

# Long-Term Preservation

## Summary

During the course of routine business, your agency generates thousands upon thousands of electronic records, from e-mail to web pages to complex e-government transactions. Most are useful for only a short period of time, but some you may need to keep permanently. For those records, you will need to implement a well-considered, well-documented plan for their preservation in order to ensure that they remain trustworthy and useful over time. Tools such as migration, conversion, metadata, and eXtensible Markup Language (XML) will help you not only preserve your records, but also realize their full value.

## Legal Framework

For more information on the legal framework to consider when developing a preservation plan for your records, refer to the *Introduction* and Appendix D of the *Trustworthy Information Systems Handbook*. Also review the requirements of:

- Official Records Act [Minnesota Statutes, Chapter 15.17] (available at: <http://www.revisor.leg.state.mn.us/stats/15/17.html>), which mandates that government agencies must keep records to maintain their accountability and specifies that the medium must enable the records to be permanent. The Official Records Act further stipulates that you can copy a record and that the copy, if trustworthy, will be legally admissible in court.
- Records Management Act [Minnesota Statutes, Chapter 138.17] (available at: <http://www.revisor.leg.state.mn.us/stats/138/17.html>), which establishes the Records Disposition Panel to oversee the orderly disposition of records using approved records retention schedules. Coordinate your records retention schedules with your preservation plan to help ensure that you store and dispose of records in accordance with the Records Management Act.
- Minnesota Government Data Practices Act (MGDPA) [Minnesota Statutes, Chapter 13] (available at: <http://www.revisor.leg.state.mn.us/stats/13/>), which mandates that government records should be accessible to the public unless categorized as not-public by the state legislature. You must be able to provide access to the stored public records, yet prevent unauthorized access to not-public records.
- Uniform Electronic Transactions Act (UETA) [Minnesota Statutes, Chapter 325L] (available at: <http://www.revisor.leg.state.mn.us/stats/325L>) and Electronic Signatures in Global and National Commerce (E-Sign), a federal law (available at: <http://thomas.loc.gov/cgi-bin/query/z?c106:S.761:>). Both UETA and E-Sign address the issues of legal admissibility of electronic records created in a trustworthy manner and the application of a paper-oriented legal system to electronic records.
- Information and Communications Technology Policy [Minnesota Statutes, Chapter 16E.04] (available at: <http://www.revisor.leg.state.mn.us/stats/16E>), which mandates state agency

compliance with Minnesota's enterprise technical architecture "to ensure that individual agency information systems complement and do not needlessly duplicate or conflict with the systems of other agencies. . . . [and to] promote the most efficient and cost-effective method of producing and storing data for or sharing data between those agencies." Section 16E.07 establishes the North Star portal as the state's official online government information service with the idea that "the greatest possible access to certain government information and data is essential to allow citizens to participate fully in a democratic system of government."

## Key Concepts

The value of your information justifies your investment in information technology. There is no point to an agency investing large sums in hardware and software if it cannot preserve the use-value of the information it creates, exchanges, and stores. In the short term, this is often not a problem. But, over time, it will be. As technology changes, hardware and software will become obsolete, and then you might face some hard choices. The challenge is to preserve the usefulness and trustworthiness of your information in an efficient and cost-effective way.

Any preservation plan for electronic records must take into account the changes in hardware and software, the limitations of storage media, and the potential use-value of your information. As you begin exploring your options, you will need to be familiar with the following:

- Needs analysis
- Physical storage options
- File format options
- Digital preservation techniques
- E-government and collaboration

## Needs Analysis

As a first step in developing your preservation plan, you should do a needs analysis to help guide your decisions. While the complexity of such an analysis will vary from situation to situation, these basic components should always be included.

First, you need to understand the value of your information. The value of your information will justify your investment in technology, over the short- and the long-term. Minnesota's enterprise technical architecture notes that information is the state's most important asset. But all information is not created equal; some has much more value than others. Some of your information, as records, will have legal and evidentiary significance and may well demand special attention. Most of the information you want to preserve will be important to your agency's mission or, increasingly, to the business of other agencies as well. As e-government develops in complexity and sophistication, more and more agencies will be expected to work

within the framework of a common technological architecture and to share the information they create.

The practical side of understanding the value of your records is determining their retention requirements. How long do you really need to keep them? Why are you keeping them? Do they have to be kept in electronic format or is there another, more cost-effective option for long-term storage? For instance, a word processing document might be printed and kept as a paper record without losing any of its value. In contrast, printing a web page means a significant loss of information and functionality.

It is also important to ascertain if access to certain data in your records is restricted by statute. The state's Government Data Practices Act and some federal laws, such as the Health Insurance Portability and Accountability Act (HIPAA), will determine if data needs to be protected as confidential or non-public. If it does, then you will need to ensure that your long-term storage and access policies account for those obligations.

In the broadest sense, the demands governing the access and use of your records will determine what preservation options are most appropriate and will dictate the metadata you should create and store along with the records. Metadata is the "data about the data," that allows you to manage, find, and evaluate your information over time. Minnesota's enterprise technical architecture includes metadata standards for GIS data, web content management, and recordkeeping. There are a number of international standards that are pertinent as well. While all-important for the long-term preservation of data, metadata takes on additional significance when you share your information because others must understand the information's structure and content in order to put it to fullest use. For more information about metadata, refer to the *Metadata* guideline in this series.

## Physical Storage Options

As mentioned, choosing the most appropriate storage option for your situation will depend upon your records' access requirements. There are basically three options available to you:

- *Online storage.* Records are kept on a server or hard drive and are immediately available for use over a network. This option is best for records that must be accessed frequently.
- *Nearline storage.* Records are stored on media such as optical disks in jukeboxes or tapes in automated libraries which are attached to a network. Because retrieval is slower than with online storage, this option is most appropriate for records that are accessed occasionally.
- *Offline storage.* Records are stored on removable media and must be manually retrieved. This option provides the slowest access and should be used for records that are only rarely needed.

If you choose nearline or offline storage, you will need to consider what media will best suit your needs. To do this, you should start by analyzing your current and projected volume of stored records, along with the size of the files themselves and any associated metadata. Also take into account any security requirements, such as viewing, use, and modification restrictions.

Different media have different storage characteristics. For instance, with CD-R, DVD-R, and DVD+R disks, data can be recorded one time only, after which the media becomes read-only. This provides protection for your records against intentional or unintentional tampering. CD-RW, DVD-RW, DVD+RW, and DVD-RAM can all be written to multiple times, although at the cost of diminished life span and the risk of possible loss or alteration of the contents. CDs can generally store up to 800 MB of data, while DVDs can store several gigabytes. Magnetic tape is an alternative to optical disk, with capacity up to several gigabytes, although retrieval is much slower since it is a sequential-access rather than random-access medium. Tapes are most often used for offline storage and backups.

Life spans also differ among media types. Under optimal storage and use conditions, optical disks and magnetic tapes will generally give reliable service for anywhere between 5 and 20 years. Under normal conditions, however, life expectancies are probably significantly shorter. For more information on media types, refer to the *Digital Media* guideline in this series.

## File Format Options

Most records are created using specific, proprietary software applications. Over time, these applications will be upgraded or be phased out altogether. Because upgraded applications may or may not be able to read files created with previous versions, backward compatibility is not a given and cannot be counted on as a preservation tool. Maintaining the software on your own is an option, but over and above the question of costs, that carries the risk the software will fail in time, leaving you with no way to access your records. One common alternative is continually to convert your files from version to version and format to format as your software environment changes.

While non-proprietary formats are the ideal for the long-term preservation of files, they are few in number and each has its limitations. ASCII or plain text will capture data in the lowest common denominator of formats, losing structure and functions in the process. Rich Text Format (RTF) is a Microsoft format, although it is supported by a variety of vendors and software applications. Portable Document Format (PDF), a popular choice for file sharing and storage, is an Adobe product. Because Adobe makes PDF's specifications publicly available, many believe that it is an open standard when, in fact, the company is under no obligation to continue this practice into the future. Furthermore, PDF has a problem with backward compatibility, with newer versions often incorrectly rendering files created with older ones. To address these problems, an archival version, currently referred to as PDF/A, has been developed and is under consideration as a potential ISO standard.

For long-term preservation and use, eXtensible Markup Language (XML) is currently the optimum choice of formats. An international standard since 1998, XML is both a file format and a text-based, self-describing, human-readable markup language that is independent of hardware and operating systems. Because it is infrastructure-independent, XML is one of the best solutions for re-purposing the content of your records and/or sharing them with others. Proper use of XML requires a certain amount of planning and up-front commitment of money and time, but its structured nature makes it suitable for automation and will allow you to more easily take advantage of whatever new open formats will follow in the future. For more information on file formats, refer to the *File Formats* guideline in this series.

## Digital Preservation Techniques

There are several approaches, some more practical than others, to ensure that electronic records remain useful over time. One is to save all of the hardware, software, and documentation needed to support the records. Known as the “computer museum” approach, it is not very realistic on a large scale because, given how rapidly hardware and software environments change, it means storing and maintaining huge quantities of outdated equipment with no assurance that any of it will work when needed.

Emulation has a similarly antiquarian flavor. Emulator programs simulate the behavior, look, and feel of other programs, thus preserving the functionality of the records in their original format without the necessity of saving the original equipment and software. However, emulation has so far proven more attractive in theory than in practice. There are few examples of success using this approach, and costs have proven high. It has a further limitation in that, at best, emulation simply reproduces earlier, less sophisticated versions of an application. Given all the expenses of technology, it seems problematic to limit the value of information by preserving it in a static framework.

Encapsulation is a third approach to preservation. It involves combining the object to be preserved with all of the necessary details of how to interpret it within a wrapper or package, all possibly formatted in XML. While appealing in its comprehensiveness, encapsulation has several drawbacks: file sizes are large because of all of the included information; format specifications must be determined; the encapsulated records must somehow be generated, usually separate from the act of record creation; and the encapsulated records must still be migrated over time.

The most common approach to preserving electronic records involves a combination of two other techniques: migration and conversion. Migration is the process of moving files to new media (also known as “refreshing”) or computer platforms in order to maintain their value. Conversion entails changing files from one format from one to another and may involve moving from a proprietary format, such as Microsoft Word, to a non-proprietary one such as a plain text file or XML. To avoid losing data in the process, you should perform initial tests and analysis to determine exactly what changes will occur and whether they are acceptable. With both migration and conversion, special attention must be paid to also maintaining the accessibility of any associated metadata. When properly planned and executed, the migration and conversion approach probably represents the easiest and most cost-effective preservation method available

today.

## **E-Government and Collaboration**

The State of Minnesota's e-government framework should influence your preservation plans. The long-term preservation of records will demand a variety of investments and decisions that will involve time, staff, technology, and specialized expertise. Practically speaking, the state probably cannot afford to have every agency make all those investments independently. Similarly, no agency, even with the best of intentions, can consistently make all those decisions correctly. Finding effective and economic solutions means working together.

The state's enterprise technical architecture reflects that. In order to facilitate the development of e-government, the architecture identifies a series of issues, approaches, and standards that will make your agency's investments in information technology more likely to succeed. At the same time, these also will facilitate the long-term preservation of digital resources through sharing services and solutions. In developing your preservation strategy, start by looking at what other agencies are doing and what you can learn from their experiences.

## **Key Issues to Consider**

The foundation of your preservation plan should be your needs analysis, as well as an analysis of the costs, benefits, and risks involved with each of the options you are studying. Your records management, information technology, and legal staff should all be involved in the process to make sure your plan meets your business requirements and fits in with your general electronic records management strategy. Be sure to document your decision-making process in addition to your choices and plans for implementation.

At the minimum, your preservation plan should include the following items:

- Rationale and requirements for your preservation program.
- List of relevant records series and their retention and access requirements.
- Explanation of the selected preservation technique(s), including schedules for preservation actions, quality assurance testing, backups, etc. and instructions for documentation.
- Pointer to a business continuity or disaster recovery plan.

Once completed, your preservation plan should not gather dust on a shelf. Rather, it should be a reference document for all preservation activities, and it should be kept up to date as your situation changes (e.g., changes in use needs, hardware, software, media, security/access requirements, retention periods, legal mandates).

## Discussion Questions

As you move from your needs analysis to the development of your preservation plan, you will face many choices. These are just a few of the questions you should ask during the process.

- How long do we need to keep these records? What will be the costs associated with such preservation tasks as migration and conversion over time?
- What best practices can we identify and apply to our situation? Can we cooperate with other agencies or organizations to share expertise or save money?
- Do we need to keep the records in electronic format or is another format, such as paper or microfilm, more appropriate? How much functionality do we need to retain over time?
- How often are these records accessed? What is the best storage solution (e.g., online, nearline, offline)?
- What is the most appropriate storage media for the records? How will we ensure that we retain the hardware necessary to handle the media? What documentation should we collect and maintain regarding the media and hardware?
- How will we ensure that the content of the records is accessible and readable over time? Is the format and necessary software proprietary or non-proprietary? What documentation should we collect and maintain regarding format and software?
- How will we perform periodic quality assurance checks to ensure accessibility and trustworthiness over time? How will we document these checks?
- What indexing and metadata schemes should we employ to ensure that the records can be easily located and evaluated for use?
- How will these records be used? Will they be shared with others inside our organization? Outside? Would XML enhance the use-value of the records?
- Have the records been compressed or encrypted? If so, how does this fit into our management plan?
- Are there data access issues that require special security measures?
- What hardware and software configurations are we moving to in the foreseeable future? How do these records fit in with that plan?
- What staff training is necessary to ensure compliance with the preservation plan?

## Annotated List of Resources

### Primary Resources

Beyers, Fred R. *Information Technology: Care and Handling for the Preservation of CDs and DVDs – A Guide for Librarians and Archivists*. NIST Special Publication 500-252. Gaithersburg, MD: National Institute of Standards and Technology; Washington, D.C.: Council on Library and Information Resources. October 2003.

<<http://www.itl.nist.gov/div895/carefordisc/>>

*This guide discusses the physical characteristics of various optical media, as well as methods for their proper care and handling to ensure longest possible use in any given environment. A useful glossary is included.*

Lawrence, Gregory W., William R. Kehoe, Oya Y. Rieger, William H. Walters, and Anne R. Kenney. *Risk Management of Digital Information: A File Format Investigation*. Washington, D.C.: Council on Library and Information Resources. June 2000.

<<http://www.clir.org/pubs/reports/pub93/pub93.pdf>>

*This publication offers detailed guidance on migration (which is defined to include conversion) as a preservation technique through a risk assessment process. A useful workbook is provided to assist users in applying quantitative risk assessment measurements to their own environment.*

Lee, Kyong-Ho, Oliver Slatterly, Richang Lu, Xiao Tang, and Victor McCrary. “The State of the Art and Practice in Digital Preservation.” *Journal of Research of the National Institute of Standards and Technology*. 107(January-February 2002): 93-106.

<<http://nvl.nist.gov/pub/nistpubs/jres/107/1/cnt107-1.htm>>

*This paper provides a concise survey of a variety of preservation techniques for digital resources, including migration, emulation, encapsulation, and the use of eXtensible Markup Language (XML), as well as some project case studies.*

Minnesota Department of Administration, Office of Enterprise Technology. *Minnesota Enterprise Technical Architecture*. Version 2.02, 2006.

<<http://www.oet.state.mn.us/>>

*The Minnesota Office of Enterprise Technology (OET) is charged with establishing and maintaining a state information architecture as specified in Minnesota Statue, Chapter 16E.04 Subdivision 2. According to the OET, “This technical architecture is established to describe technology components of the State’s information infrastructure and their individual principles, practices and standards that are to be used to guide the development and delivery of all information systems services. The architecture will provide a reference so that various groups of government IT professionals have a consistent view of the information systems infrastructure and the methods that they employ to develop and deliver information systems services.” Chapter 4, Data and Records Management, describes the framework for managing information resources, as*

*well as the standards and guidelines that apply.*

World Wide Web Consortium (W3C). Extensible Markup Language (XML).

<<http://www.w3.org/XML/>>

*The W3C is the international body responsible for the development and ongoing refinement of the XML family of standards. This site provides links to the specification itself, as well as pointers to working groups and other resources.*

### **Additional Resources**

Minnesota Historical Society, State Archives Department. *Preserving and Disposing of Government Records*. St. Paul: Minnesota Historical Society, May 2008.

<[http://www.mnhs.org/preserve/records/docs\\_pdfs/PandD\\_may2008.pdf](http://www.mnhs.org/preserve/records/docs_pdfs/PandD_may2008.pdf)>

*Developed for Minnesota government agencies, this overview of the basic principles of records management includes chapters on defining a government record, taking inventory of your records, developing records retention schedules, preserving archival records, disposing of records, and setting up a records storage area. A list of resources for more information is included, as well as information about applicable state law regarding electronic records management. Originally published by the Minnesota Department of Administration in July 2000, the guide was updated jointly by the Minnesota Historical Society and the Minnesota Government Records and Information Network (MNGRIN) in 2008.*

Minnesota Historical Society, State Archives Department. *Trustworthy Information Systems Handbook*. Version 4, July 2002.

<<http://www.mnhs.org/preserve/records/tis/tis.html>>

*This handbook provides an overview for all stakeholders involved in government electronic records management. Topics center around ensuring accountability to elected officials and citizens by developing systems that create reliable and authentic information and records. The handbook outlines the characteristics that define trustworthy information, offers a methodology for ensuring trustworthiness, and provides a series of worksheets and tools for evaluating and refining system design and documentation.*