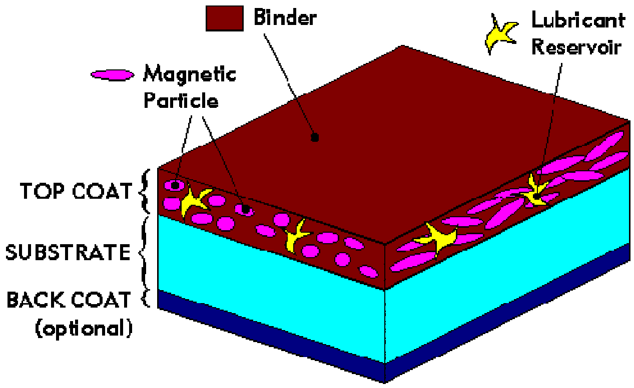


Figure 5: Cross section of magnetic tape (from Van Bogart 1995)

Magnetic Disks

Magnetic disks contain a flat, circular, substrate layer that is covered with a magnetic recording substance. The surface of the substrate is separated into concentric rings that are called tracks (Saffady 2002, 21). Within the tracks are sectors, which contain the information in question (Hunter 2000, 23). The magnetic read/write head reads the tracks to either record or playback the information that the user wants. The bits that are recorded on the disk are horizontally on each track. The substrate on magnetic disks can either be rigid or elastic. Appropriately, the former is referred to as a hard disk and the latter is called a floppy disk (Saffady 2002, 21).

The most common type of hard disk is the rigid fixed magnetic disk drive, which is popularly known as a hard drive. Unlike floppy disk drives, the storage device and the recording medium on a hard drive are considered a single unit (Saffady 2002, 21). The bits for accessing a single file are not stored together, but are often spread throughout the disk (Hunter 2000, 23). To reassemble the data to display a single, coherent file, the outermost part of the disk contains a list of addresses. A controller consults this subdirectory and reassembles the correct order before sending it on to the central processing unit (CPU) (Hunter 2000, 23-24). Hence, when a file is erased from a hard drive, it will remain on the disk until new data replaces the sectors used by the old file. Information that has been accidently deleted from a hard drive can be retrieved if the user looks for the file in the sectors rather than the disk directory. An example of a laptop hard drive can be seen in Figure 6 (Hunter 2000, 24).

Figure 6: Example of a hard drive for a laptop computer

Floppy disks are defined as "circular pieces of polyester coated with a magnetizable material." The recording material can be made of a variety of substances, including gamma ferric oxide, cobalt-modified iron oxides, and barium ferrite (Saffady 2002, 25). Introduced in 1971, they were the most popular type of storage device until they were rendered obsolete by the USB flash drive (Hunter 2000 24). Although floppy discs were originally designed to be low-cost alternative for hard drives in early microcomputers, they quickly became the principle external storage device for first generation home computers (Saffady 2002, 25).

Figure 7: 8-inch floppy disk drive next to a 3.5-inch disk

The first floppy disks had a diameter of 8 inches and had storage capacities of 80 kilobytes. Floppy disks of this size were occasionally referred to as being of the standard size, to differentiate them from the 5.25 and 3.5-inch disks that were subsequently developed. However, the moniker "standard" was misleading; not only were 8-inch disks formatted to be used by only one operating system, but many disks could only be used by a single application. Eight-inch floppy disks were used until the early 1980s when they were superseded by 5.25-inch floppy disks, also known as minifloppy disks (Saffady 2002, 26). Figure 7 depicts an 8-inch floppy disk and drive next to a 3.5-inch floppy.

Figure 8: 5.25-inch floppy disk

5.25-inch floppy disks were originally created for low priced microcomputers in educational or home settings, while the 8-inch floppy was preferred by the business market. The format finally gained greater acceptance when IBM installed 5.25-inch disk drives in its Personal Computers (PC) in the early 1980s. The first 5.25-inch disks were single-sided with maximum storage capacities that ranged from 70 to 120 kilobytes. Later, a double-sided, double density disk was developed that had a storage capacity of 360 kilobytes (Saffady 2002, 26). Figure 8 shows a 5.25-inch floppy disk.

By the early 1990s, 5.25-inch floppy disks were rendered obsolete by the 3.5 floppy disk. The first 5.25 inch disks were formatted for use either with the MS-DOS or Macintosh operating systems. Disks for the former could hold 720 kilobytes, while those formatted for the latter were capable of storing 800 kilobytes. Later, the standard format became a double-sided, high-density 1.44 megabyte disk (Saffady 2002, 26).

Although there were sporadic attempts to introduce smaller floppy disks with higher storage capacities (10 to 20 megabytes) during the late 1980s and early 1990s, none were commercially successful. These devices, known as floptical disks, worked by using "optically encoded control information to greatly increase track density" although the recording layer was magnetic, not optically based (Saffafy 2002, 26). The most successful high-density floppy disk was the Zip disk, which was developed in 1994 by the Iomega Corporation. Zip disks were also 3.5 inches in length, but were noticeably thicker and were not compatible with the drives of standard 3.5-inch floppies. By the time Zip disks became obsolete in the early 2000s, they had a maximum storage capacity of 750 megabytes. Unlike the floppy disk types mentioned previously, the Zip disk was a proprietary format (Saffafy 2002, 27).

Magnetic Tape

Magnetic tape is defined as “a long strip of polyester film coated with a magnetized recording material (Hunter 2000, 25).” Until the introduction of the floppy disk, magnetic tape was the primary external storage device for digital information. However, as Saffafy says, “[magnetic tape’s] serial access characteristics rendered them unsuitable for on-line applications requiring rapid retrieval of information in unpredictable sequences.” Although floppy disks eventually replaced magnetic tape as the primary storage device for most home computer applications, the former is still used extensively in electronic recordkeeping, especially as data backup. Magnetic tape is also widely used for mainframe and minicomputers (Saffady 2002, 28).

Of all of the storage devices discussed in this paper, magnetic tape has the most formats. The most commonly used type for mainframe computers is the nine-track tape drive, shown in Figure 9. This consists of a 10.5-inch plastic reel that contains 2,400 feet of magnetic tape. The data bits are encoded horizontally across the tape in nine parallel tracks. Eight of the tracks contain data, while one is used for error checking. Tape density is determined by the number of bits per inch (bpi) along the tape’s length. Although the storage capacity of nine-track tape reels are limited when compared to newer forms of storage media (i.e., solid state media), its position as a mature medium means that there are no concerns about backward compatibility resulting from continued product development (Saffady 2002, 28).

Figure 9: Nine-track tape drive

Another magnetic tape format utilized by mainframe computers is half-inch data cartridges. They were originally developed by IBM in 1984 and were later adopted as a standard by other computer manufacturers. The original format is the 3480 type, which consists of a 4 inch by 5 inch by 1 inch cartridge with a chromium-dioxide recording surface. The tape is 0.5 inches wide and contains two parallel sets of nine tracks. According to Saffady, tapes in the 3480 series contain “550 feet of magnetic tape. The normal cartridge capacity is 200 megabytes, which is equivalent to five reels of nine-track magnetic tape recorded at the 1,600 bpi or 1.25 reels recorded at 6,250 bpi.” More recent types of half-inch cartridges include the 3490 (1989) and the 3490E (1991). The 3490 is identical to the 3480 format, except that it allows data compression which triples the tape’s storage capacities. When uncompressed, the 3490E stores up to 800 megabytes and when compressed it can store 2.4 gigabytes (Saffady 2002, 29). Since the 1980s, these three formats, often referred to as the 34XX formats, have been steadily replacing nine-track tapes as the standard type of tape used for mainframe and minicomputer data storage (Saffady 2002, 30).

Digital linear tape (DLT) is the preferred format for network server installations and midrange computers. This format was first developed by the Digital Equipment Corporation as an alternative to 34XX formats. DLT has the same physical dimensions as half-inch cartridges and the recording material consists of a high density metal particle material. Older DLTs had a maximum storage capacity of 40 gigabytes, though modern versions can store up to 330 gigabytes when uncompressed (Saffady 2002, 31).

Linear tape-open (LTO) technology was developed by IBM, Hewlett-Packard, and Seagate Technology in the early 1990s as a way of providing data backup for the increasing number of large network servers (Saffady 2002, 31). Two formats exist for LTO technology: Ultrium and Accelis. The former is notable for its high storage capacity, whereas the latter has the ability to access data quickly. The Ultrium tape consists of a single reel with half-inch tape that can hold up to 800 gigabytes of uncompressed data. The Accelis tape contains a double reel with eight-millimeter tape and a storage capacity of 25 gigabytes (Saffady 2002, 32).

Quarter-inch cartridges, as seen in Figure 10, were developed in the 1970s as smaller, low cost alternatives to nine-track tapes for minicomputer applications. Since then, these tapes are primarily used as data backups for small to mid-sized computers. While the earliest quarter-inch cartridges were proprietary, modern versions are manufactured using open standards created by Quarter-Inch Cartridge Drive Standards Incorporated that are known as QIC formats. The tapes and drives that are compliant with these standards are called QIC products (Saffady 2002, 32). Despite the presence of these standards, there are many formats, some of which are now obsolete, for quarter-inch cartridges. Hence, some cartridges may not be compatible with all tape drives (Saffady

2002, 34). Quarter-inch cartridges come in two different sizes. Cartridges of the first size have dimensions of 4 inches by 6 inches by 0.625 inches and are used in 5.25-inch tape drives. Those of the second type are 2 inches by 3 inches by 0.5 inches and are for 3.5-inch tape drives. The first quarter-inch cartridges could contain less than 200 megabytes of data, though modern versions can have the same storage capacities as a hard drive (Saffady 2002, 33).

Figure 10: Quarter-inch tape

According to Saffady, eight-millimeter cartridges are “cassettes that contain a metal particle or evaporated metal tape specifically designed for high-density data recording.” These cartridges have dimensions of 3.7 inches by 2.5 inches by .6 inches and can be an alternative to quarter-inch tapes and digital tape. They are mainly used for data backup and archiving in midrange computer operations. Eight-millimeter data cartridges use helical scan technology, rather than the longitudinal method that is employed by half and quarter-inch tapes. Saffady compares and contrasts the two recording techniques, stating:

> [Longitudinal recording uses] stationary magnetic heads [that] record data in parallel tracks that run the entire length of a tape. Helical scan recording technologies, by contrast, record computer processible information in narrow tracks positioned at an acute angle with respect to the edges of a tape. As their principle advantage for electronic recordkeeping, helical scan technologies offer higher densities than are possible with longitudinal tape recording.

Eight-millimeter data cartridges are manufactured by Sony and the Exabyte Company, but the products that these companies make are not compatible with each other (Saffady 35, 2002). Exabyte cartridges can hold up to 60 gigabytes, while Sony's have a maximum capacity of about 800 gigabytes (Saffady 2002, 36).

Digital audio tape (DAT) was originally developed for audio recording and was eventually adopted for computer data storage in 1988. DAT consists of "tape [that] is four millimeters wide and features a high-coercivity metal recording material," where coercivity is defined as "the amount of force, usually measured in oersteds, that is required to orient magnetic particles." The Digital Data Storage (DDS) Manufacturers are responsible for establishing standards for DAT. DAT can store up to 50 gigabytes of data and the newer tape drives are backward compatible with older versions. This is the format of choice for home and small business data backup and archiving (Saffady 2002, 36). An example of DAT can be seen in Figure 11.

Figure 11: Digital Audio Tape cassette

The Preservation Problems of Magnetic Media

One of the most common problems that afflict magnetic tape is binder degradation. This occurs when the binder becomes soft, brittle, or loses its lubricant and/or shape (Van Bogart 1995, 4). Environments that have high humidity rates can generate binder hydrolysis when water from the air causes the binder molecules to become shorter. These smaller molecules do not provide the same degree of binder integrity. The end result of binder hydrolysis is soft binder coating, a higher rate of friction between the tape and the magnetic head, and gummy tape surface residues. Van Bogart states:

> A sticky tape can exhibit sticky shed, produce head clogs, result in stick slip playback, and in extreme cases, seize and stop in the tape transport. Tape binder debris resulting from binder deterioration will result in head clogs that will produce dropouts on a VHS tape when played back. The sticky tape syndrome will result in the squealing of audio-tapes as the tape very rapidly sticks to and releases from the playback head.

Although it is possible to temporarily reverse the effects of binder hydrolysis by baking distressed tapes at 122 degrees Fahrenheit for three days, this technique only works on reel-to-reel tapes. For other types of magnetic tape, baking may aggravate the situation (Bogart 1995, 5).

Another common problem with magnetic tape is lubricant erosion. Lubricant is added to the binder of magnetic tape to reduce the level of friction between the recording head and the tape. Because a small amount is transferred to the recording head and pins each time it is played, the lubricant on the tape eventually begins to dissipate. This dissipation occurs even when the tape is not being used, due to the natural process of evaporation. Although tapes can be re-lubricated, this process requires the work of a trained professional, as over-lubricated tapes lose their information (Van Bogart 1995, 6).

Finally, magnetic tapes regularly shed their magnetic particles, which lead to a gradual loss of information. The degree to which magnetic particles are lost varies, depending on the type of tape in question. For example, tapes containing iron oxide and cobalt-modified iron oxide shed less frequently than those made of metal particulate and chromium dioxide, although the former two are used for medium to low grade tapes and the latter two are used in high grade tapes. Regardless of the material used for the particles, all tapes experience magnetic deterioration; it is simply a physical property of magnetic media. Archivists can decelerate deterioration by storing tapes in an environment that is temperature controlled (Van Bogart 1995, 7).

Optical Storage Devices

Optical storage devices use laser induced light to record information on the surface of a platter-shaped recording medium. The laser records data by creating variations in the way that light reflects off the recording surface. These variations represent ones and zeros, respectively (Hunter 2000, 27). Two lasers are used, one for recording and one for playing the information. The latter operates at a lower power level or on a different wavelength than the former to prevent the accidental overwriting of information. Optical media is characterized by a high storage capacity, because the data bits are recorded very closely together. There are three categories of optical media: magneto-optical disks, compact disks, and DVDs (Saffady 2002, 43).

Although all magnetic media are capable of being recorded, optical media is available in read-only or recordable formats. Read-only disks contain pre-recorded information that can neither be erased nor replaced by new content. Recordable disks are blank and are capable of being recorded. Optical disks of this type are further classified according to the number of times in which information can be written on them. Write-once Read Many (WORM) disks refer to those disks for which the contents are not erasable once the information is recorded. Rewriteable disks are those whose contents can be erased and re-recorded many times. Optical disks can be classified into three broad categories: magneto-optical disks, compact disks, and DVDs (Saffady 2002, 44).

Magneto-Optical Discs

Magneto-Optical disks, such as the kind depicted in Figure 12, employ a hybrid technology in which information is stored magnetically, but recorded and read using a laser. The substrate of a magneto-optical disk is covered with a thin layer of iron and other metals. The recording layer is then covered with a protective layer that prevents foreign matter from entering the disk. All of the magnetic particles on a blank disk have the same magnetic charge and are inert when stored at room temperature. During the recording process, a laser heats up the disk, while an electromagnet moves the particles into the spaces where they will represent either a one or a zero. The manner in which different parts of the disk reflect light represent the ones and zeros of the digitally encoded data. Optical disks with a 5.25 diameter are the most common, although 3.5-inch and 2.5 disks are also available (Saffady 44, 2002).

Figure 12: Magneto-optical disk

Compact Discs

Compact disc (CD) is a general term that refers to disks and media readers that use the optical technologies that were jointly developed by Sony and Phillips in the late 1970s and early 1980s. Most CDs have diameters of 4.75 inches and are further classified by the type of information that they contain (Saffady 46). Commercially produced CDs are in the compact disc-digital audio (CD-DA) format. Compact disc-read-only-memory (CD-ROM) consists of a CD that has information that is computer readable, but not recordable. CD-ROMs are widely used as a storage medium for software, digital art, and multimedia works (Saffady 2002, 46).

Compact disc-recordable (CD-R) and Compact disc-recordable-rewrite (CD-RW) are blank CDs that can be used to record information. As their names suggest, the former can be recorded once, while the latter can be recorded multiple times (Saffady 2002, 46). CD-Rs are recorded using dye-based technology. This involves using a layer of light-absorbent organic dye covered with transparent polymer as a recording material. The recording laser diffuses the dye into the absorption layer, which creates changes in the way light is reflected (Hunter 2000, 29). The reflective layer is made of silver, silver alloy, or gold to prevent internal corrosion (O'Kelly 24). Figure 13 depicts a CD-R disc.

Figure 13: CD-R

CD-RWs contain layers of a semi-metal alloy vacuum between two insulating layers, rather than a dye-based layer. The heat from the recording laser causes the alloys to switch between a highly reflective crystalline state and a less reflective amorphous state. These different states represent ones and zeros. To rewrite the disc, the entire alloy layer is reheated to its crystalline state. This method of recording is known as phase changing recording. It is the same technology used for DVD-R. CD-RWs can be rewritten approximately 1,000 times (O'Kelly 32).

DVDs

Digital Versatile Disks (DVD) were developed to store large quantities of high quality data in a device the size of a CD. DVDs were preceded by the laserdisc, which was developed in the 1970s, as an alternative to VHS and Betamax tapes. However, laserdiscs were too bulky for the average consumer and could only contain analog information (O'Kelly 34). To create a device capable of storing digital video, the data pit sizes were decreased and the number of data tracks (the concentric rings that contain the data pits) was increased. The former was accomplished by decreasing the wavelength of the reading laser vis a vis the one used for CDs (650 nanometers versus 780 nanometers) and molding the disc from two separate halves. This process enabled researchers to:

> make each half of the DVD hold an inner semi-transparent layer to allow the laser to focus either on the reflective surface or the semi-reflective surface in the middle. This allowed [manufacturers] to produce four different types: 1 side, 1 layer; 1 side, 2 layers; 2 sides, 1 layer on each; and 2 sides, 2 layers each (O'Kelly 36).

The video format for DVDs is MPEG-2 and is of a much higher picture quality than that of VHS cassettes. The storage capacity of DVDs varies depending on the number of layers and sides that are employed. For example, DVD 5, which is single layered and single-sided, can contain 4.8 gigabytes, whereas DVD 18, which is double-sided and double layered, has a maximum storage capacity of 17.0 gigabytes (O'Kelly 37).

Like CDs, DVDs are available in read-only and rewritable versions. DVD-Rs use the same dye-based recording technology as CD-Rs. There are two versions of DVD-Rs, one for home use and the other for copyrighted materials. The former uses a recording laser that is of the same wavelength as the one used for home DVD players (650 nanometers). The latter is used to make master copies on advanced audio-visual equipment. Its recording laser is 635 nanometer, which makes it incompatible with many general use DVD drives (O'Kelly 40). DVD-RW employs phase change recording technology like CD-RW. They have a lifecycle of about 1,000 rewrites. Because DVD-RWs are not as reflective as DVD-Rs, some DVD players may have difficulty reading them or completely reject them (O'Kelly 41).

DVD-RAM (Random Access Memory) discs are designed primarily for data backup and archiving, although they can also be used to record videos. In addition to data pits, there are also sector pits, each with an assigned address that enables information to be located quickly. To the naked eye, these pits appear to be small perpendicular lines, as seen in Figure 14. The disc itself spins at a constant speed, which further reduces searching time, as the reading laser does not have to change speeds for different operations. DVD-RAMs are contained in protective cartridges (O'Kelly 42).

Figure 14: Close-up of a DVD-RAM

DVD technology advancements have made it possible to record high definition video on optical discs. Blu-ray technology, which was developed in 2002, is the standard for high definition DVDs. These discs, such as the kind seen in Figure 15, get their name from the violet-colored laser used to record information. The recording laser has a wavelength of 405 nanometers, which is much shorter than the ruby red ray used to read and record CDs (780 nanometers) and conventional DVDs (650 nanometers). The shorter wavelength enables a greater number of smaller data pits to be placed on the disc than those found on a standard DVD (O'Kelly 60). As with other types of optical discs, it is possible to purchase recordable Blu-ray discs: BD-R, BD-R LTH, and BD-RE (rewritable).

Figure 15: Single-layer Blu-ray disc

Figure 16: Xbox 360 with a built-in HD DVD drive

Discs using the now defunct HD DVD standard were constructed to be more similar to conventional DVDs, so no special hardware needed to be purchased. A single layered HD DVD could hold 15 gigabytes, while a double-sided HD DVD could contain 30 gigabytes. An advanced compression logarithm shrank high definition digital video files down to a size that easily fits on a single disc (O'Kelly 61). HD DVD lost the high definition DVD "format war" to Blu-ray in 2008, though many repositories and individuals may have discs and equipment, such as the Xbox 360 in Figure 16, relating to this obsolete format.

Preservation Problems Associated with Optical Media

The most pressing short-term preservation concern with optical discs is the readability problems that result from incompatible formats. While most DVD-ROM drives can read CD-ROMs and CD-RWs, some cannot read CD-Rs. Due to incompatible methods of some proprietary dye-based recording technologies, some DVD-R drives cannot read CD-Rs. However, all DVD-RW drives can read CD-R and CD-RW discs (Saffady 2002, 47-48). Blu-ray discs cannot be read on conventional DVD drives, though HD DVDs can. Given the backward compatibility problems associated with optical media, it may be necessary for repositories to have several sets of disc drives to enable the continued readability of information stored on these discs.

Most optical discs suffer from decay, particularly CD-Rs and CD-RWs. This degradation results from exposure to heat, humidity, light, oxidation, accumulated scratches, and internal chemical reactions (O’Kelly 63). Although manufacturers claim that optical discs can last 50 to 212 years, these numbers are based on light and heat endurance tests that are performed in controlled laboratories. Such experiments do not take into account the negative effects of constant, and often careless, handling. Therefore, the actual life expectancy of optical discs could be between five and 20 years (O’Kelly 54). Currently, the archival properties of DVDs and Blu-ray discs have not been determined (Brown 2008, 5).

Flash Memory Devices

Flash memory devices have replaced floppy discs as the primary storage device for digital information. Although these devices have become popular only within the last decade, flash technology itself is quite mature. In the mid-1960s, metal-oxide semiconductor (MOS) technology made it possible to create high density and high performance memory chips. However, MOS chips were extremely volatile – that is, the stored information would be erased when the device in question was not powered. Flash technology was developed to create a storage device that would be non-volatile; although flash memory devices lose power when they are removed from a PC, the information stored on them remains until it is erased by the user. Flash technology is called so because the entire contents of the device can be erased at once in a process that is known as Fowler-Nordheim tunneling. Examples of storage devices that use flash memory technology include USB drives, memory sticks, and SmartCards, as Figure 17 illustrates (US Byte 2006a).

Figure 17: Selection of flash memory devices positioned next to a match for size comparison

Preservation Problems Associated with Flash Memory Devices

Because flash memory devices are relatively new compared to magnetic media or optical discs, the archival attributes of this type of media are not fully understood (Brown 2008, 5). Although flash memory sticks and cards are considered to be stable, since they do not have moving parts, the information contained on them appears to degrade over time. Consequently, flash memory devices are not suitable for long-term storage. Digital files that are stored on flash memory devices that are to be retained for long periods of time should be migrated to more robust storage mediums (Brown 2008, 6).

Conclusion

The devices used to store digital information are not as robust as paper and other analog mediums. Paper, when properly cared for, can persist for centuries, whereas magnetic tape has a maximum lifespan (if manufacturer claims are to be believed) of thirty years, and five to ten years for magnetic disks, CD-Rs, CD-RWs and flash memory devices. Storage devices for digital information can succumb to deterioration that may not always be discernable to the naked eye. Even if the storage media remains usable for several decades, changes in hardware will make it difficult for the information contained on the device to be accessed. Bogart sums up the situation when he writes:

> Some gold plated/glass substrate digital optical disk technologies promise 100-year lifetimes. However, a 100-year life expectancy is irrelevant when the system technology may be in use for no more than ten or twenty years (or less) (Bogart 1995, 15).

Older versions of the storage devices may not be compatible with the most recent media drives for that particular device. This can be of particular concern for corporate archives that may contain decades-old digital and electronic records on various forms of magnetic media and optical discs. Thus, even if the bitstreams of the original text, graphics, audio, or moving images of a museum's multimedia remain intact, a decayed storage device will render the information inaccessible. Should this occur, it would be impossible to recreate the project in question. To help remedy this situation, more emphasis needs to be placed on library science education programs that identify the archival properties of digital storage devices and how to extract the files contained on them when they begin to decay.

CHAPTER 6

THE CURRENT STATE OF DIGITAL PRESERVATION

In and of itself, the obsolescence of hardware is unimportant; technologically speaking, new hardware always has more capability than old hardware. So, in a sense, technological obsolescence introduces more capable machines into the public sphere. However, technical obsolescence carries a social cost. As computers and their storage media fail, knowledge is lost. Digital photographs, sound files, movies, and other important cultural artifacts may be lost forever if they reside on machines and in formats that fall into obscurity. Thus, for the sake of future generations, archivists and computer scientists must devise techniques to safeguard the long-term accessibility of digital information.

However, within the archival community, controversy exists as to which method of digital preservation will prove to be the most durable. As mentioned in Chapter 5, digital information exists in many different types of media, each with preservation requirements that are specific to a particular medium. Determining what these requirements are is dependent on a number of factors, including "the reason the record is being preserved, how long it needs to be preserved, the context and history of the record, and its original format (Digital Preservation Testbed 2003, 6)." Currently, four options are viewed as viable methods for long-term digital preservation: hardware conservation, migration, emulation, and XML. Although each of these methods has its own set of strengths and weaknesses, none can act as a "stand-alone" means of preservation.

Consequently, institutions that hold digital assets should draft preservation plans that utilize a variety of strategies.

This section consists of five parts. The first part describes hardware conservation, the practice of maintaining obsolete computer systems for the purpose of accessing files. The second part details migration, the transferring of digital files from an obsolete format to one that is more compatible with a modern computer system. The third portion explains emulation, the creation of a computer program that recreates obsolete operating systems and applications on modern computers (Besser 2000, 160). The fourth part focuses on XML, an open standards mark-up language that has shown promise as a means of archiving digital information. The section concludes with an examination of how institutions can create a multi-faceted long-term digital preservation program based on current best practices.

Hardware Conservation

The most conservative method of digital preservation is hardware conservation. This approach involves preserving obsolete hardware so digital files that are compatible with said systems can be accessed in their original environments (Digital Preservation Testbed 2003, 8). Some archivists have proposed establishing "computer museums" that would collect and maintain obsolete hardware systems on a large-scale. For a fee, users could bring obsolete storage devices, such as punch cards, floppy disks, and zip drives to the computer museum and access the desired information using the original platforms. Although this approach is the one that best preserves the integrity of documents, a number of serious problems with hardware conservation make it unfeasible (Borghoff et al. 2007, 14).

The success of a computer museum is dependent on the ability of the institution to assemble every computer ever made, from “hobby computers” like the Altair 6000 to modern MacBooks. Large mainframe computers that use data cards would have to be collected as well as desktop and laptop computers. Given the rapid evolution of technological devices, it would be difficult for a computer museum to adequately store and care for the multitude of operating systems, peripherals, and applications that would be needed to access every known file format. A computer museum would also have to accumulate different versions of software and a plethora of peripherals, such as joysticks, printers, and keyboards that are specific to certain platforms and often require their own third-party software and applications. Because computer manufacturers have shown little interest in servicing obsolete hardware, it is doubtful that old computers could be kept operational for an extended period of time (Borghoff et al. 2007, 15-16). Although the establishment of computer museums is not a feasible option for long-term digital preservation, large corporations and private individuals often keep obsolete computer hardware to retrieve files from these systems (Digital Preservation Testbed 2003, 8).

Migration

Migration is the most common method of digital preservation, as well as the most controversial. This approach involves converting bitstreams from an obsolete file format to a modern counterpart. An example of migration would be the conversion of the contents of a vinyl album to a MP3 format (NASCIO 2007, 3). It can also entail moving files from one hardware system to another. Migration has primarily been performed on simple digital objects, such as word processing files and bitmap images. It is unclear as to whether more complex digital files can be successfully migrated without substantially distorting the appearance and/or interface of the document (Wheatley 2007).

While the most basic definition of migration refers to the act of transferring the bitstreams of files encoded in obsolete platforms into new formats, a wide range of preservation activities fall under the blanket term “migration.” The simplest form of migration is data refreshing. This process is the creation of a copy of a digital file that is identical to the original, in terms of format and bitstream content. Refreshing is done when it is believed that a loss of data is about to occur because the original storage device is in danger of physical deterioration. Although data refreshing eliminates the fear of losing information from changing formats, it does not solve the problems associated with format obsolescence (Borghoff et al. 2007, 38).

In addition to data refreshing, Paul Wheatley of the CAMiLEON Project has identified five other kinds of migration. The first type is minimum migration. It entails retaining a copy of the digital object’s bitstream and slightly altering it to make the data easier to read. An example of this process is converting a Microsoft Word file to ASCII characters. This form of migration requires the least amount of human labor. Although this method preserves the informational content of a digital file, it eliminates the formatting and structure from the document. For this reason, minimum migration is not appropriate for more complex digital documents, such as audio, visual, or moving picture files. However, it may be appropriate as an emergency technique for preserving the intellectual content of text-based documents.

The second type of migration is preservation migration, which attempts to preserve the intellectual content of digital documents while retaining some of the visual and physical attributes of the original. At a minimum, this involves taking screen shots of the obsolete software in use. Annotated preservation migration uses screen shots and incorporates relevant metadata about the document that informs future users about the look and feel of the original interface. Complex preservation migration provides more detailed information about how the original file was accessed. For example, a word processing file might be accompanied by a video that explains how the original software and/or hardware functioned.

Recreation is the third type of migration. It necessitates re-coding a digital file in a modern computing environment. For a text file, recreation would entail retyping the information in a modern word processing file and then adding the proper formatting, so it resembles the original. More complex digital files would have to be completely re-coded. In some instances, a portion of the original file, usually the data, is included in the final object while a new front end is generated by the new platform. While recreation ensures that the look and feel of a file is faithfully reproduced on a new platform, it may be impractical for many institutions because it is time and labor intensive (Wheatley 2001).

Digital information can also be migrated to non-digital media. The most common method of digital to analog migration is printing out digital information to paper. While this approach may work for text-based documents and certain types of images, attempting to print out more complex digital documents, such as web sites and databases, loses significant behavioral and navigational aspects that cannot be recreated in a linear monograph format. The option of printing digital information to paper should be considered a temporary solution until a more comprehensive long-term digital preservation plan can be established (Digital Preservation Testbed 2001, 6-7).

Although migration is the most widely used method of digital preservation, it has a number of weaknesses. For example, it has traditionally been limited to relatively small files, such as word processing files, bitmap images, and spreadsheets. It is unclear how interactive and multimedia digital objects could be converted into new formats or if this would even be possible. Since the functionality of complex digital documents is heavily dependent on the format in which they originate, migration renders these objects unusable by transferring them to a foreign data environment (Wheatley 2001).

Another problem with migration is that the bitstream of the digital object must be transformed to preserve it. Migrating a document requires transforming the bitstreams of digital objects to make them readable in the new format. The result is that the archival integrity of digital objects is compromised by the changes that occur during the migration process. Even the most meticulous digital archivist cannot avoid altering the bitstream in such a way that does not change the content, structure, or appearance of the original object. As the document ages and more migrations are performed, the less the preserved copy will resemble the original. Even after the migration process is completed, there is still the possibility that the new software could introduce further changes to the bitstream when the file is accessed. This circumstance creates serious questions about the long-term integrity of digital data that has been subjected to multiple migrations (Digital Preservation Testbed 2003, 9-10).

Finally, migration is a labor intensive process that requires each file to be individually updated into the current technology. The corresponding metadata associated with each file must also be migrated and may have to be updated to accommodate the new platform. As it is often difficult to ascertain when a particular format has become obsolete, a platform can remain operational long after the software has ceased to be updated. Therefore, it is difficult to develop a cost model for migration activities, which in turn complicates the ability of the archival institution to formulate an accurate budget (Digital Preservation Testbed 2001, 11).

Emulation

Emulation, known in the computer science field as porting, keeps the original digital files and accesses them by recreating the original application and/ or operating system on a modern computer. It requires that a working copy of the original bitstreams of a given file be preserved, while the computer system that the document was created in is “migrated” to a modern platform. “Digital-born” documents, such as spreadsheets, hypertexts, and electronic carts, depend on their original rendering systems to provide context and functionality. Complex digital files often lose both of these characteristics when migrated. Many digital files now contain programming code, such as Javascript, PHP, or SQL, which makes them part of their computing environments. The purpose of any type of preservation activity is to safeguard the physical appearance and intellectual content of an artifact. While migration achieves the latter, it cannot do the former. Since migration often leads to a loss of formatting and the corruption of the bitstreams, proponents of emulation argue that the only way to accurately preserve digital documents is to ensure that future generations can access these files using a facsimile of the original platforms (Digital Preservation Testbed 2003, 18).

Although the concept of emulation as a long-term digital preservation technique is relatively new, emulators have been used for decades in the computer science field. Emulators test new software on older platforms while the hardware itself is still in development. In other instances, emulators might be placed in a new computer system so users can access files from a recent, but new incompatible format. For example, when the

Apple PowerPC initially came on the market, it was outfitted with an emulator for the obsolete Motorola 68000 CPU to accommodate applications and files that relied on the older processor to operate. Emulators are also commonly used by video game enthusiasts to play classic games on modern computers (Borghoff et al. 2007, 63).

Emulation has many advantages when compared to migration. First, it preserves the bitstream of the digital object without having to transform it, as is the case with migration. Bitstreams simply need to be preserved so the emulator can read them on the new platform (this assumes that bitstreams can be preserved indefinitely, a statement that has yet to be proven). Since migration cannot accurately preserve complex digital objects, emulation would be an ideal method for dynamic, multimedia documents. It would be perfect for documents that contain executable code that is specific to a certain operating system or application (Digital Preservation Testbed 2003, 22).

Second, emulation accurately preserves the “look and feel” of complex digital objects. While important platform specific aspects of a file may be lost during the migration process, emulation can correctly represent every facet of an obsolete computer system. Preserving the behavior and functionality of digital files is important for the purposes of ascertaining the authenticity of documents. Because migration alters a digital document’s bitstream, it can become difficult to determine the authenticity of a digital file as the number of migration cycles increase. As emulation leaves the bitstreams untouched, digital files that are accessed using emulation can maintain their integrity (Borghoff et al. 2007, 77).

Third, emulation acts as a “one size fits all” preservation method. Unlike migration, emulation can be used for all digital file formats, including text, hypertext, and code. Even if emulation is not used as the primary preservation method, it could be used as a back-up system, especially for file formats that are too obscure for other methods to recognize. Emulation is useful for documents that are particularly dependent on observing the behavior or appearance of the original computing environment (Digital Preservation Testbed 2003, 46).

Despite these advantages, emulation is inferior to migration in several ways. First, it interferes with the provenance and original order of digital objects. The creation of emulators sets a troubling precedent in which digital documents are classified according to their formats, rather than by their provenance and original order. Using emulators also eliminates the context in which digital objects originated. If emulators are to be used on an extensive basis, there must be some way to reconcile format-specific emulation technology with the need to preserve the original order and provenance of the documents (Thibodeau 2002).

Second, emulation requires a more substantial initial investment to develop the required technology for its implementation. Separate emulators for each of the obsolete technologies that are to be preserved also have to be produced, which requires further research and development. Therefore, it is not a realistic choice for documents that only need short-term preservation or for smaller institutions. However, expenses for emulator creation and maintenance could be shared between institutions as part of a consortium or other partnership (Oltmans and Kol 2007).

Third, even if emulators do prove successful at preserving continuous access to digital files, it is not known if future computer users would understand how to interpret the obsolete interfaces. Although keyboards, mice, and graphical user interfaces (GUIs) have been part of the de facto design of home computers for more than thirty years, this may not be the case in 50, 100, or 200 years. If researchers in the year 2300 use an emulator to view a Word 2007 file, it is uncertain whether they would know how to interpret a scrollbar, graphical icons, a menu bar, or any of the other common features of early twenty-first century computers. Since the purpose of an emulator is to faithfully recreate the look and feel of an older computer system, it is uncertain whether it would be possible to use an emulator without a certain level of hardware preservation to ensure that users are communicating with the computer in the same way that people did when the original system was in common use. However, if no one in the twenty-fourth century knows how to use a keyboard or a mouse, using an emulator could require a steep learning curve for would-be researchers (Digital Preservation Flatbed 2003, 28).

Finally, there is the problem of the rapid pace of technological obsolescence, both for computer hardware and software, and for the emulators themselves. Each emulator is only built to recreate one particular platform (e.g., Apple II, Commodore 64, Windows 95). With new software, application systems, and operating systems being released at a staggering pace, an institution could potentially create hundreds or even thousands of separate emulators to accurately preserve all of the digital documents it may have accumulated. This scenario does not take into account the difficulties that would be required to emulate documents that originated on devices other than desktop computers, such as virtual machines or client-server applications. Emulators themselves eventually become obsolete. Once this occurs, emulators require migration to a newer platform or the creation of one that is compatible with the latest technology. Therefore, it may not be practicable to create emulators for such a wide diversity of operating systems, programs, and computing machines (Thibodeau 2002).

Presently, emulation is more of a theoretical method of long-term digital preservation than one that is widely used in archival repositories. Though computer scientists have used emulation for decades, it has only been within the past ten to fifteen years that it has been seriously considered as an archival preservation tool. The purpose of emulators in the computer science field is to test hardware and software, not to provide accurate representation of files written in obsolete formats for a prolonged period of time (Digital Preservation Testbed 2003, 47). In fact, migration is the main technique used for the preservation of large amounts of digital objects. Stewart Granger notes that leading archival entities have no intention of using emulation as their principal digital preservation tool. While this does not mean that there is no place for emulation in "real world" conservation activities, archival institutions should take into account that emulation is more of an experimental technology than migration (Granger 2006).

XML

eXtensible Mark-up Language (XML) has shown promise as a non-platform specific form of digital preservation. XML is a web standard that was first adopted by the World Wide Web Consortium in 1998 and revised in 2000. The primary purpose of XML is to provide digital documents with a clearly delineated structure and meaning. XML is easy for humans and computers to read, which means that it is likely that this language will continue to be used on the computers of the future. Because XML is an open source standard, it can be incorporated as part of a larger migration and/or emulation strategy (Slats and Verdegem 2004, 8).

XML is related to the more commonly used Hypertext Mark-up Language (HTML), but is much more powerful. XML and HTML are both derived from Standard General Markup Language (SGML), an influential, but seldom used markup language that was established as an independent standard in 1985. Unlike HTML, the form and content of XML-based documents are strictly separated. This means that it is possible to view the code for a single document using different representations. The main difference between XML and HTML is that element definitions are not fixed in the former as they are in the latter. Hence, users can create their own elements and structure documents according to the needs of their particular files. However, XML must conform to basic syntax rules, or the file will not validate (Borghoff et al. 2007, 104-105).

Because XML is independent of any particular platform or software, it has quickly become the “lingua franca” of digital data exchange, similar to the status of Latin in the Middle Ages. This has occurred because the bulk of the proprietary information in a document is related to its outward appearance. Since XML separates form from content, the intellectual content of a file may be able to persist for long periods of time (Digital Preservation Testbed 2002, 25).

XML can be used as a long-term digital preservation solution in several ways. For example, a digital document can be encapsulated in a XML “wrapper” that contains metadata about how to access a particular file. Instructions written in XML could be included with an emulator that contains information about how to access a file (Digital Preservation Testbed 2002, 27). It is also possible to migrate files in obsolete formats to XML. Because XML files are organized in schema such as DTD and XSLT, it is easy to migrate structured documents like databases and spreadsheets to XML (Digital Preservation Testbed 2002, 28).

The major concern regarding the use of XML as a long-term digital preservation strategy is whether it will be in use 20, 50, or 100 years from now (Digital Preservation Testbed 2002, 26). Given the continuous evolution of computer languages in general, it is not known whether documents written in 2008's version of XML will be usable in the decades to come or if it will be superseded by another language. Since new programming languages are invented on a regular basis, XML itself could become obsolete. However, even if XML does become a "dead language," it might remain understandable by future generations of humans and computers if the proper documentation is retained. Such a situation is analogous to how documents written in the dead languages of antiquity can still be accessed today (with the proper training), even though these languages are not in regular use (Digital Preservation Testbed 2002, 25).

Like other forms of migration, converting a file to XML changes the look and feel of a document. XML eliminates extraneous data and replaces it with metadata. Thus, the end product of XML migration will look very different from the original document. From a legal perspective, this creates concerns about how to ascertain the authenticity of a digital record that has been migrated (Digital Preservation Testbed 2002, 29).

Converting documents to XML can also be a costly endeavor. Compared to simple binary-based documents, XML files are much larger, and therefore more expensive to store. Consequently, institutions storing digital documents in XML would have to compress them. Not only does compression further alter the structure of file bitstreams, but there is also the possibility that future users would not understand how to decompress the files. This would necessitate creating an algorithm that would tell future computers how to extract the data from the compressed files. Although it is possible to write such algorithms, doing so takes a considerable amount of time, money, and effort, which is probably impractical for most institutions (Borghoff et al. 2007, 112).

Hence, while XML has a number of advantages that make it an ideal solution for many long-term digital preservation issues, more research needs to be conducted on the ability of XML-based documents to survive over long periods of time and on the degree of functionality XML can preserve. For example, during the course of experiments conducted by the Digital Preservation Testbed, it was discovered that:

> … [although] the conversion to XML is suitable to represent the context, content and structure of the database itself…[W]e were not able to preserve behavior of database systems for the longer term using migration or XML…Hardware emulation could be a potential approach in this respect, but has not been implemented with an archival focus.

These findings indicate that it may not be possible to preserve the functionality of all types of digital formats using XML (Slats and Verdegem 2004, 16).

Conclusion

To sum up, no single method of preservation can be used for every type of digital document. Because digital preservation is a new field, more research needs to be done to determine how the four methods discussed here can be used in tandem to best protect the integrity of digital files. However, archivists and librarians cannot wait for the results of research that has yet to be conducted when deciding how to protect the digital files in their care. As the next section about the history of the Legacy Project will illustrate, in some instances, a complete reinterpretation of digital information may be necessary to save the intellectual content of a digital object, even if it means abandoning the original interface altogether.

CHAPTER 7

THE LEGACY PROJECT: A CASE STUDY

During the 1990s, many museums became intrigued by the ways in which digital technology could enhance the museum experience. Of particular interest were the ways in which digital technology could revamp the museum's ability to function as a vehicle for storytelling. This is particularly important in the case of history museums, where the goal is to make large, impersonal events meaningful to the visitor on an individualized, interpersonal level. According to Mouw and Spock**:**

> [M]useum visitors most readily connect to history through the personal stories of others. Cognitively, what takes place is a series of comparisons between one's own experiences and another's…People are less interested, at least at first, in what happened, why and when. Rather they want to understand, through a process of personal introspection or through sharing an experience with others…

An effective museum exhibit, be it artifact-based or digital, tells a compelling narrative that creates an experience that has more in common with the theater than the classroom. Museum visitors are more likely to learn from the exhibit when they relate personally to the individuals who actually experienced the historical event in question. The interactive and immersive qualities of digital media make them ideal for creating an effective narrative for the museum environment (Mouw and Spock 2007, 47).

The views noted above were part of the reasoning behind the construction of the Legacy Project, a multimedia installation that told the stories of Holocaust survivors that settled in the Atlanta area. Legacy was intended to teach young people about the Holocaust in a way that a traditional exhibit could not, by letting survivors speak about their experiences in their own words via streaming video. Listening to an actual person speak about how the Nazi regime affected them as an individual creates a narrative experience that is lacking from a text-based resource that speaks of the Holocaust in generalized terms. However, approximately three years after Legacy was completed, it became inaccessible due to hardware and software failure and technical obsolescence.

This chapter is divided into six parts. The first section is a general history of the William Breman Jewish Heritage Museum and the Legacy Project. The second section describes Legacy's interface. This part contains many screenshots of the various sections and sub-sections that comprised the Legacy Project. The third section details Legacy's technological specifications. The fourth section describes the technical problems that led to Legacy's final demise. This segment also provides information on the Breman's plans to relaunch Legacy as a web-based exhibition. The fifth section compares Legacy with the Birmingham Civil Rights Institute's Richard Arrington, Jr. Resource Center, a digital installation that was the inspiration for the Legacy and had almost identical technological specifications. Unlike Legacy, the Richard Arrington, Jr. Resource Center has remained operational since it was established in 1998. This section concludes with a discussion of the obsolescence of site-based installations.

The Genesis of the Legacy Project

Figure 18: Exterior of the William Breman Jewish Heritage Museum

The William Breman Jewish Heritage Museum (henceforth known as "the Breman") was established in 1992 to chronicle Jewish history in Atlanta. The genesis of the Breman was a 1983 exhibit organized by the Jewish Federation of Greater Atlanta entitled *Jews and Georgians: A Meeting of Cultures, 1733-1983*. The exhibit, which detailed the contributions of the Jewish community to Georgia's history, received acclaim from critics and the public. However, after the exhibit ended, the displayed artifacts, which had been culled from the attics and storage closets of private homes, synagogues, and businesses, were returned to their owners. This situation highlighted the need for a permanent facility with which to preserve, chronicle, and exhibit Atlanta's Jewish history (SEMC Presentation, 1).

The Breman opened in its current space as seen in Figure 18 with two permanent exhibitions: *Creating Community: The Jews of Atlanta from 1845 to The Present* and *Absence of Humanity: The Holocaust Years* (see Figure 19). The latter exhibit ended with a section entitled *New Lives,* which gave a brief overview of how Holocaust survivors built new lives in Atlanta after liberation (SEMC Presentation, 1). Because of budgetary constraints, the *New Lives* section was not as effective as it could have been; aside from a short loop of interviews with survivors discussing their experiences, this section consisted primarily of low-quality photographs that had been expanded and turned into museum exhibition panels. (SEMC Presentation, 2).

By the late 1990s, the Breman was interested in adding an interactive component to its offerings. This interest, combined with a sustained commitment to document the experiences of Holocaust survivors in the state of Georgia, made a multimedia Holocaust exhibit a natural choice. The purpose of what would become the Legacy Project was to create an interactive educational resource and memorial to Atlanta's Holocaust survivor community. The urgency of this project was based on the dwindling number of Holocaust survivors and the museum's desire to record their experiences for future generations. Hence, the idea was proposed to create a multimedia exhibit that would integrate text, photographs, video, and web-based resources to teach visitors, particularly young people, how Holocaust survivors created new lives in the Atlanta area (SEMC Presentation, 2-3).

Figure 19: Photography of a portion of the Holocaust Gallery: Absence of Humanity: the Holocaust Years, 1933 - 1945

Alek and Halina Szlam, both of whom were children of survivors, provided initial funding for Legacy (SEMC Presentation 2). The Hewlett-Packard Corporation and the Fulton County Commission, under the guidance of the Fulton County Arts Council, provided additional funding. Work began on Legacy in 2000, and it opened to the public on April 7, 2002. Legacy consisted of four computer workstations that played video clips of Holocaust survivors, displayed family photographs, and contained text-based biographies (see Figure 20). Interactive maps detailed the history and fates of Europe's Jewish communities and provided information about the ghettos and concentration camps. A fifth computer workstation was Internet accessible and enabled visitors to find web-based resources pertaining to the Holocaust (Legacy Project Blurb-1).

Figure 20: Two of the Legacy Project's workstations from the period in which it was operational

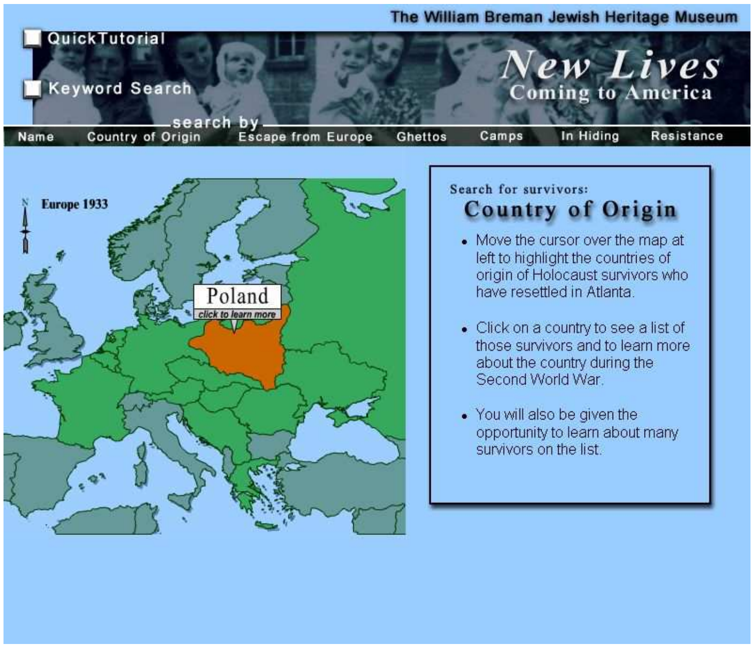
The Legacy Project's Interface

Legacy's interface was designed to be understandable to individuals with limited computing experience. The background of the opening interface consisted of an enlarged photograph of a Jewish refugee family onboard a ship that was en route to the United States (see Figure 21). Three other photographs were positioned in a column on the left-hand side of the screen. The top of the page had a tabbed horizontal menu bar that contained different methods of browsing for survivor profiles. Although the navigation bar identified these tabs as portals for searching, it would be more accurate to say that they enabled users to organize the survivor lists according to pre-determined criteria. Users could sort the biographical content by name, country of origin, or by the manner in which the individual survived the Holocaust: escaping from Europe either before or after World War II started, imprisoned in a ghetto and/or concentration camp, by hiding, or as a member of the Resistance. Above the navigation bar were two additional links that led to a user tutorial and a search engine with which to conduct a keyword search.

Figure 21: Screenshot of the opening interface for the Legacy Project

The interfaces for the browse sections either contained interactive maps or a mixture of text and images. “Country of Origin,” “Ghettos,” and “Camps” were in the former category. Each of these sections contained an interactive, color-coded map of inter-war Europe. The color scheme for the maps was as follows: red for Greater Germany and Occupied territories, pink for German allies or dependent states, and gray for neutral countries. Users in the “Country of Origin” section could click on a specific country to get a list of the Atlanta-based Holocaust survivors from the nation (see Figure 22). Each country page also contained information about how the Holocaust manifested itself in that particular nation (see Figure 23). The “Ghettos” and “Camps” pages used the same map as the “Country of Origin” section, but employed iconography to indicate the locations of Jewish ghettos and concentration camps. The icon used for the “Camps” section (see Figure 24) was a stylized watchtower, while the “Ghettos” section (see Figure 26) used a building covered with barbed wire. When an icon belonging to a specific camp (see Figure 25) or ghetto (see Figure 27) was clicked, users were directed to a page in which they could find out more about the institution in question and were provided with a list of individuals who were imprisoned there.

The other sections consisted of lists of survivor profiles integrated with text and expository text. The “Name List” section was simply an alphabetized list of all of the survivors. Although this option brought back the most number of results, there was no way of finding more specific information about the names unless the user clicked on each profile individually. The “Escaping Europe” portion had a list of individuals who survived the Holocaust by escaping the Continent. There was a brief essay about the logistical, financial, and emotional difficulties in leaving Europe in an environment in which Jewish refugees were viewed with suspicion and hostility. The survivors in this section included individuals who had escaped Europe prior to the outbreak of World War II, those who left the Continent via underground networks, and those who illegally immigrated to British Palestine. Similarly, the “Resistance” section contained a short explanation of the ways in which Jews fought against the annihilationist Nazi regime, in addition to a list of survivor profiles (see Figure 28). The part entitled “In Hiding” included an essay about how some Jews survived the Holocaust by being hidden by private individuals, organizations, or entire communities, juxtaposed with a list of survivor profiles (see Figure 29). It should be noted that these survivor lists were not mutually exclusive, as some survivor profiles could be found on more than one list.

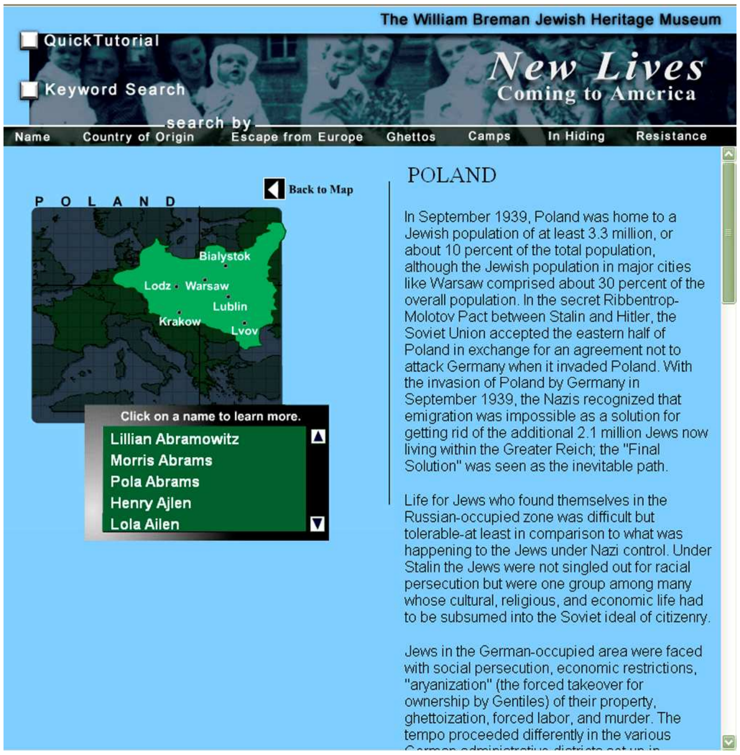


Search for survivors:

Country of Origin

- Move the cursor over the map at left to highlight the countries of origin of Holocaust survivors who have resettled in Atlanta.
- Click on a country to see a list of those survivors and to learn more about the country during the Second World War.
- You will also be given the opportunity to learn about many survivors on the list.

Figure 22: Screenshot of the "Country of Origin" section



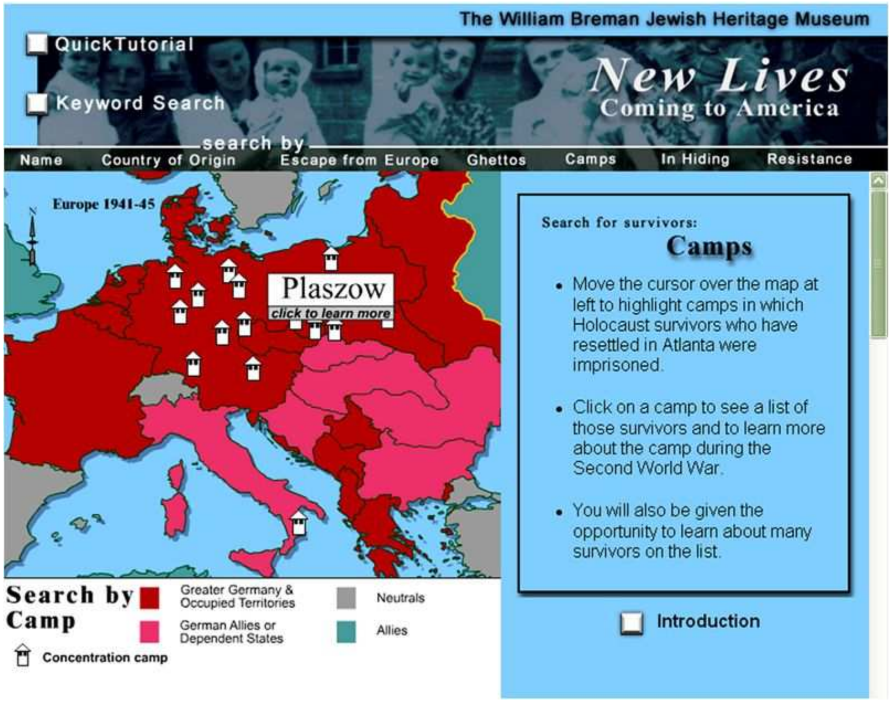
POLAND

In September 1939, Poland was home to a Jewish population of at least 3.3 million, or about 10 percent of the total population, although the Jewish population in major cities like Warsaw comprised about 30 percent of the overall population. In the secret Ribbentrop-Molotov Pact between Stalin and Hitler, the Soviet Union accepted the eastern half of Poland in exchange for an agreement not to attack Germany when it invaded Poland. With the invasion of Poland by Germany in September 1939, the Nazis recognized that emigration was impossible as a solution for getting rid of the additional 2.1 million Jews now living within the Greater Reich; the "Final Solution" was seen as the inevitable path.

Life for Jews who found themselves in the Russian-occupied zone was difficult but tolerable-at least in comparison to what was happening to the Jews under Nazi control. Under Stalin the Jews were not singled out for racial persecution but were one group among many whose cultural, religious, and economic life had to be subsumed into the Soviet ideal of citizenry.

Jews in the German-occupied area were faced with social persecution, economic restrictions, "aryanization" (the forced takeover for ownership by Gentiles) of their property, ghettoization, forced labor, and murder. The tempo proceeded differently in the various

Figure 23: Screenshot of the survivor list and country profile of Poland



Search for survivors:

Camps

- Move the cursor over the map at left to highlight camps in which Holocaust survivors who have resettled in Atlanta were imprisoned.
- Click on a camp to see a list of those survivors and to learn more about the camp during the Second World War.
- You will also be given the opportunity to learn about many survivors on the list.

Figure 24: Screenshot of the interface for the "Camps" section

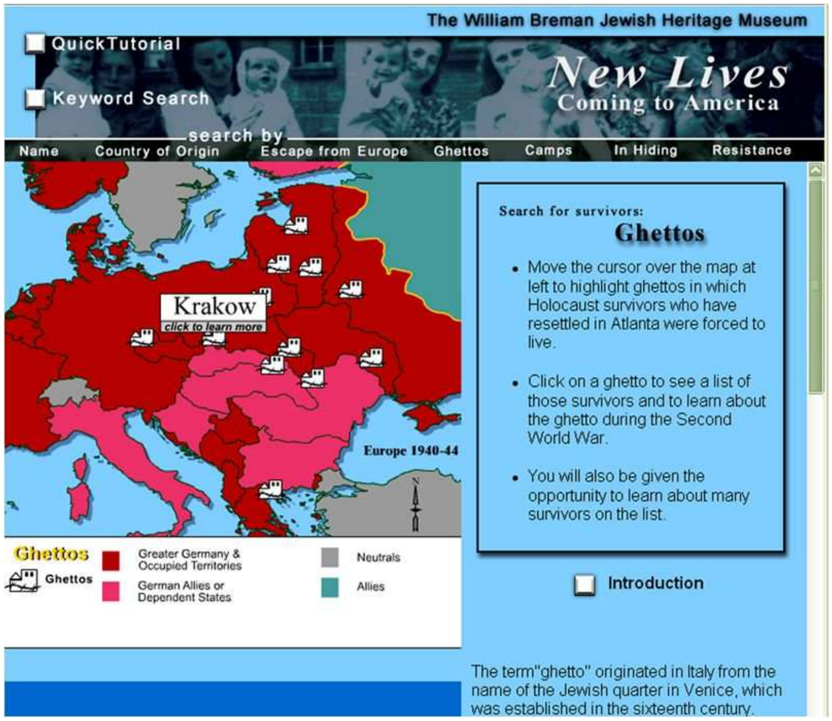
Jewish women at forced labor pulling hopper cars of quarried stones in Plaszow.

PLASZOW

Plaszow was established in the summer of 1942 in a suburb of Krakow, Poland. It was originally a forced labor camp, but in 1944 it became a concentration camp. The camp was built near a quarry next to a railroad station and atop two Jewish cemeteries that were razed. The Germans desecrated the Jewish graves by stealing the tombstones, which were then broken up and used as paving stones for roads in the camp.

Originally, Plaszow held about 2,000 prisoners, but after the liquidation of the Krakow ghetto on March 13 and 14, 1943, six thousand Jews from the ghetto were sent there. Plaszow continued to grow in size and number of prisoners until it covered almost 200 acres and contained, at its peak period of operation, between 22,000 to 24,000 prisoners, including about 7,000 Jews from Hungary. Poles were also imprisoned there as well as several dozen families of gypsies and German criminals, who were employed

Figure 25: Screenshot of the sub-section about Plaszow concentration camp



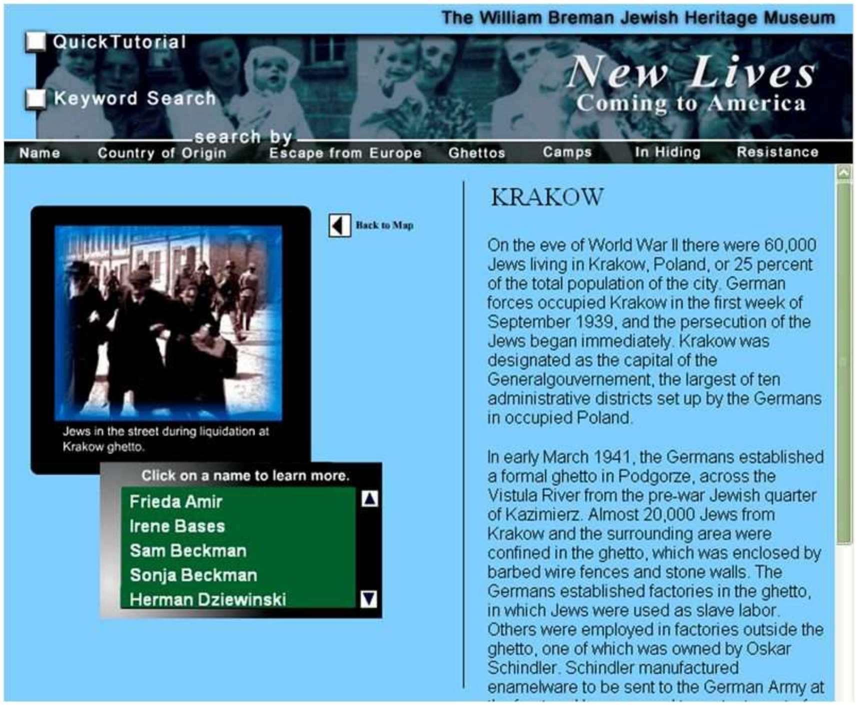
Search for survivors:

Ghettos

- Move the cursor over the map at left to highlight ghettos in which Holocaust survivors who have resettled in Atlanta were forced to live.
- Click on a ghetto to see a list of those survivors and to learn about the ghetto during the Second World War.
- You will also be given the opportunity to learn about many survivors on the list.

The term"ghetto" originated in Italy from the name of the Jewish quarter in Venice, which was established in the sixteenth century.

Figure 26: Screenshot of the interface for the "Ghettos" section



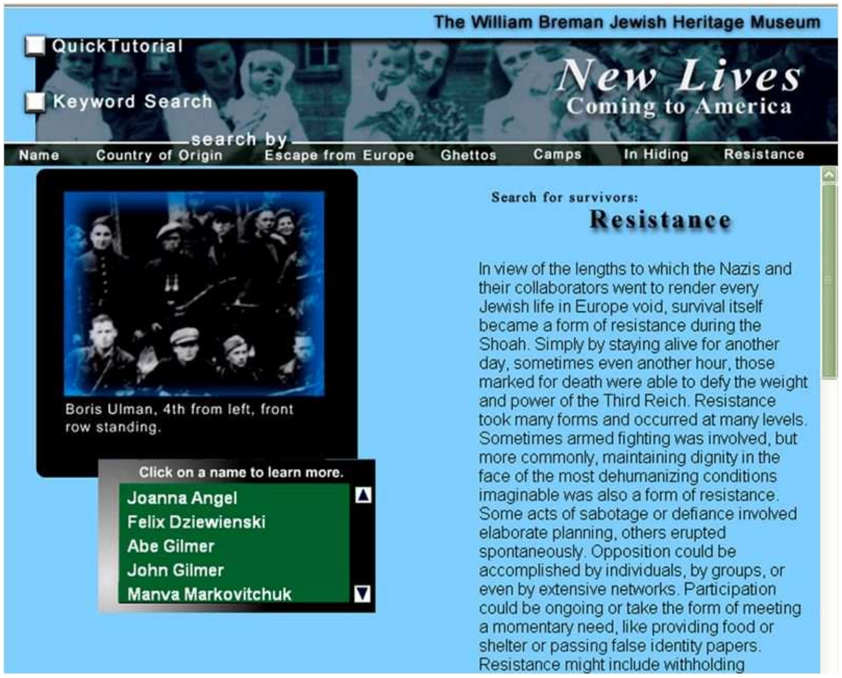
Jews in the street during liquidation at Krakow ghetto.

KRAKOW

On the eve of World War II there were 60,000 Jews living in Krakow, Poland, or 25 percent of the total population of the city. German forces occupied Krakow in the first week of September 1939, and the persecution of the Jews began immediately. Krakow was designated as the capital of the Generalgouvernement, the largest of ten administrative districts set up by the Germans in occupied Poland.

In early March 1941, the Germans established a formal ghetto in Podgorze, across the Vistula River from the pre-war Jewish quarter of Kazimierz. Almost 20,000 Jews from Krakow and the surrounding area were confined in the ghetto, which was enclosed by barbed wire fences and stone walls. The Germans established factories in the ghetto, in which Jews were used as slave labor. Others were employed in factories outside the ghetto, one of which was owned by Oskar Schindler. Schindler manufactured enamelware to be sent to the German Army at

Figure 27: Screenshot of the sub-section about the Krakow ghetto

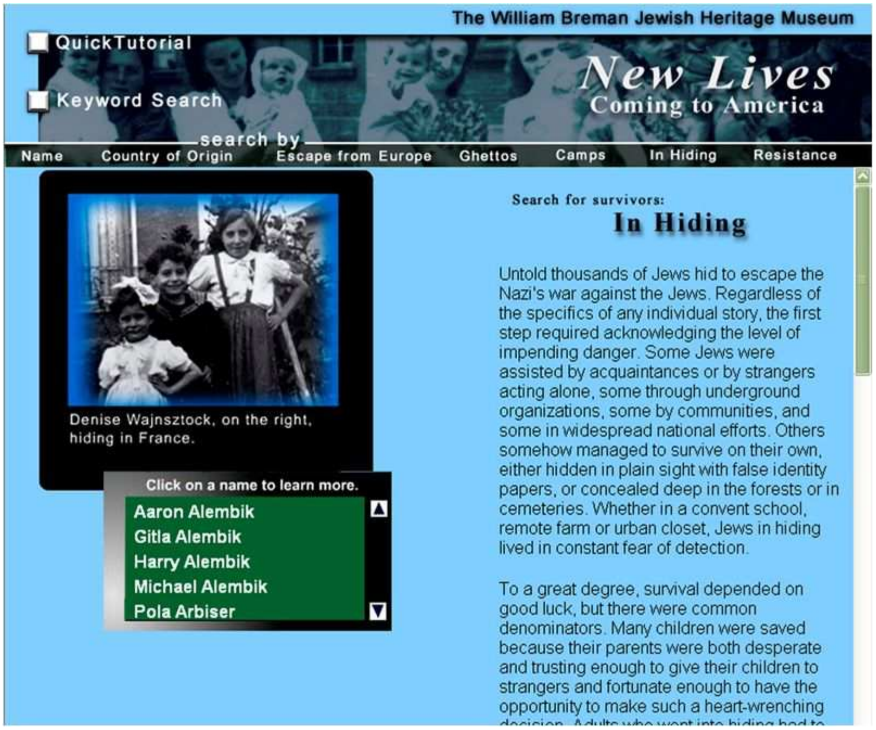


Search for survivors:

Resistance

In view of the lengths to which the Nazis and their collaborators went to render every Jewish life in Europe void, survival itself became a form of resistance during the Shoah. Simply by staying alive for another day, sometimes even another hour, those marked for death were able to defy the weight and power of the Third Reich. Resistance took many forms and occurred at many levels. Sometimes armed fighting was involved, but more commonly, maintaining dignity in the face of the most dehumanizing conditions imaginable was also a form of resistance. Some acts of sabotage or defiance involved elaborate planning, others erupted spontaneously. Opposition could be accomplished by individuals, by groups, or even by extensive networks. Participation could be ongoing or take the form of meeting a momentary need, like providing food or shelter or passing false identity papers. Resistance might include withholding

Figure 28: Screenshot of the "Resistance" section



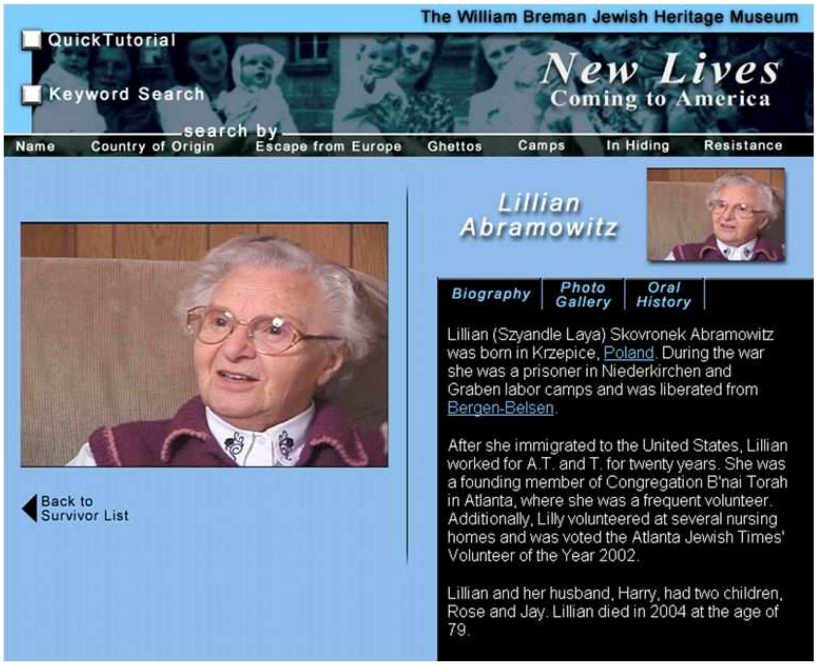
Search for survivors:

In Hiding

Untold thousands of Jews hid to escape the Nazi's war against the Jews. Regardless of the specifics of any individual story, the first step required acknowledging the level of impending danger. Some Jews were assisted by acquaintances or by strangers acting alone, some through underground organizations, some by communities, and some in widespread national efforts. Others somehow managed to survive on their own, either hidden in plain sight with false identity papers, or concealed deep in the forests or in cemeteries. Whether in a convent school, remote farm or urban closet, Jews in hiding lived in constant fear of detection.

To a great degree, survival depended on good luck, but there were common denominators. Many children were saved because their parents were both desperate and trusting enough to give their children to strangers and fortunate enough to have the opportunity to make such a heart-wrenching decision. Adults who went into hiding had to

Figure 29: Screenshot of the "In Hiding" section

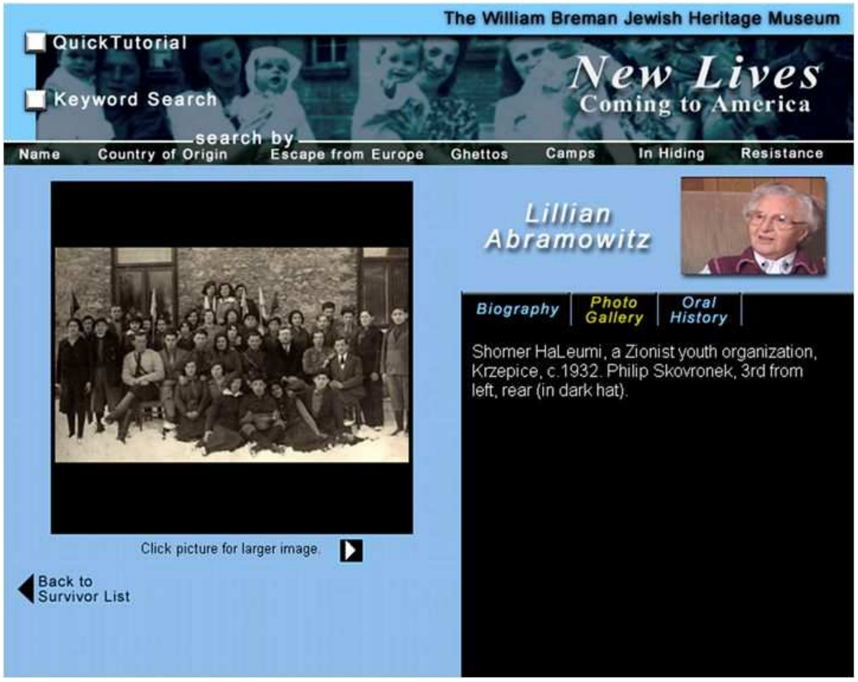


Lillian (Szyandle Laya) Skovronek Abramowitz was born in Krzepice, Poland. During the war she was a prisoner in Niederkirchen and Graben labor camps and was liberated from Bergen-Belsen.

After she immigrated to the United States, Lillian worked for A.T. and T. for twenty years. She was a founding member of Congregation B'nai Torah in Atlanta, where she was a frequent volunteer. Additionally, Lilly volunteered at several nursing homes and was voted the Atlanta Jewish Times' Volunteer of the Year 2002.

Lillian and her husband, Harry, had two children, Rose and Jay. Lillian died in 2004 at the age of 79.

Figure 30: The "Biography" portion Lillian Abramowitz's survivor profile



Shomer HaLeumi, a Zionist youth organization, Krzepice, c.1932. Philip Skovronek, 3rd from left, rear (in dark hat).

Figure 31: Screenshot of the "Photo Gallery" section of Lillian Abramowitz's survivor profile

Shomer HaLeumi, a Zionist youth organization, Krzepice, c.1932. Philip Skovronek, 3rd from left, rear (in dark hat).

Figure 32: Enlarged photograph from the "Photo Gallery" section of Lillian Abramowitz's survivor profile

Survivor profiles were multimedia in nature. Each biography was divided into three parts: "Biography," "Photo Gallery," and "Oral History." The "Biography" section contained a text-based biographical sketch of the individual in question (see Figure 30). The "Photo Gallery" contained photographs of the individual's life prior to the Holocaust (see Figures 31 -32). The "Oral History" section provided users with digitized video of the survivor speaking about his or her experiences in his or her own words.

Technical Specifications

Legacy consisted of six main parts: the desktops, the video server, the external RAID array for the video server, the web server, the SQL server, and the peripherals. Each computer was a Dell OptiPlex GX110 desktop, priced at $1250 each. The Dell OptiPlex contained a Pentium III processor, running at 933 MHz, with a 256 full-speed cache. The computers also contained a 1.44 MB 3.5 floppy drive and a 12X DVD-ROM drive. There were integrated Intel 3D graphics with direct AGP and 4MB Display Cache. One of the computers was Internet accessible via an Integrated 3Com Etherlink 10/100 with ACPI and Remote Wake-up only. Audio was provided by an integrated sound blaster compatible sound (AC97 Audio). The operating system utilized was Windows 2000 Professional (SP1) (Hardware Specs 1).

The video server contained a Pentium III Xeon CPU unit with a 700MHz/1M cache that cost $6,780. Like the computers, the video server included a 1.44MB disk drive for 3.5-inch floppy disks. However, rather than a DVD-ROM drive, the video server had an internal 650MG CD-ROM drive. The RAID 1 set had a 2x 18GB hard drive. The video server contained two hard drives. The first was an 18GB 1" Ultra3 SCSI, 10K RPM hard drive with a dual channel 128 MB RAID card. This hard drive also contained two channels, one internal and one external channel. The second was an 18GB 1' Ultra3 SCSI 10K RPM hard drive. The operating system was a Windows2000 server and a 4GB utility partition. The video server had a 512MB SDRAM – 4DIMMS memory (Hardware Specs 1-2).

The external RAID 5 array for the video server was a Dell PowerVault 20xS SCSI 8-drive external SCSI Storage system. It cost $6,352 and contained four hard drives. All four were 36GB SCSI 10,000 RPM. The SCSI support options were Ultra3SCSI, cluster ready. A second redundant power supply was also available (Hardware Specs 2).

The web server was a Dell PowerEdge 2650. The CPU unit was a 1 GHz Intel Pentium III with a 256K cache. Like the video server, it had four hard drives, each of which was an 18GB 1'' Ultra3 (Ultra160) SCSI 10K RPM Hot Plug hard drive. The total cost was $6,617. It also contained a 1.44MB 3.5-inch floppy disk drive (Hardware Specs

2). The data was backed up with an internal 20/40GB DDS-4 tape drive with a 39160 SCSI controller. The tape drive backup software was CA Arcserve Enhanced. The web server contained a 24X IDE Internal CD ROM drive. It ran on a Windows2000 operating system and a single non-redundant 330-watt power supply (Hardware Specs 3).

The SQL server was a Dell PowerEdge 1550. The CPU unit was a Pentium III 1GHz with 256K cache, and it contained a single processor and a 256MB SDRAM, 4 DIMMS memory. This server had two hard drives, each of which was an 18 GB Ultra3, 1', 10K RPM SCSI hard drive. The primary controller was a PERC3-DCL RAID card with 64MB cache, and a 1 Int/1 Ext channel. As with the other servers, the SQL server had a 1.44 MB 3.5-inch floppy disk drive. It also contained a 24x IDE CD-ROM drive and a front keyboard mouse connected via a Y cable. The SQL server ran on a Windows2000 server operating system (Hardware Specs 3).

Legacy also consisted of peripheral devices, in addition to the hardware. These included five Samson QS headphone amps, ten Telex DH3V headsets, five Kensington trackballs, five Ketronic classic keyboards (each covered with a plastic "skin"), five TrippLite 700VA UPS, and five KDS 18" kiosk mount touch screens. The workstation that was Internet accessible had a Samsung 17" LCD flat screen monitor (Hardware Spec 4).

Legacy's multimedia content was created using Macromedia Producer. An HTML framework with an SQL backend enabled the multimedia content (text, photographs, video) to be generated dynamically for each page. The video server maintained the streaming video, while the web server contained the text and photographs. The workstation that was Internet accessible contained a custom-made proprietary browser that was developed to prevent users from accessing inappropriate content. (personal communication, R. Einstein, February 24, 2009).

The Legacy Project Falls Victim to Technological Obsolescence

Figure 33: The Legacy Project as it appears today, dormant and unusable

Two years after its completion, the Legacy Project began to experience technical problems. The software for the video server had been problematic since the time of its installation, necessitating the complete reloading of the video files on several occasions. Two of the video server's hard drives eventually failed and the entire system became inoperable in 2005. Since the video server is the only part of the backend that does not work, the non-video portions of the Legacy Project are theoretically still accessible. However, of the five computer workstations that comprise Legacy, only one works on a consistent basis (personal communication, R. Einstein, February 5, 2009). Because all of the components of Legacy were custom-made, the long-term sustainability of the project was only possible if the businesses that manufactured the parts remained in operation. Neither the software nor the hardware could be fixed because the company that manufactured them, Openshaw Media Group, had gone out of business by the time the technical problems manifested (R. Einstein, personal communication, September 18, 2008).

Although the original Legacy Project workstations lie dormant at the Breman, there are plans to resurrect it as an exhibition. Since the textual, photographic, and video data from Legacy were derived from existing information, the same materials will be used, but reformatted for the Internet. This endeavor is part of a larger effort to redesign and repackage the entire Breman Museum web site. Free Range Graphics, an Atlanta-based web design firm, has been hired to spearhead this venture (R. Einstein, personal communication, September 18, 2008). The web-based Legacy Project will be Flash-based with a redesigned interface that will match the overall theme of the new museum site. Unlike the original, the new Legacy Project will have the flexibility to add new profiles as they become available. It is hoped that, in addition to functioning as a web-based resource, the new Legacy Project will be integrated into the museum's Holocaust Gallery (R. Einstein, personal communication, February 5, 2009). Free Range Graphics will also be responsible for the reinterpretation of Legacy (R. Einstein, personal communication, September 18, 2008).

Comparison with the Birmingham Civil Rights Institute's Richard Arrington, Jr. Resource Center

Figure 34: The first Richard Arrington, Jr. Resource Center at the Birmingham Civil Rights Institute

The decision to create Legacy was based on a similar installation that had been created for the Birmingham Civil Rights Institute (BCRI) in 1998. However, unlike the Legacy Project, the BCRI's exhibit remained operational, until the original installation was completely refurbished in 2008. Because the Openshaw Media Group was responsible for creating the installations for both the Breman and the BCRI, and the technical specifications for both were similar, the author speculated as to why the latter still works, but not the former.

The BCRI is a combination museum, archive, and research facility located in Birmingham, Alabama. It was established in 1992 to document the history of the Civil Right Movement in Birmingham (Birmingham Civil Rights Institute 2007). The BCRI is located in the city's Civil Rights District, which is also home to the Sixteenth Street Baptist Church, the Kelly Ingram Park where Civil Rights marches were often held, and the historic black business district (Woolfolk 2007).

Figure 35: Close-up of one of the workstations at the first Richard Arrington, Jr. Resource Center at the Birmingham Civil Rights Institute

The purpose of the Richard Arrington Jr., Resource Center (henceforth referred to as “The Resource Center”) was to integrate archival content with multimedia technology and to provide visitors with an “experience” that could not be conveyed by traditional research materials. The design phase for the Resource Center lasted from February 1997 to December 1998. It was opened to the public in January 1999. Users could examine digitized photographs, magazine articles, newspaper clippings, and radio and television broadcasts from Birmingham’s Civil Rights era. The Resource Center was much bigger than Legacy, containing 16 workstations as opposed to five. As all of the material for the Resource Center was produced from the archives, none of the computers were Internet-accessible. Unlike Legacy, where users communicated with the computer via a keyboard and tracking ball, the Resource Center used a touch screen interface. It was believed that touch screen technology would be more intuitive for individuals who were not accustomed to using computers (W. Coleman, personal communication, February 9, 2009).

Although all 16 of the original workstations were still operational by 2008, the design of the Center was considered outdated. Thus, the BCRI decided to upgrade the Resource Center to give it a more modern appearance, both in terms of physical demeanor and interface. The new Resource Center contains flat-screened monitors and runs on the Windows 2003 operating system. The touch screen system found in the original design was abandoned in favor of a traditional mouse and keyboard interface. The materials that users have access to have been greatly expanded, reflecting additions to the BCRI's archival holdings. The BCRI's goal is to make this resource as expansive as possible. A condensed version of the materials that are available at the Resource Center will soon be available as an online resource. Currently, the Resource Center is the BCRI's primary digital asset. However, the institute hopes to add a digitization project to its web site in the near future. The company responsible for the redesign of the Resource Center was the Atlanta-based Melia Design (W. Coleman, personal communication, February 9, 2009).

Unlike Legacy, the decision to upgrade the Resource Center was purely aesthetic, and was not based on technical obsolescence or hardware failure. Although the Resource Center did experience some hardware problems, these difficulties were never as severe as those that plagued Legacy. This suggests that the Breman may have received a defective video server. Another advantage that the BCRI had was that Openshaw Media Group was based in Birmingham, meaning that when technical problems did arise, they could be quickly resolved by the manufacturer (W. Coleman, personal communication, February 9, 2009).

Conclusion

Although Legacy was meant to integrate the latest digital technology in the museum environment, it was already obsolete when it was completed. By the early 2000s, site-based digital installations had been rendered outmoded by online exhibitions and digitization projects. The percentage of Americans who had regular Internet access had increased to the point that most users had come to expect a digital asset like Legacy to be available in an online form. It was assumed that family and friends of survivors would come to the Breman specifically to use Legacy and that other visitors would be willing to sit for hours to use the installation. However, this was seldom the case, because much of the same information about the Holocaust could be accessed via a home Internet connection. The fact that Legacy included a workstation that was specifically set aside for Internet access illustrates that many of the assumptions that the Breman made about how this resource would be used were incorrect. Therefore, it would have been more sensible for Legacy to have been built as an online exhibition from its conception (R. Einstein, personal communication, November 10, 2008).

Although the Richard Arrington, Jr. Resource Center is also a site-based digital installation, it is large enough to accommodate an entire classroom, which encourages school groups to come to the BCRI to do research. Unlike Legacy, which is cordoned off in a small corner at the end of the Breman's Holocaust Gallery, the Resource Center is seamlessly integrated into a larger exhibit on human rights, which makes the latter more inviting for visitors to use. The success of the BCRI's Resource Center vis a vis Legacy suggests that museums that are interested in pursuing digital projects should consider the needs of their visitor base, as well as the current state of digital technology.

CHAPTER 8

ISSUES RAISED BY THE LEGACY PROJECT

The story of the Legacy Project raises several issues concerning long-term digital preservation that have not been addressed in existing literature. Although most discussions of digital preservation concern themselves with how to maintain both the intellectual content and the technical context in which the digital file in question originated, these two goals are often mutually exclusive, and the former usually triumphs over the latter. This section explores some of the implications of integrating digital technology into the museum setting. The first section explains how reinterpretation complicates traditional archival theory. The second section details the role of digital installations in the museum environment. The third section illustrates the difficulties involved in trying to preserve web sites.

Reinterpretation as a challenge to traditional archival theory

In general, discussions of long-term digital preservation revolve around debating the relative merits of migration, emulation, or XML. The solution that the Breman chose to rescue its digital asset was reinterpretation – recreating the work using modern technology. Reinterpretation constitutes the most radical form of digital preservation, because archivists have free reign in determining how the work should be reconstructed (Depocas et al. 2003, 97). However, for some digital objects, reinterpretation may be the only way of saving a lost digital work, especially when the obsolete hardware is replaced by another device that performs the same function (e.g., when the teleprinter was replaced by e-mail and cell phones). Since musical and dramatic performances are reinterpreted each time they are performed, some authors have suggested that the ephemeral nature of digital files makes them more akin to time-based performance art, than physically instantiated artifacts (ERPANET 2004, 6).

Because reinterpretation violates the most basic principles of archival theory, it is seldom discussed in the literature. Indeed, when one examines the properties of digital files using the five archival principles outlined by Gilliland-Swetland, it becomes obvious that the standards that have guided preservation activities for almost 200 years are no longer relevant. Gilliland-Swetland lists the following as being the foundational principles of archival theory:

- the sanctity of evidence
- respect du fonds
- the life cycle of records
- the organic nature of records
- hierarchy in records (Gilliland-Swetland 2000, 9)

Sanctity of evidence refers to the preservation of the context in which the records in question were created. This is accomplished by maintaining a paper trail that documents the custodial history of the records (Gilliland-Swetland 2000, 10). In the context of digital objects, this evidence is usually metadata that describes the intellectual and administrative characteristics of the artifact in question. However, during the digital preservation process, much of this metadata can be lost or corrupted, especially in the case of migration (Gilliland-Swetland 2000, 11). During the reinterpretation process, the metadata of the original becomes irrelevant, because an entirely new version of the digital object has been created. Although the end product is based, to some degree, on the original, the reinterpreted object is a distinct object with its own evidential metadata. In the context of reinterpretation, the evidential process is only useful as a blueprint to create new versions of the digital object at a later data.

Respect du fonds, also known as provenance or original order, states that records should be classified according to the institution from which they originated. This also means that archivists should endeavor to keep the previous owner's original order intact as much as possible (Gilliand-Swetland 2000, 12). However, in the digital environment, respect du fonds loses its meaning. Since digital files can be organized in various ways on a storage device or hard drive, it would be a mistake to say that these objects have an original order. For example, on the Windows XP operating system, it is possible to view the contents of a folder in five ways: thumbnails, tiles, icons, list, and details. The view that a user chooses has nothing to do with the order or manner in which the files were created; it is simply a matter of personal preference. Because many organizations, including museums, keep digital records on a common drive, it can be difficult to ascertain the exact origin of a specific folder or file from within the organization. When a complex digital object is reinterpreted, the place of origin may be completely different from where the original was created. For example, the first Legacy Project was created by Openshaw Media Group, whereas the new version will be produced by Free Range Studios. Although the web-based Legacy is supposed to be an updated version of the original, it is essentially a new creation with a different set of metadata.

The life cycle of records is a model that represents the different stages in the life of a record. The life cycle begins with the creation and use of the record in question. As the record ages, the information contained in it will become less relevant, and it will be referenced more infrequently, until it becomes inactive. When this occurs, the record is either deemed superfluous and subsequently destroyed, or it is sent to an archive (Gulliland-Swetland 2000, 14). When the latter option is chosen, the record is integrated with other records of the same provenance (Gulliland-Swetland 2000, 15).

The life cycle principle assumes the existence of documents that are readable without the aid of machines (except for microfilm or celluloid film), that records can be ignored for years or even decades without being processed, and that the archive is a "dumping ground" for records that are no longer needed by the donating organization. Most noticeably, the decision of whether or not to archive paper-based records is conducted when the document is no longer in active use or when it is donated for historical posterity. However, it is imprudent for an archivist to wait until the end of the life cycle of a digital file, because the document may already be inaccessible. As the Legacy Project illustrates, when a digital object reaches the end of its life cycle, that probably means that it has become inaccessible, not that its organization of origin has finished using it (Day 1999). Since storing digital objects in a physical repository like a paper-based record is not possible, reinterpreting helps extend the life cycle of a digital object so it can remain accessible.

The organic nature of records refers to the relationships that exist between records in the same collection, as well as the various contexts (e.g., historical, legal, procedural) in which they were created (Gilliland-Swetland 2000, 16). However, it is not always possible to discern the relationships between the digital files in a particular folder or storage device. Although files that are part of the same complex digital object are stored in the same folder, there is no way of determining the relationship between these files, unless an information architecture chart is included. When a complex digital object is reinterpreted, the previous relationship that existed between the files is no longer valid. For example, the web-based version of the Legacy Project will have a different interface and more profiles than the original. It will be completely recoded using Actionscript and an updated version of SQL. The navigational differences between the two versions of Legacy will be so great, that no organic relationships can be said to exist between them.

Last, hierarchy in records refers to the order that is imposed on records by virtue of the filing and organization process (Gilliland-Swetland 2000, 18). Archival finding aids reflect these hierarchies by providing detailed descriptions of each layer of the collection. Finding aids document the provenance, arrangement, and custodial history of the records in question and provide a logical way to find materials at the item level. (Gilliland-Swetland 2000, 19). Since digital files are not organized in the same way that paper-based records are, the hierarchy that arises naturally for the former is not appropriate for the latter. Although some digital objects can be described using a markup language like XML, the relationship of digital files to each other is generally too diffuse for a strictly hierarchical finding aid to be helpful.

While reinterpretation may violate the bedrock principles of archival theory, the experiences of both the Breman and the BCRI illustrate that it can be a sensible method of digital preservation, especially if additional material will be added or subtracted from the end product on a regular basis. Unlike books, paintings, films or other physically instantiated objects which have a fixed period of time in which the creator works on them until they are considered finished, complex digital objects, such as web sites, social networking sites, or digitization projects are constant works in progress; the content of these objects continues to change, until the author(s) decides to cease work on the undertaken endeavor. When this occurs, the digital object usually becomes inaccessible, because the author no longer wishes to invest the time or resources needed to maintain it. Unlike paper-based records or other physically instantiated artifact, the end of the lifecycle for the digital object typically means the end of its existence. Thus, reinterpretation is the norm rather than the exception for complex digital objects.

Digital projects and museums

The fate of the Legacy Project also raises questions about the role of digital projects in history museums, namely whether such creations should be considered tools or artifacts. If a digital asset is simply a tool, then the creating institution should be able to change the content or appearance of the object to suit whatever needs the museum may have at a given point in time. However, if a digital asset is an artifact, then it must be subjected to the same preservation criteria as a physically instantiated object. Based on the evidence gathered at the Breman and the BCRI, most museums seem to view digital installations as tools.

Digital projects are created to fulfill specific institutional goals, which include (but are not limited to): outreach, access to collection holdings, activism, and education. Like physically instantiated exhibits, digital projects must be maintained over time to ensure that the institution gets a suitable return on its investment (Steinbach 2007, 110). Because digital projects are large investments, particularly for a small niche institution like the Breman, a museum has to make the most use of its digital assets. Therefore, if a problem occurs with the project, the best interests of the museum are to keep it functional, even if that means completely reinterpreting the object in question. Many institutions do not view their digital assets as potential artifacts, because the very term "artifact" indicates that an object has reached the end of its life cycle. A museum would probably not consider a digital installation or web site to be an item of enduring value that merits continuous preservation work after its utilitarian role for the institution has ceased.

Issues pertaining to the long-term preservation of web sites

Although the Breman's plans to recreate the Legacy Project as an online exhibition solve the short-term problem of making the original accessible again, this decision does not address the issue of ensuring Legacy's long-term viability. Web sites have the shortest lifespan of any type of digital media. The average web page remains online for 75 days, giving archivists and librarians little time to preserve the document of web sites that detail major political, historical, or cultural events. For example, the official web site of the 2000 Sydney Olympics disappeared soon after the games ended. Although the National Library of Australia was able to save several iterations of the Sydney Olympics site before the home page went offline, thousands of sites of enduring value (e.g., campaign sites, war blogs, photo galleries) are permanently lost. The short lifespan of web sites is largely due to the relative ease in which they can be set up and removed; while it may take six months to several years to write a book, a simple HTML-based web site can be online in less than a day and taken down in a matter of minutes. Thus, if web site creators become bored with the subject matter of their web sites or blogs, removing then takes little effort (Day 2003, 6). Although it is unlikely that the future web-based Legacy Project will one day disappear because of lack of interest on the part of the Breman, it is likely that Legacy could migrate to a different technological form, rendering the web site obsolete.

Another problem with preserving web sites is that they are dynamic creations. As mentioned previously, the content of a web site is in a constant state of flux. Given the changing nature of web aesthetics, individuals and organizations often completely change the appearance of their respective home pages to give the site a more modern appearance. To effectively preserve even a single web site would necessitate archiving a version of the site in each iteration. However, it is not possible to view previous versions of a web site unless one has access to the original code (Day 2003, 7). For example, the content of the original Legacy Project was codified during the planning stages, whereas new materials will be added to the web-based version on a regular basis.

Web sites can also suffer from technological obsolescence. Although many web standards have remained consistent, the way in which web-based content is delivered has changed. The content for many web sites is created using dynamic databases, which cannot be easily replicated in an archival environment. Other sites use browser plug-ins or non-standard features that an archive may have difficulty finding to ensure that the web site in question looks as it should (Day 2003, 7).

Conclusion

Although reinterpreting the Legacy Project as a web-based exhibition solves the short-term problem of keeping the intellectual content accessible to the public, it does not ensure that it will be accessible in 50 or 100 years. Like all digital technologies, web standards and technological changes occur very rapidly. The World Wide Web itself may become obsolete, thus forcing the web-based Legacy Project to be reinterpreted in a yet to be developed future digital environment. As it works to make the Legacy Project accessible again, the Breman must ask itself how long it plans to invest in the effort and what function this resource should play for the institution. A plan for what steps should be taken when the museum's digital assets become obsolete or experience technical failures should be outlined during the planning phase, as well as in the institution's emergency management plan. Doing so will prevent the digital project in question from entering into a long period of inoperability, as was the case with Legacy.

CHAPTER 9

CONCLUSION

Guaranteeing the continuous access of complex digital artifacts will necessitate changing many long-standing definitions of archival theory and practice. The goal of traditional preservation activities is to ensure that the artifact in question remains the same, whereas digital preservation almost always requires the alteration of the object. Currently, whether an institution uses migration, emulation, XML, or reinterpretation, it is impossible to preserve the intellectual content of a digital object, as well as the context in which the object was created. Even emulation can only recreate the hardware and software of an obsolete computer system, not the peripherals needed to communicate with the interface (Chen 2001, 3). Since the archival properties of physically instantiated artifacts, regardless of the specific material, are very different from their "born digital" counterpart, forcing digital files to fit into the traditional archival paradigm is akin to trying to fit a square peg into the proverbial round hole.

Rather than attempting to make digital information adhere to traditional archival theory that was developed specifically for physically instantiated artifacts, the archival community should create a new set of principles based on the unique archival properties of digital media. Any proposed digital archival theory needs to be based on the following principles: authenticity, detailed metadata, and preservation of intellectual content.

Assuring the authenticity of digital files is a major concern because of the ease in which digital information can be copied or altered. This situation is particularly problematic since many digital files are "born digital" with no analog original with which to compare. Migrating digital files also compromises authenticity by unintentionally corrupting data bitstreams or software functionality (Dollar 2000, 25). Ensuring the authenticity of digital files requires the establishment of a trusted digital archive that allows users access to these files without allowing them to be altered (Dollar 2000, 26). Authenticity can be further guaranteed by documenting the technological specifications of each piece of hardware and software that is involved in the file creation process (Saffady 2002, 105). Each stage of the process should be recorded as evidence in the metadata logs. This includes media handling, storage conditions, password controls, and lists of which individuals have access to the file in question (Saffady 2002, 106). These procedures will ensure that digital files will retain their intellectual and legal authenticity.

Digital archivists should keep detailed metadata records for digital files to allow future users to understand the context in which the files were created. Metadata can also be used to recreate a digital file or object if the original bitstreams are no longer available. The types of metadata that would be retained include bibliographic data (e.g., author, title, date of publication, publisher, keywords, location), data format, and copyright information. When appropriate, a migration log detailing the number and type of digital environments that the file has been saved in should be kept. Likewise, if an emulator is utilized, there should be metadata about the original platform (including the peripherals that accompany the hardware) and a description of how the emulator software operates (Borghoff et al. 2005, 9). Existing metadata standards such as Encoded Archival Description (EAD) could be used or expanded to achieve this end (Gilliland-Swetland 2000, 25).

Although archivists should try to save as much of the digital context of a file as is possible, priority should be given to preserving the intellectual content. Digital files are heavily dependent on certain platforms and applications to provide context, and it will probably be impossible to preserve the digital environment in which they were created for long periods of time. For example, the main priority for archivists working on Portico (a digital archive of electronic peer reviewed journals) is to preserve the intellectual content (e.g., text, images, tables) of the articles. While Portico attempts to save some of the functionality of the original, its look and feel is seldom retained (Library of Congress). Archivists need to take screenshots of digital files in their original environments to provide future users with documentary evidence of how they functioned on their native platforms, if this information enhances the intellectual content. However, in most cases users will be interested in the intellectual content, not the appearance of digital files.

The emerging field of knowledge management already utilizes many of these concepts noted. Gilliland-Swetland defines knowledge management as:

> the practices, skills, and technologies associated with creating organizing, storing, presenting, retrieving, using, preserving, disposing of, and re-using information to help identify, capture and produce knowledge. Knowledge management activities can include data and metadata mining as well as digital asset management.

Although knowledge management has been primarily confined to the corporate world, it could be the next stage in the evolution of archival theory; digital archives exist not only as repositories for materials of historical value or to ensure access to information, but also to safe keep administrative and legal documents that can affect an organization’s profits and governing. As is the case with digital archives, knowledge management systems often contain a mixture of paper-based, digitized, and “born digital” records as well as digital images, audio, and video (Gilliland-Swetland 2000, 26). The knowledge management paradigm considers issues pertaining to digital file retention at every stage of the information creation process. In some systems, digital files are automatically transferred to an asset management system where a digital project manager or archivist prepares them for secondary use. Digital archivists and librarians should investigate the knowledge management approach as a basis with which to administer digital files and objects (Gilliland-Swetland 2000, 27).

With this in mind, archivists need to work with the creators of museum-conceived digital assets to ensure that the lifespan of the end product is as long as possible. The input of the archival perspective during the planning and construction stages can help an institution plan the long-term maintenance of a digital project, rather than simply focus on short-term gains. Archivists should also become more educated about the technology behind digital information to allow them to better understand the care needed to protect the digital assets in their custody. Professional organizations such as the Society of American Archivists should offer continuing education courses in topics like XML, HTML, SQL, and digital storage device management. As the number of "born-digital" objects of enduring value proliferates, information professionals must be prepared to preserve them to ensure that they are accessible for future generations.

REFERENCES

ARMA International. 2009. “What is Records Management? Why Do I Care?” http://www.arma.org/pdf/WhatIsRIM.pdf (accessed April 5, 2009).

Barnes, Ian. 2006. “Preservation of word processing documents.” http://www.apsr.edu.au/publications/word_processing_preservation.pdf. (accessed January 21, 2009).

Battalina, Michael. 2004. “The GIF Controversy: A Software Developer's Perspective.” http://purl.org/mcb/id/giflzw (accessed January 25, 2009).

Bearman, David. 1999. “Reality and Chimeras in the Preservation of Electronic Records.” *D-Lib Magazine* 5 (4). http://www.dlib.org/dlib/april99/bearman/04bearman.html (accessed February 5, 2009).

Bergeron, Bryan. 2002. *Digital Dark Ages II: When the Digital Data Die*. Upper Saddle River: Prentice Hall.

Besser, Howard. “Digital Longevity.” In *Handbook for Digital Projects: A Management Tool for Preservation and Access*, edited by Maxine K. Sitts. Andover: Northeast Document Center, 2000. 155-166.

Besser, Howard. “Longevity of Digital Art.” 2001. http://www.gseis.ucla.edu/~howard/Papers/elect-art-longevity.html (accessed September 18, 2008).

Birmingham Civil Rights Institute. “Frequently Asked Questions.” *Birmingham Civil Rights Institute Home Page.* http://www.bcri.org/general_information/FAQs/faqs.htm (accessed February 21, 2009).

The Blue Ribbon Task Force on Digital Preservation and Access. 2008. “Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation.” http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf (accessed February 5, 2009).

Borghoff, Uwe M, Rodig, Peter, Scheffczyk, Jan, and Schmitz, Lothar. *Long-Term Preservation of Digital Documents: Principles and Practices*. New York: Springer, 2005.

Brown, Adrian. 2003. “Digital Preservation Guidance Note: Graphics File Formats.” http://www.nationalarchives.gov.uk/documents/graphic_file_formats.pdf (accessed January 24, 2009).

Brown, Adrian. 2008. “Digital Preservation Guidance Note 3: Care, handling, and storage of removable media: National Archives." http://www.nationalarchives.gov.uk/documents/removable-media-care.pdf (accessed February 8, 2009).

Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving. Washington, D.C.: Council on Library and Information Resources and Library of Congress, 2002.

The CAMiLEON Project. "BBC Domesday." *The CAMiLEON Project Home Page.* http://www.si.umich.edu/CAMILEON/domesday/domesday.html (accessed September 30, 2008).

Chen, Su-Shing. "The Paradox of Digital Preservation." http://www.media-matters.net/docs/resources/Digital%20Preservation/ParadoxOfDigitalPreservation.pdf (accessed February 28, 2009).

Colorado Digitization Program Digital Audio Working Group. 2006. "Digital Audio Best Practices." http://www.bcr.org/cdp/best/digital-audio-bp.pdf (accessed January 30, 2009).

Conway, Paul. "Overview: Rationale for Digitization and Preservation." In *Handbook for Digital Projects: A Management Tool for Preservation and Access*, ed. Maxine K. Sitts. Andover: Northeast Document Center, 2000. 5-20.

The Commission on Preservation and Access and the Research Libraries Group. 1996. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*. http://www.ifla.org/documents/libraries/net/tfadi-fr.pdf

Darlington, Jeffrey, Finney Andy, and Pierce, Adrian. "Domesday Redux: The rescue of the BBC Domesday Project videodiscs." http://www.ariadne.ac.uk/issue36tna/. (accessed February 5, 2009).

Day, Michael. 1999. "Metadata for digital preservation: an update." *Ariadne* (22). http://www.ariadne.ac.uk/issue22/metadata/ (accessed February 5, 2009).

Day, Michael. 2003. "Collecting and preserving the World Wide Web: A feasibility study undertaken for the JISC and Wellcome Trust." http://www.jisc.ac.uk/uploaded_documents/archiving_feasibility.pdf (accessed February 26, 2009).

Deegan, Marilyn and Tanner, Simon. 2006. Key Issues in Digital Preservation. In *Digital Preservation* edited by M. D. a. S. Tanner. Cornwall: Facet Publishing. 1-31.

Depocas, Alain and others, ed. "The Variable Media Approach: Permanence through Change." New York: The Solomon R. Guggenheim Foundation, New York, and The Daniel Langlois Foundation for Art, Science, and Technology, Montreal, 2003.

http://www.variablemedia.net/pdf/Permanence.pdf (accessed September 8, 2008).

Deitel, P.J. Deitel and H.M. 2007. *Java: How to Program*. Seventh Edition ed. Upper Saddle River: Prentice Hall.

Digital Preservation Coalition. 2002. "Digital Preservation – Strategic Overview." *Digital Preservation Coalition Home Page.* http://www.dpconline.org/graphics/digpres/stratoverview.html (accessed September 12, 2008).

ERPANET. 2004. "The Archiving and Preservation of Born-Digital Art Workshop." http://www.erpanet.org/events/2004/glasgowart/briefingpaper.pdf (accessed February 26, 2009).

Gagnier, Richard, Lafaille, Madeleine, Tolmatch, and Elaine, Zeppetelli. 2008. "New Media Art in Museum Collections: A Report from the DOCAM." Cataloguing and Conservation Committees, in J. Trant and D. Bearman (eds.). Museums and the Web 2008: Proceedings, Toronto: Archives and Museum Informatics. http://www.archimuse.com/mw2008/papers/gagnier/gagnier.html (accessed February 2, 2009).

Gilheany, Steve. 1998. "Preserving Information Forever and a Call for Emulators." http://www.archivebuilders.com/aba010.html. (accessed February 10, 2009).

Gilliland-Swetland, Anne J. 2000. *Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment.* Washington, D.C.: Council on Library and Information Resources.

Granger, Stewart. 2000. "Emulation as a Digital Preservation Strategy." *D-Lib Magazine*. http://www.dlib.org/dlib/october00/granger/10granger.html (accessed December 6, 2008).

Hoeven, Jeffery van der. 2007. "Dioscuri: Emulator for Digital Preservation." *International Preservation News* (43):23-24.

Hunter, Gregory. *Developing and Maintaining Practical Archives*. New York: Neal-Schuman Publishers, 2003.

Hunter, Gregory. *Preserving Digital Information*. New York: Neal Schuman Publishers, 2000.

Jones, Maggie and Beagrie, Neil. 2001. *Preservation Management of Digital Materials: A Handbook*. London: The British Library.

Lesk, Michael. "Preserving Digital Objects: Recurrent Needs and Challenges." *Home Page of Michael Lesk.* http://www.lesk.com/mlesk/auspres/aus.html (accessed January

21, 2009).

Leskien, Hermann. “Introduction.” Borghoff, Uwe M, Rodig, Peter, Scheffczyk, Jan, and Schmitz, Lothar. *Long-Term Preservation of Digital Documents: Principles and Practices*. New York: Springer, 2005.

Library of Congress. “Digital Preservation Pioneer: Eileen Fenton.” http://www.digitalpreservation.gov/partners/pioneers/detail_fenton.html (accessed March 8, 2009).

Mackintosh, Barry. “Chapter 3 – Audiovisual Innovations.” *Interpretation in the National Park Service*. http://www.nps.gov/history/history/online_books/mackintosh2/directions_av_innovations.htm (accessed February 28, 2009).

MacLean, Margaret and Davis, Ben H, eds. *Time and Bits: Managing Digital Continuity*. Los Angles: The J. Paul Getty Trust, 1998.

O’Kelley, Terence. “Reference Guide for Optical Media.” *Memorex, Inc. Home Page.* http://www.memorex.com/downloads/whitepapers/Reference%20Guide%20for%20Optical%20Media%209-9.pdf (accessed February 10, 2009).

Oltmans, Erik and Kol, Nanda. “A Comparison Between Migration and Emulation in Terms of Cost.” *RLG DigiNews Home Page.* http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file959.html (accessed November 18, 2008).

O'Shea, Greg. March 1996. “Keeping Electronic Records: Issues and Strategies.” *Provenance: The Electronic Magazine* (No. 2), http://www.netpac.com/provenance/vol1/no2/features/erecs1a.htm (accessed February 4, 2009).

Pearce-Moses, Richard. “Life Cycle.” *SAA: Glossary of Archival and Records Terminology Home Page.* http://www.archivists.org/glossary/term_details.asp?DefinitionKey=247 (accessed March 1, 2009).

Rinehart, Richard. “The Media Art Notation System: Documenting and Preserving Digital/Media Art.” http://www.coyoteyip.com/rinehart_leonardo.pdf (accessed February 20, 2009).

Rothenberg, Jeff. January 1999. *Avoiding Technological Quicksand: Finding a Viable Technological Foundation for Digital Preservation*. Washington, D.C.: Council on Library and Information Resources.

Rothenberg, Jeff. “Ensuring the Longevity of Digital Information.”

http://www.clir.org/pubs/archives/ensuring.pdf.

Saffady, William. 2002. *Managing Electronic Records*. 3rd edition ed. Lenexa: ARMA International.

Sitts, Maxine K., ed. *Handbook for Digital Projects: A Management Tool for Preservation and Access*. Andover, Mass.: Northeast Document Conservation Center, 2000.

Slats, Jacqueline and Verdegem, Remco. 2004. "Practical Experiences of the Dutch Digital Preservation Testbed." http://www.iwaw.net/05/papers/iwaw05-hoeven.pdf (accessed November 30, 2008).

Steinbach, Leonard. 2007. "Analyzing Return on Investment: Process of Champions." *The Digital Museum: A Think Guide*. Ed. Herminia Din and Phyllis Hect. Washington, D.C.: American Association of Museums.

Stumm, Casey. "Preservation of Electronic Media in Libraries, Museums, and Archives." *The Moving Image*, Volume 4, Number 2, Fall 2004, pp. 38-63.

Technical Advisory Service for Images. 2005. "An Introduction to the Vector Image Format." http://www.tasi.ac.uk/advice/vector/pdf/vector-intro.pdf (accessed January 28, 2009).

Thibodeau, Kennneth. "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years." http://www.clir.org/pubs/reports/pub107/thibodeau.html (accessed October 5, 2008).

Unicode , Inc. *The Unicode Standard: A Technical Introduction.* http://www.unicode.org/standard/principles.html (January 19, 2009).).

USByte. "Flash Memory." 2006a. *USByte Home Page*. http://www.usbyte.com/common/Flash_memory_1.htm (accessed February 12, 2009).

USByte. "Flash Memory." 2006b. *USByte Home Page*. http://www.usbyte.com/common/Flash_memory_3.htm (accessed February 12, 2009).

Van Bogart, John W. C. 1995. "Magnetic Tape Storage and Handling: A Guide for Libraries and Archives." http://www.imation.com/america/pdfs/AP_NML_doc_magtape_S_H.pdf (accessed January 15, 2009).

van Nispen, Annalies, Kramer, Rutger, and van Horik, Rene. 2005. "The eXtensible Past: The Relevance of XML Data Format for Access to Historical Datasets and a Strategy for Digital Preservation." *D-Lib Magazine* 11 (2). http://www.dlib.org/dlib/february05/vannispen/02vannispen.html (accessed December 6, 2008).

Vitale, Tim. 2007. “Digital Image File Formats - TIFF, JPEG, JPEG2000, RAW and DNG.” http://aic.stanford.edu/sg/emg/library.pdf/vitale-digital-image_file_formats.pdf (accessed January 23, 2009).

Wheatley, Paul. “Migration – a CAMiLEON discussion paper.” *The CAMiLEON Home Page.* http://www.ariadne.ac.uk/issue29/camileon/intro.html (accessed November 4, 2008).

Woolfolk, Odessa. “General Information.” *The Birmingham Civil Rights Institute Home Page.* http://www.bcri.org/general_information/history_and_mission/history.htm (accessed February 21, 2009).

Wilson, Andrew, Wright, Richard, Plfreman, Malcolm, Anderson, Sheila, Tanner, Simon, Beer, Emma. 2006. “Digital Moving Images and Sound Archiving Study.” http://www.jisc.ac.uk/media/documents/programmes/preservation/moving%20pictures%20and%20soundarchving%20final%20version.doc (accessed January 30, 2009).

Yeung, Tim Au. 2004a. “Digital Preservation: Best Practices for Museums.” http://www.chin.gc.ca/English/Pdf/Digital_Content/Digital_Preservation/digital_preservation.pdf (accessed February 2, 2009).

Yeung, Tim Au. 2004b. “Digital Preservation for Museums: Recommendations.” http://www.chin.gc.ca/English/Pdf/Digital_Content/Preservation_Recommendations/preservation_recommendations.pdf (accessed February 1, 2009).