North Carolina Department of Cultural Resources

State Library of North Carolina

State Archives of North Carolina

Best Practices for Digital Permanence

Version 1.0

July 2013

Contents

1 Introduction

Today, many public records are either born digital or have been digitized. Regardless of format, these records need to be retained based on the record's retention and disposition schedule. Some of these records will have enduring or historical value; therefore, their retention schedule will require that the document be preserved for longer periods of time. Electronic records have a life cycle that includes the creation, management, and use of the digital object. Preserving a digital record throughout this life cycle presents unique challenges for records creators and users. This document discusses the threats to digital materials and strategies for digital records preservation and access.

This paper will cover the following:

- Relevant North Carolina Statutes
- Discussion on the permanence of electronic records
- Threats to digital records
- Types of digital media
- How-to preserve a digital record through its life cycle

1.1 North Carolina Statutes

Some public records are required to be retained "permanently." State statutes direct public employees to indefinitely retain records with enduring and historical value. In the North Carolina, those statutes include:

- ***NCGS § 132, Public Records Act*** tasks state employees with maintaining public records, which are the property of North Carolina citizens. Public records are defined as: " all documents, papers, letters, maps, books, photographs, films, sound recordings, magnetic or other tapes, electronic data-processing records, artifacts, or other documentary material, regardless of physical form or characteristics, made or received pursuant to law or ordinance in connection with the transaction of public business by any agency of North Carolina government or its subdivisions."[1] This law includes all records created in the course of public business regardless of format; therefore, electronic records must be maintained under the conditions of this law. Statute 132 can be found here: http://www.ncga.state.nc.us/gascripts/statutes/StatutesTOC.pl?Chapter=0132

- ***NCGS § 121, Archives and History Act*** instructs the Department of Cultural Resources to help public officials manage their public records, including providing assistance on preserving electronic records such as email, databases, and website. The Department of Cultural Resources will work with the agencies to create retention and disposition schedules that regulate the destruction of public records. Public records with enduring value will be transferred to the State Archives. Statute 121 can be found here: http://www.ncleg.net/gascripts/statutes/StatutesTOC.pl?Chapter=0121

- **NCGS § 147-33.89, *Business continuity planning*** tasks state agencies to "develop and continually review and update as necessary a business and disaster recovery plan with respect to information technology."[2] This means that state agencies must assess the types of disaster that could affect their technical

[1] G.S. §132-1. Public Records Act.
[2] G.S. §147.33-89(a). Business continuity planning.

infrastructure and then take the appropriate measures to mitigate the risk to data loss during those disasters. Creating a disaster recovery plan is an important component to long-term digital preservation.

If your agency makes a commitment to keep records permanently, you must properly manage and maintain those records in their original form. North Carolina requires state agencies to create human-readable preservation duplicates of analog records that have permanent value as identified in the records and disposition schedules. As described in §132-8.2, "preservation duplicates shall be durable, accurate, complete and clear..."[3] Read more about the public records requiring human-readable preservation duplicated here: http://www.records.ncdcr.gov/guides/Humreadabledupspolicy050217.pdf

Similarly, if an electronic record is listed as having permanent value in your department's retention and disposition schedule, it is your responsibility to retain access to that record over time. Preservation of an electronic record includes ensuring authenticity of the original record and retaining the prescribed metadata. To read more about metadata as public record, click here: http://www.records.ncdcr.gov/guides/Metadata_Guidelines_%2020101108.pdf

1.2 What do we mean by Permanent?

The practical application of retaining electronic records can be more complicated than the law suggest. Records management professionals have determined the practical implications of permanent retention of paper records but the digital age has ushered in a range of new dependencies and considerations. Digital materials' longevity is dependent on a host of elements within a records manager's control such as logical file naming conventions and choosing stable file formats. But there are also elements outside of our control that need to be managed such as storage media, hardware, operating systems, and software applications. Unfortunately, many public and private organizations are playing "catch-up" as a large amount of data has already been lost to the various threats. Section *2 Threats to Permanence of Digital Materials* discusses these threats in more depth. However, it is important to note that rediscovering and recreating digital information is expensive. The Blue Ribbon Task Force on Sustainable Digital Preservation and Access highlighted the economic advantage of a preservation plan over having to recreate lost digital items. The report notes that the "benefits of preservation may be most compellingly expressed in terms of *negative benefits*—the costs incurred if data are not preserved. These costs may reflect the time and effort needed to recreate the information or, if it cannot be recreated, the kinds of uses that would then not be possible."[4] The big take-away is that maintaining digital permanence is an ever-evolving task. Once a commitment is made to preserving digital materials, it is important to stay apprised to changing technology and standards in order to ensure the longevity of record use; otherwise, it is easy to succumb to data loss.

1.3 Definitions

Born-digital records: Information created in electronic format. Examples include documents created in Microsoft Word, databases, and online content such as websites.

Checksums: An error-detection mechanism in which a transmitted digital message is accompanied by a numerical value based on the number of set bits in the message. Once that message makes it to its destination, the same formula is applied to the message to checks to

[3] G.S. §132-8.2. Public Records Act.
[4] "Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information." The Blue Ribbon Task Force on Sustainable Digital Preservation and Access, Feb. 2010. Web.

make sure the accompanying numerical value is the same. If not, the receiver can assume that the message has been altered.

Digital object: It is a single unit of digital content, such as a document, a photograph, or an audio file that is accessible through electronic format. A digital object is made up of multiple components including code that comprises word or images, and metadata that helps describes the object.[5]

Digitized records: Records that have been converted from an analog copy to a digital form, through scanning or other forms of digital reproduction.[6]

Electronic (or digital) record: An electronic record is a record that can be stored, transmitted or processed by a computer; an electronic record is maintained in a coded format and can only be accessed by using a computer that converts the codes into human-readable text, images, or sounds.[7]

Migration: The process of moving data from one information system or storage medium to another to ensure continued access to the information as the system or medium becomes obsolete or degrades over time.[8]

Metadata: Metadata is structured information that describes, explains, and/or locates an electronic file. Metadata provides answers to questions like "what is it," "where did it come from," and "who created it"?[9]

Preservation plan: Addresses an institution's overall preservation goals for electronic records and provides a framework that defines the methods that will be used to reach those goals.[10]

Records Series: A group of similar records that are related by being created, received, or used in the same activity.[11] Retention and Disposition schedules divide records into series for the purpose of determining the length of retention.

Retention and Disposition Schedule: A document that identifies and describes an organization's records, usually at the series level, and provides instructions for the disposition of records throughout their life cycle.[12]

Trustworthiness: The quality of being dependable and reliable. For electronic records, trustworthiness often implies that the system is dependable and produces consistent results based on well-established procedures.[13]

---

[5] Millar, L. (2010). *Archives: principles and practice*. New York: Neal-Schuman Publishers. 208.
[6] Ibid.
[7] Millar, L. (2010). *Archives: principles and practice*. New York: Neal-Schuman Publishers. 208.
[8] Migration, *A Glossary of Archival and Records Terminology*, http://www2.archivists.org/glossary/terms/m/migration. Accessed April 2013.
[9] Department of Cultural Resources guide "Metadata as a Public Record in North Carolina: Best Practices Guidelines for Its Retention and Disposition," November 2010 http://www.records.ncdcr.gov/erecords/default.htm.
[10] Electronic Records Management Guidelines: Digital Media. Minnesota Archives. March 2012.http://www.mnhs.org/preserve/records/electronicrecords/erdigital.html
[11] Series, *A Glossary of Archival and Records Terminology*, http://www2.archivists.org/glossary/terms/r/retention-schedule. accessed April 2013.
[12] Retention Schedules, *A Glossary of Archival and Records Terminology*, http://www2.archivists.org/glossary/terms/r/retention-schedule. accessed April 2013.
[13] Trustworthiness, *A Glossary of Archival and Records Terminology*, www2.archivists.org/glossary/terms/, accessed March 2013.

2 Threats to Permanence of Digital Materials

Electronic records face challenges that have not been issues for the preservation of paper records. These new threats will require that records' creators and managers vigilantly address these issues in order to ensure the long-term preservation of and access to electronic records. Digital materials cannot wait until they are transferred to the State Archives before the preservation process begins. Instead, records' creators have to take an active role in beginning the preservation process. Part of beginning this process is to be aware of some of the issues that plague electronic records and taking measures to evade those problems. The following section lists some of the problematic issues related to digital materials.

2.1 Application Obsolescence

Traditionally preservation meant keeping items unchanged and in their original format. Due to the constant evolution of hardware and software, digital information is becoming increasingly vulnerable to obsolescence. Obsolescence occurs when old technology is replaced by a newer version and materials created on the outdated technology are no longer accessible on the new technology. In today's competitive market, hardware and software companies come and go at a rapid rate. As such, much of the hardware and software used today may not be available in the near future, and digital public records may be unreadable by new systems. File formats and applications must also be compatible with replacement systems. If digital records are not converted to formats that are congruent to new systems, there is risk of loss by obsolescence.

When possible, use open-standard software or save files you want to keep in open-standard file formats. An open-standard or sustainable format is one that increases the likelihood of a record being accessible in the future. To assist you, the Department of Cultural Resources has published *File Formats Guildelines for Management and Long-Term Retention of Electronic Records*. Recommended formats include TIFF for photographic image files, PDF/A-1-a and XML for text files, and AVI for video files. View the document here: www.digitalpreservation.ncdcr.gov

Additionally, to prevent records from becoming incompatible with modern systems and applications, records' users can employ migration techniques. When new hardware is purchased, immediately transfer data from the old hardware. Also, if new software is used, move digital records into the new programs and applications so that the records do not remain in the superseded structures. Migrating large amounts of records can be difficult; however, the staff from the Digital Services Section of the State Archives and the Digital Information Management Program of the State Library are available to provide assistance. Click here for more information: http://www.archives.ncdcr.gov/

Migrating digital records is important to retain content, but it is also beneficial to maintain the original version in its native format. By maintaining the integrity of the source data, one lessens the chance of losing the data altogether and increases the chances that migration will produce a successful copy. Researchers have argued that it is easier to recover data from its original source than from its copies, especially after several migrations.

2.2 Corruption

Paper materials decay over time, but professional preservation techniques reverse some of the corrosion and slow future processes of degradation. Digital records do not decompose

like paper and other analog records, but they do face a unique set of preservation issues. Digital materials are made up of 'bits' – a series of digits that stand for the material's information or content. These strings of digits are read by a device such as a computer and displayed or communicated for the human eye. However if one of these bits is corrupted, the entire record becomes corrupted. One alteration in bits can result in the record's effective death.

Corruption also relates to the vulnerability of digital materials. Ensuring trustworthiness in digital records requires consideration of network access, encryption settings, and system protections. It is recommended that technology professionals and public employees run regular virus scans and employ appropriate password protections to shield permanent records from potentially harmful access. Even a seemingly small alteration such as changing the last access date of a record can alter its reliability and usefulness as a permanent record. Minimizing substantive modifications of the records after production or migration is also recommended to maintain integrity.

2.3 Completeness

Another important consideration for the preservation of digital materials is the retention of accompanying metadata or other contextual information. Metadata is structured information that describes, explains, and/or locates an electronic file. Examples include the file name, creation date, and modification history. Metadata can help fill in the gaps so that records are more comprehensible and relevant to the user. When preserving electronic files, it is recommended that metadata be kept in a manifest or spreadsheet so that future users understand where a file is located and its purpose. For more information on metadata see Metadata in North Carolina: Best Practices Guidelines for Its Retention and Disposition. http://digitalpreservation.ncdcr.gov/policies.html

2.4 Findability

There is a proliferation of digital materials in the workplace. Proper management of electronic records has never been more important. Digital files can become essentially lost and useless if one cannot locate them or if multiple copies are on different drives outside of the architecture developed by IT professionals. Active management practices of digital files ensure that they can be found for use in your office or in the case of a public records request. Employees need to be able to differentiate between the many digital records in order to carry out a productive preservation plan for permanent records. One of the simplest steps public employees can take to improve the management of their digital records is to employ conscientious file-naming customs. Consistent file-naming helps employees organize and locate records. Please refer to the following best practices document for tips regarding file-naming: http://digitalpreservation.ncdcr.gov/policies.html

Also, view these short informational videos on file-naming:
http://digitalpreservation.ncdcr.gov/tutorials.html

2.5 Mutability of Electronic Records

Electronic records can easily be changed and updated. However, this convenience also presents challenges for record-keepers. An electronic record can be edited with little indication of if and when those changes were made. To create an authentic record, it is important that the electronic record be preserved in a stable manner so that changes are not made either purposely or accidently. Again, practicing descriptive and consistent file-naming is important to indicating major changes to a record. For each new version, the file can be saved with the

version number in the file name. Additionally, descriptive file naming provides context to the content of the electronic records. This context will give clear indication when major changes to the record have been made.

3 Strategies for Digital Preservation

3.1 Types of Digital Media

Electronic records are saved on three main types of media—magnetic, optical, and solid state.

Magnetic: Digital information is encoded as microscopic magnetized needles on the surface of the magnetic medium being used.[14] Two commonly used magnetic mediums are (1) disks and (2) tape.

(1) Magnetic disks are the most common type of permanent data storage. Disks include a computer's internal hard drive, which saves your computer programs and documents; and external hard drives that connect to your computer through a port and provide additional storage options. The advantage to magnetic disk storage is that it relatively inexpensive and it allows for fast access to data. The disadvantage to magnetic disks is that it can be affected by environmental factors including magnetic fields and dust. Over time, hard disks can fail, which can lead to data loss.

(2) Magnetic tape can be reel-to-reel or cartridge format. It is used primarily for data storage of large amounts of information because it is relatively inexpensive. It is slightly more cumbersome since it provides sequential access, meaning you have to go through all preceding data before finding the information you may need. Sequential access differs from random access which allows the user to access the data at an arbitrary period of time. Magnetic tape has a life span of about 15 to 30 years.[15]

Optical: Optical drives use focused lasers to create microscopic holes on the surface of the medium. These holes represent coded data. The lasers are used to both write and read data.[16] CDs, DVDs, and Blue-Ray disks are all examples of optical disks. Advantages to optical disks are that they are portable, fairly inexpensive, and durable. Some disks, such as CD-R and DVD-R, do not allow data to be overwritten. Disadvantages to optical disks include the limited amount of storage space and they can be expensive compared to other types of media. From a preservation standpoint, optical disks present several challenges. First, they require other drives to read and write on the disks, and therefore, there can be compatibility issues, especially at the rate technology changes. Second, the life expectancy of a CD/DVD is about 5 years, which does not make it a solution for long-term preservation of digital materials.

For recommendations related to physical protection of DVDs and CDs, please refer to the following document: www.ncdcr.gov/archives

[14] Electronic Records Management Guidelines: Digital Media. Minnesota Archives. March 2012.http://www.mnhs.org/preserve/records/electronicrecords/erdigital.html
[15] Ibid.
[16] White, R. (2008). How computers work (9th ed.). Indianapolis, IN: Que Pub.

Solid State Storage Device (SSD): Solid state media uses flash memory for data storage.[17] Flash memory can be erased and reprogrammed; however, there are limitations on the number of times this can occur before the device will begin to fail. Examples of SSDs include flash memory cards, USB flash drives, and solid state hard drives. These devices connect to a computer through a card reader or USB port to exchange the saved data. Advantages to SSDs are that they are durable, they have a longer life expectancy than most optical drives, and they retrieve data quickly. One disadvantage of SSDs is that they are expensive; however, with technological advances, SSD are getting cheaper and have increased their data storage capacity. From a preservation standpoint, one of the disadvantages of SSDs is that they are vulnerable to magnetic fields and electric or static charges leading to potential data loss. Additionally, media devices that are portable are more likely to come into contact with other environmental factors that will increase risks for data loss. Therefore, they are not recommended for long-term preservation.

Steps to maintaining digital media

Studies have shown that under optimal conditions the life expectancy of magnetic media ranges from 10 to 20 years and up to 30 years for optical media; however, be cautious of vendors that claim longer life expectancy rates than industry standards.[18] Additionally, wear and tear of these materials can lower their life expectancy. Recognizing the type of data storage you are using helps in the long-term planning of your electronic records. You may be required to migrate your data from one device to another to ensure that your records remain usable and authentic. Additionally your office can take the following precautions:

- Use high quality storage media and batch test newly purchased storage devices to ensure they are not defective.
- Prohibit eating and drinking in areas where storage devices are being held.
- Store these devices in a cool, dry place and keep the area free from dust and other environmental contaminants.

For additional information on types of digital storage media, see the GeoMAPP Storage Primer located here: http://www.geomapp.com/docs/GeoMAPP_Storage_Primer_final_20111231.pdf

3.2 Cloud Computing and Storage

The National Institute of Standards and Technology (NIST) defines cloud computing as a "model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."[19] Cloud computing allows you to retrieve, use, and store records regardless of computing device or location. The service provider maintains the equipment used to create and store the data. Cloud computing can be used as another source to store a copy of your data. However, there are several challenges to using cloud storage including managing all additional copies and syncing them so that the copy saved in cloud matches the copy stored locally. As long as you have a local copy, you can use cloud storage as part of your data management strategy. Section 3.3 discusses in more detail the benefits and challenges of saving multiple

[17] Electronic Records Management Guidelines: Digital Media. Minnesota Archives. March 2012.http://www.mnhs.org/preserve/records/electronicrecords/erdigital.html
[18] Electronic Records Management Guidelines: Digital Media. Minnesota Archives. March 2012.<http://www.mnhs.org/preserve/records/electronicrecords/erdigital.html>
[19] Peter Mell and Timothy Grance, The NIST Definition of Cloud Computing (Draft), NIST, January 2011. http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf

copies. Additionally, cloud vendors do not guarantee data integrity; therefore, it is recommended that there are checksums in place prior to moving files out of the cloud.[20]

Different cloud vendors will provide varying degrees of long-term preservation functionality. It is important to consider what you need from a cloud service in order to have it fit into your preservation plan before choosing a vendor. Some providers will have more robust options for managing and storing metadata, searching content, and verifying checksums. Once a vendor is chosen, the contract should be written in a way that is consistent with your preservation plan. Additionally, this contract should include clear expectations if data is to be moved from one provider to another. Some cloud vendors make it difficult and costly to move data from one cloud to another.[21]

Currently, State agencies must adhere to Session Laws of North Carolina, SL 2011-39. §11(c), which mandates that "State agencies developing and implementing information technology projects/applications shall use the State infrastructure to host their projects."[22] However, an exception to this requirement may be granted if approved by either the State Chief Information Officer on the basis of technology requirements or by the Office of State Budget and Management based on cost savings. If your office decides to maintain records in the cloud, it is important to review the possible implications to your records management strategy. For more information, review the Department of Cultural Resources' "Best Practices for Cloud Computing Records Management Considerations Guide:" www.ncdcr.gov/archives

3.3 Preservation of a Digital Record through its Life Cycle

Much like a paper record, digital records have a life cycle. The life cycle of a digital record includes its creation, management, and use and re-use. It is important to take an active role in preserving the document at each stage of the record's life cycle.

Creation:
Preservation of an electronic record begins with its creation. Therefore, the record creator is integral to the long-term longevity of the data she produces. The creator should add relevant metadata; save the record in a recommended file format; and give the record a descriptive file name.[23]

As noted in other places in this document, there are many resources available to assist creators in taking steps to preserve their electronic records.

- Metadata as a Public Record in North Carolina: Best Practices Guidelines for Its Retention and Disposition: http://digitalpreservation.ncdcr.gov/policies.html
- File Format Guidelines for Management and Long-Term Retention of Electronic Records http://digitalpreservation.ncdcr.gov/policies.html
- Best Practices for File-Naming: http://digitalpreservation.ncdcr.gov/policies.html

[20]. *Report on Digital Preservation and Cloud Services (Public).* Minnesota Historical Society and Instrumental, Inc March 3013. < http://www.mnhs.org/preserve/records/docs_pdfs/Instrumental_MHSReportFinal_Public_v2.pdf>
[21] Ibid.
[22] Session Laws of North Carolina 2011-391 (HB 22).
[23] North Carolina Department of Cultural Resources. Digital Preservation Best Practices and Guidelines: I create Files. Digital Preservation Education for NC State Government Employees. http://digitalpreservation.ncdcr.gov/

Management:
Electronic record creators may become managers of those records or another employee may become responsible for maintaining records created by others in her office. Effective electronic records management includes understanding the scope of the materials you have to manage; running regular virus checks to ensure the digital records are saved in a safe environment; storing more than one copy in multiple locations; and ensuring that the file formats are still readable on current technology.[24] If an electronic document is about to become obsolete, it is the responsibility of the manager to employ a preservation strategy to migrate or emulate the record.

LOCKSS
As part of an electronic records management strategy, it is recommended that more than one copy of the record be stored in multiple locations. This model is known as "Lots of Copies, Keep Stuff Safe (LOCKSS)."[25] Preserving electronic records in a distributed manner helps reduce potential technical threats faced by digital materials and ensures longevity. However, when employing LOCKSS, it is important to have a workflow in place to keep track of the various copies especially if they are stored on multiple drives or in cloud storage.

Use and Re-Use:
Users can assist records managers in preserving digital materials by providing feedback if there is trouble finding or accessing an electronic record. The symbiotic relationship between the user and record manager will help ensure that digital materials are cared for in a manner that promotes a long life span of the record.

For more information about digital preservation during a record's life cycle, visit the Digital Preservation Education for NC State Government Employee webpage on Digital Preservation Best Practice and Guidelines, located here: http://digitalpreservation.ncdcr.gov/

Additionally, you can print the "State employee checklist for digital preservation" one-sheet to assist with the daily management of electronic records.
http://digitalpreservation.ncdcr.gov/checklist_dig_pres.pdf

3.4 Electronic Records as Public Records

Public records need to be regularly backed up, especially those records with permanent retention. Files can be backed up at a remote location or on a network drive. If digital records are copied onto an external format such as a CD or microform, multiple copies should be made and stored in different locations. Having multiple copies reduces the risk of losing the primary content of the records. For most permanent records, preservation duplicates are required and must be stored in an off-site location. These preservation duplicates must be in a human-readable format – paper hard copy or microfilm. The State Archives stores preservation duplicates for North Carolina agencies and local governments. For more information on human-readable preservation duplicates, please see: www.ncdcr.gov/archives

3.5 Disaster Preparedness

Another essential piece to the planning process is disaster preparedness and response. Since agencies' permanent records are considered essential, disaster protections should be

[24] Ibid.
[25] Stanford University. *Lots of copies, keep stuff safe*. <http://www.lockss.org/>

among the first established. Consult an IT professional to ensure that all electronic records are being backed-up regularly. Additionally, each agency should have policies in place to outline specifications for data backups including how often backup files are made and the length those backup files are kept. However, IT backups are designed to aid in a recovery situation, not ensure preservation or permanence. Disaster preparedness is just one of the initial steps to long-term preservation of electronic records.

4 Conclusion

The key to compliant and responsible record-keeping is planning. Digital preservation is based on risk and access management—guaranteeing future usability of and accessibility to digital content. This process warrants attention to the issues discussed in this document, among others. Agencies' unique concerns should be worked through in the preservation planning process according to priority.

5 Additional Resources

North Carolina Guidelines for Managing Trustworthy Public Records Produced, Version 2.0
http://www.ncdcr.gov/Portals/26/PDF/guidelines/guidelines_for_digital_public_records.pdf

North Carolina Department of Cultural Resources "Digital Preservation" webpage
http://digitalpreservation.ncdcr.gov/index.html

North Carolina Department of Cultural Resources "Best Practices for Management and Preservation of Digital Media" document www.ncdcr.gov/archives

Library of Congress "Digital Preservation" webpage
http://www.digitalpreservation.gov/index.html

OCLC Digital Archive Preservation Policy and Supporting Documentation
http://www.oclc.org/support/documentation/digitalarchive/preservationpolicy.pdf

Electronic Records Management Initiative (ERM)
http://www.archives.gov/records-mgmt/initiatives/erm-overview.html

GeoMapp Storage Primer:
http://www.geomapp.com/docs/GeoMAPP_Storage_Primer_final_20111231.pdf