

CHAPTER 6 DIGITAL PRESERVATION

The preservation of cultural heritage materials is a cornerstone of our work as professionals. Ensuring the longevity of the materials entrusted to our care is a concern no matter what other activities you undertake. Clay tablets, stored properly, will last thousands of years. Good rag paper, if kept away from pests and in the right environment, will last hundreds of years. The proof of their longevity lies in museums and rare book collections around the world. Someday, future curators, archivists, and librarians will know how long a floppy disk, a CD, or hard drive will last, but until then we have only estimates.

The creation of digital surrogates in no way alleviates these concerns. Instead, it presents a new arena with preservation concerns. This chapter discusses the preservation of digital materials. It does not cover issues associated with traditional preservation. Those measures will not be discussed in light of the voluminous material already available to meet those challenges. The life of inadequately stabilized and housed original documents, artifacts, published works, or works of art will not be extended through the process of digitization. Digitization will do nothing to help the condition of the original. In short, digitization is not a preservation measure for originals, even though it may mitigate against further damage by providing access to surrogates.

Institutions can invest a great deal of effort and expense in the digitization of their collections and in the presentation of these digitized collections online. Until recently, the digital world has not given high priority to the preservation and storage of its content. The very nature of the Web is one of impermanence and fluidity, and digital information presented via the Internet has grown with little regard for preservation.

Fortunately, cultural institutions by their very nature think long-term, and preservation of digital material has emerged as a major concern in the cultural heritage community. It is necessary that digital objects created remain accessible for as long as possible both to the intended users and the wider community. Digitization projects are costly and time consuming and the digitization process can subject original materials to potentially damaging exposure to light and excessive handling. Because of these realities, it is essential to see that the digitization process need not be repeated. Doing it right the first time, following the "scan once methodology," and properly preserving the digital items produced saves money, time and originals from additional handling and unnecessary wear and tear.

More than any other aspect of a digitization project, digital preservation, by its very nature, requires vigilance on the part of cultural heritage professionals. Digital preservation is a new and developing field. Cultural heritage professionals must monitor research being conducted in the fast-paced world of digital technology. In addition, the issues surrounding the preservation of digital objects are of a more immediate imperative than traditional materials. This is due to the lack of stability in digital storage media and the necessary equipment required to interpret and to access digital materials. This chapter seeks to equip you with the ability to do both. It introduces the challenges of digital preservation and provides information and recommendations to guide the decision making for digital preservation strategies.

Digital Preservation Challenges

At its most fundamental level, digital material is comprised of ones and zeros. Software programs are written to create and interpret this binary structure into a variety of forms that we use such as images, text documents, databases, video, sound files, etc. Without an appropriate machine, even the most simplistic of these are impossible to decipher. Consider the variety of computer hardware and software available readily and how fast popular computers and programs become obsolete.

Hardware obsolescence

Hardware obsolescence refers to the maintenance of the appropriate equipment to read the digital storage media selected. For instance, just five years ago, a 3 1/2-inch floppy disk was used for virtually everything digital anyone wanted to save or transport. Yet in today's computers, 3 1/2-inch floppy disk drives need to be specially ordered in order to be installed on a standard PC. Even if your storage media remains unharmed over the ravages of time, eventually there will be no computer extant that has the appropriate mechanism to read that medium. Without the appropriate hardware, files contained on that floppy disk are lost – irretrievable.

In order to prevent this kind of preservation problem, institutions and companies are maintaining working machines that will read the variety of medium that have been supplanted with current or advanced technologies. This kind of access, though, can be expensive. Maintaining the hardware requires space and expertise and outsourcing obsolete storage media for access requires funds that may not be readily available.

Software obsolescence

Closely connected to the issues of hardware obsolescence is the development of software applications. In interacting with those ones and zeros, software programs use their own systems for creation and interpretation. Often times, a file created in one software program cannot be interpreted by another because the program does not know how to read this particular configuration of ones and zeros. Updates to newer versions of software or changes from one to another can create obstacles to access for those digital objects.

Fragility

Both hardware and software obsolescence underscore the fragility of digital objects. Their very nature requires machines that are supplanted as technology advances and they require programs that are updated as new functionality becomes available. However, there is another kind of fragility that needs to be considered as well, for fragility is no just an operational matter. As with our more traditional objects, digital media have their own specific environmental concerns, and like the material it stores, degradation to the digital media can be more subtle and is often not identifiable to the naked eye.

Keeping Current

Digital preservation is a fast-growing area in research for information science and technologies. The Commission on Preservation and Access (CPA), the Digital Preservation Consortium, CLIR (Council on Library and Information Resources), and the Digital Library Federation (DLF) are working to develop methodological frameworks, and to ensure continuing research and development in the little understood areas of digital preservation.

A series of publications and additional initiatives can be found on the preservation pages of DLF, available at: <http://www.diglib.org/preserve.htm>.

Cornell University has created a thorough tutorial on digital preservation management available at: http://www.library.cornell.edu/iris/tutorial/dpm/eng_index.html.

Additional resources can be found in the Further Reading section of this chapter and the Resources section at the end of these Guidelines.

Creating a Digital Preservation Strategy

How can an institution keep up with all the various types of digital files, programs, and computers being used in-house much less keep abreast of emerging technologies? How much preservation is enough? How much is too much? What are the deciding factors for your institution? These are all issues that need to be considered when devising a digital preservation strategy. Here are four places to start:

- 1) **Software/hardware migration.** Because of the issues of obsolescence, all digital objects must be migrated at some point, at the very least to a file format that the latest technology can recognize. If you have chosen to preserve the whole system, then operating systems and functional software must be migrated as well. Full system migration must be carried out frequently to ensure access and usability. In order to protect your digital assets, you want to formulate a migration policy that is implemented on a regular basis rather than a reactionary action to new software or hardware. After migration, it is crucial to test your documents to ensure that functionality has been preserved.
- 2) **Physical deterioration of digital media.** As with other formats, all digital media deteriorates over time. This process will be more rapid if storage conditions are bad, such as a damp basement, or as a pile of CDs stacked one on top of another. Correct storage (e.g. in racks that enable the disks to be stored separately) and an environmentally controlled location will help to optimize stability and protect digital information from loss. Digital media should be checked and refreshed regularly to ensure that the data is still readable, and this process should be part of your preservation policy. Preserve your data on a medium where the hardware exists to transfer to a later medium if the original becomes obsolete. Remember that it is costly to use a data recovery agent to move files from an obsolete medium, so make sure your preservation policy will prevent this from happening, and migrate while the process is still straightforward. Digital media should also be part of an institution's disaster preparedness plan. See **Project Management** for more details on disaster plans.
- 3) **Metadata.** Information about the creation and maintenance of your digital objects is crucial to their preservation. The NC ECHO Preservation Metadata for Digital Objects (<http://www.ncecho.org/presmet/index.htm>) provides the basic elements that should be recorded to inform you on the properties of your digital objects. Many collection management or digital asset management systems have incorporated this into their structure, but other digitization projects will be created outside of these systems. It is important to assure that this metadata is recorded somewhere.
- 4) **User needs and preferences.** This is a complex issue that may cause certain formats to become effectively obsolete even while they remain technically functional. User

acceptance—and its decline—will be one of the key “trigger events” that will compel migrations to new delivery versions of digital collections.

The Digital Preservation Process

An institution can easily become overwhelmed by the avalanche of issues that impact the process of planning for digital preservation. This section seeks to address several issues at the core of digital preservation including digital storage during the digitization process, migration of digital material, storage media for the short term and long term, and trends for the future. It includes guidelines on what is currently considered minimum recommended practice and best practice for digital preservation with the understanding that these standards are fluid and require revisiting often. The very best result that cultural institutions can hope to achieve for the long-term sustainability of digital material will be accomplished through good digital preservation planning and vigilant management. There are essentially five main storage applications that occur during the digitization process: production, data transportation, presentation to the public, backup or archiving, and migration.

Production

The production or creation of digital material generally requires sufficient hard disk capacity to store working files while they are being manipulated and developed. If the collection is considerable and there is a large production environment, a Redundant Array of Inexpensive Disks (RAID) may be the most appropriate however storage for active files can generally be handled by a large hard drive. Be warned that determining what will reside on your hard drive and what will be forwarded to a server in networked environments can often be difficult because multiple versions of files can become confusing. It is important to outline the various processes you need to perform on the same images and then determine how many “active” files you need at any one time. This demonstrates the role of file management and file naming in the preservation process.

Data Transportation

Generally, moving digital information is handled by portable storage devices such as recordable CDs (compact disks) and more recently by DVDs (Digital Video Disks or Digital Versatile Disks). The capacity of the CD and DVD is greater than that of the tape drive, an early favorite for data transportation, though its transfer rate is slower than tape. Another feature of the CD is its compatibility across platforms. The CD-R (a CD which can only be written upon once) is a secure format; its “write once mechanism” does not allow overwriting. CD-RW (CD Read/Write) is less secure but more versatile.

Presentation to the Public

Most institutions making images of their collections available to the public via the Internet make use of in-house servers or rent space on commercial servers.

Backup/Archiving

Digital collections should be backed up in a format that is easily accessible and stored remote from the original source on a routine basis. When evaluating storage for backup, the inevitable dilemma is between speed and cost. Most managers prefer tape for backup, as it may be used at non-peak hours, when speed is not an issue. For small networked systems, tape backup is the common practice. Remember that no digital media is considered permanent.

Migration

In the context of digital preservation, migration refers to the shifting of digital objects from old media formats and software programs to newer ones. Migration of backed-up digital material needs to be as easy and cost-effective as possible for institutions to buy into a system. The continual drain on fiscal resources to repeatedly upgrade equipment and software can be borne by some institutions, but others will find it difficult to stay abreast of continual migration. Decisions must be made in every institution concerning what information will be saved and migrated and what will not based on a combination of cost effectiveness, intellectual necessity, and moral and professional obligation. When making these tough decisions, refer to the four preservation strategy issues discussed above.

Types of Digital Storage Media

Longevity of a digital medium depends on many factors – the type of media (CD, DVD, tape, etc.), how often and the way in which the media is handled, and how the media is stored. It is important to keep in mind that even with proper maintenance and great luck, no digital format is permanent or archival. The very best result that cultural institutions can hope to accomplish is long-term sustainability of digital material through good preservation planning and vigilant management. The storage media is an essential part of that process.

There are two types of digital storage media - portable and non-portable. Each has advantages and disadvantages for long-term storage.

Portable Media

CD (CD-R and CD-RW)

CD-R or Compact Disk Recordable is a format that requires a CD-ROM drive to read and to write. The CD-R format is an inexpensive way to store digital object masters, which typically require many megabytes of storage. Currently the CD-R disk will store 650 MB, though approximately 100 MB of those bytes should remain free to allow for manipulation of the data. These disks, unlike the common music CDs, are more susceptible to scratches, to fingerprints and to extremes in temperature and light. They should be handled and stored with great care. They are also susceptible to other destructive agents. If writing on the disk, only a water-based felt-tip pen should be used. An alcohol-based felt-tip pen can migrate through the protective layer and possibly affect the integrity of the data.

CD-R conforms to the ISO standard 9660, which is an established standard that allows a file system to be used under a variety of operating systems. The standard

applies only to the data track of a CD-ROM and not to audio tracks or any other media, such as erasable-optical drives. Thus CD-Rs may be read by any of a variety of operating systems such as UNIX and MS-DOS.

CD-RW or Compact Disk-Rewritable format allows the re-writing of information on the disk, requiring a special CD-RW drive. These disks will not run on a standard CD-R drive unless they are specially formatted to run on that drive. Unless an institution requires the convenience of re-write, the CD-R format is the better choice. Currently many prefer the CD-R (Compact Disk-Recordable) format for archival storage, though there is debate regarding its archival quality. To store data on a CD-R generally requires that the data be gathered on a hard drive and then "written" to the CD-R.

DVD – (DVD-R and DVD-RW)

DVD technology (Digital Video Disk or Digital Versatile Disk) is a recent addition to the growing optical disk technology market. DVD is backwardly compatible, so it may be used to read CD disks. But CD-R and CD-RW drives cannot read a DVD disk. A DVD-ROM drive will be needed to read DVD-R disk, but some DVD-R disks do not play on some machines. DVD-ROM is different than DVD-VIDEO. The former handles data, while the latter is reserved largely for the commercial video market. A DVD-R disk will hold approximately 4.7 gigabytes. The enormous storage available on a DVD makes it appealing as a storage device; gold standard DVDs have been developed to meet archival standards.

Improving the lifespan of CD's and DVD's

Always	Avoid	Never
Store media in controlled archival environment	Damage to the upper and lower surfaces and edges of the disk	Attach of fix anything to the surface of a disk
Store media in jewel case or protective sleeve when not in use	Scratching and contact with surfaces that might result in grease deposits (e.g. human hands)	Write on any part of the disk other than the plastic area of the spindle
If using sleeves, use those that are of low-lint and acid-free archival quality	Exposing disks to direct sunlight	
Wear gloves when handling the master disks		

DAT Tape, DLT Tape, ZIP® and JAZ® drives

Tape, ZIP® and JAZ® drives are all magnetic media, and magnetic media is NOT recommended for long-term storage. Tape is, however, an excellent intermediate medium, particularly for transport of data and for backup.

Improving the lifespan of DLTs

Always	Avoid	Never
Keep tape in its protective case when not in use	Placing tape near magnetic fields	Stack tapes horizontally
Move tape in its case	Moving the tapes about	Put adhesive labels on the cartridge
Store the tapes in appropriate archival environment	Exposing disks to direct sunlight	Touch the surface of the tape
Store tape vertically		Put a tape that has been dropped in a drive without first visually inspecting it to make certain that the tape has not been dislodged or moved

The above charts, modified from tables in the *NINCH Guide to Good Practice* (available at, <http://www.nyu.edu/its/humanities/ninchguide/XIV/>), can assist in making sure digital storage media lasts as long as possible.

Non-portable media

Network Servers (drives)

The minimum storage space recommended for network servers changes every three to six months. Suffice it to say that if a server is required, it should be purchased to be adequate for the first two years of the project. Depending upon the size of their digital holdings, larger institutions may need to upgrade on an annual basis, especially if production levels of digital materials are high.

Hard Drives (PC disk drives)

It is recommended that institutions purchase the largest hard drive they can afford. If it is possible to purchase two hard drives, this will provide a more flexible storage system. If managers of digital projects use hard drives for image storage, they should defragment them on a regular basis to maintain optimum performance. Hard drives are not recommended for long-term storage.

The following recommendations are based on available hardware in the medium price range:

- Minimum Processor: Pentium II (300-450 MHz)*
- Recommended Processor: Pentium IV (up to 512 Mhz)
- Minimum configuration: 40-80 GB for one hard drive and expansion slot for additional hard drive.
- Recommended configuration: two hard drives, 40-80 GB each, and/or a shared storage option.

Other Digital Storage Concerns

The amount of storage required depends on a number of inter-related issues including but not limited to the size of your digital holdings, your institution's budget, and your institution's digital preservation strategy. There are a number of storage issues other than cost, amount, and media permanence that cultural repositories should factor into decisions regarding the "long term" storage of digital materials. These include labeling, file management, and metadata issues.

As soon as digitization projects get going, the number of images piles up. As noted above, CDs and DVDs hold immense amounts of data, and it is unlikely that contents lists on the jewel cases will be palatable in the long-run, as more and more storage devices are used. Therefore, in addition to the labeling that would occur on the outside of a CD storage case, managers of digital projects should maintain "preservation metadata" for each image. The information necessary is explained in the NC ECHO Preservation Metadata for Digital Objects (<http://www.ncecho.org/presmet/index.htm>).

It is also recommended that the file naming conventions follow these standards:

- Attempt to conform to ISO 9660 naming standard (a standard that defines a file system usable under a variety of operating systems)
- Establish a file naming convention and the extensions later made to it
- Base names on accession numbers or unique IDs
- Avoid case sensitivity

NC ECHO recommended storage standards

- Master file storage:
 - Minimum recommendation: Gold CD-R
 - Best practice recommendation: Redundant Hard Disk storage and/or Hard Disk with Tape Backup
- CD names are simple date/time stamps (e.g., 19990412_1628)
- ISO 9660 standard is used as strictly as possible

Conclusion

Digital preservation seeks to achieve longevity of the digital object with all its original properties intact. This is a daunting task and one not easily tackled or mastered. Many questions in the field of digital preservation remain unanswered, and many more questions will emerge as technology relentlessly forges ahead with new developments. Whether your institution has only the means to preserve the minimum content of your digital creations or can afford to preserve the whole discovery and display system, policies should be put in place to ensure the long-term sustainability and accessibility of the digital content you have chosen to be preserved.

Further Reading

Benford, Gregory. *Deep Time: How Humanity Communicates Across Millennia*, New York: Avon, 1999.

Conway, Paul. "The Implications of Digital Imaging for Preservation." In *Preservation of Library and Archival Materials*, 2nd ed. Edited by Sherelyn Ogden. Andover, MA: Northeast Document Conservation Center, 1994.

Conway, Paul. Preservation in the Digital World, available at:
<http://www.clir.org/pubs/abstract/pub62.html>.

Development of a Testing Methodology to Predict Optical Disk Life Expectancy Values (Summary) <http://palimpsest.stanford.edu/byorg/nara/nistsum.html>

Digital Preservation Coalition <http://www.dpconline.org>

Digital Projects Guidelines. Attachment 11, Arizona State Library, Archives and Public Records, available at http://www.lib.az.us/digital/dg_a11.html

"Long-Term Usability of Optical Media - The National Archives and Records Administration and the Long-Term Usability of Optical Media for Federal Records: Three Critical Problem Areas" <http://palimpsest.stanford.edu/bytopic/electronic-records/electronic-storage-media/critiss.html>

Rothenberg, Jeff. "Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation," <http://www.clir.org/pubs/reports/rothenberg/contents.html> (January 1998).