

# **Staying Digital: Recommendations on Preserving New Jersey Government Information in the Digital Age**

Report of the State Documents Interest Group of the  
Documents Association of New Jersey  
Report of the State Documents Interest Group  
Ad-Hoc Committee for the Preservation of Documents in Electronic Form

## **Contributors to the Report:**

Susan Kadezabek  
Susan Lyons – Report Editor  
Shelley Myer  
Susan Sabatino  
Dorothy Warner

## **Committee Participants:**

Geetali Basu, County College of Morris  
Carole Bruce, Seton Hall University  
Mary Alice Cicerale, Rutgers University  
Mary Fetzer, Rutgers University  
Ann Grice, East Brunswick Public Library  
Susan Kadezabek, Monmouth University  
Susan Lyons, Rutgers Law School - Newark (Chair, Ad-Hoc Committee)  
Shelley Myer, Rutgers SCILS Graduate Student  
Peggy Murtha, Rowan University  
Laura Saur, Newark Public Library  
Susan Sabatino, William Paterson University (Chair, State Documents Interest Group)  
Ella Strattis, Rowan University  
Dorothy Warner, Rider University

## **Documents Association of New Jersey 2001 Executive Committee**

A. Hays Butler – President  
Susan Lyons – Vice-President/President-Elect  
Ann B. Grice – Secretary  
Wen-Hua Ren – Treasurer/Web Page Editor/Past-President  
Mary Alice Cicerale – Member-at-Large  
Laura Saur – Chair, Federal Documents Interest Group/Archivist  
Susan Sabatino – Chair, State Documents Interest Group  
Mary Fetzer – Chair, International Documents Interest Group  
Geetali Basu – Membership Committee Chair  
Ellen Boncarti – Newsletter Editor  
Jan Wanggard – Liaison to Rutgers SCILS

## Executive Summary

### **Key Points:**

- The State of New Jersey now produces large amounts of information in digital form. A policy to protect and preserve that information is needed to ensure permanent public access to this information. Developing a plan for permanent public access will safeguard the State's investment in information and save New Jersey taxpayers money in the long run.
- Digital information is fragile. To preserve digital information over time, the information must be periodically copied or "refreshed." After a period of three to five years digital information must be "migrated" to work with newer software and hardware or the information will be inaccessible.
- Digital documents may be less expensive than analog documents to create, but are more expensive to maintain over time, given the costs of refreshing and migrating data.
- Creators of digital information can enhance longevity through the use of metadata that will make migration or emulation of data easier and less expensive.
- Documents that have enduring legal, historical or cultural value are best protected by distributing them in both tangible and electronic formats. The Internet may offer greater initial public access, but paper and microform are the only media proven to last for periods of 50 to 100 years and beyond.

### **Key Recommendations:**

- The State should develop a plan to preserve government information from the point of its creation and throughout its life cycle. We recommend that digital documents include descriptive metadata (e.g., XML) to facilitate migration of data to new software and hardware environments.
- Government documents are created by many state and municipal entities. A central repository is needed to ensure permanent access to these documents. The State Library has taken a leadership role in providing Internet access to documents created by many state agencies. Additional funding and staffing are necessary for the Library to properly store and safeguard valuable government information created in both electronic and tangible formats.
- Documents with enduring legal, historical or cultural value should continue to be distributed in both tangible and electronic formats. Our current system of state depository libraries provides the best method of preserving valuable government information for long-term use.

## **Introduction**

This report is the product of discussion among librarians who oversee New Jersey state government document depositories. Like many state governments, New Jersey now publishes many government documents directly on the Internet. As librarians we applaud these efforts to provide citizens with greater access to government information. We are also concerned for the long-term viability of such documents. Without careful planning, much of the digital information created today will be gone tomorrow.

The major concern of this report is government publications, that is, substantive reports and monographs that have traditionally been distributed to state depository libraries, as opposed to agency records, press releases and memoranda that are the concern of the State Archives. But the Internet has blurred that traditional distinction as agencies post both publications and records on their websites. From the perspective of library patrons, an electronic publication is anything they can find on the Internet. Internet postings of agency memoranda and press releases are frequently cited in newspapers and scholarly publications. As librarians are called upon to locate and verify these citations, our concern must embrace all electronic documents that have continuing long-term value. It is hoped that the recommendations made in this report will be of use to all state officials responsible for the creation, dissemination, and preservation of state government information, as well as the State Library and State Archives.

Though the digital age has barely begun, we have already lost tremendous quantities of data. Digital documents created and stored in legacy software such as COBOL, C/PM, D-Base, Wordstar, and even MS-DOS, are now inaccessible to most

computer users. The hardware necessary to view information stored on 8” and 5 ¼“ floppies, 8-track and betamax tapes, and other legacy formats has largely disappeared.

Government information is not immune from the threat of technological obsolescence. The original raw data from the 1960 decennial census was stored on a then state-of-the-art UNIVAC computer. When the Census Bureau turned the data over to the National Archives in the mid-1970’s UNIVAC computers were long obsolete. Heroic and costly rescue efforts recovered most, but not all, of the data. Other items lost to the digital black hole include much of the data from the Viking mission to Mars and pre-1979 Landsat images of the earth. In neighboring New York, all of the computerized data from a comprehensive 1960’s study that mapped land use and environmental data throughout the entire state was lost. The study had employed customized computer software that no longer existed when the computer tapes were turned over to the New York State Archives.<sup>1</sup>

Many people imagine digital documents to be indestructible. They don’t crumble or fade like paper and they can be copied quickly and inexpensively. In reality, digital media are far more fragile than paper or microform. All digital documents are stored as computer files on magnetic or magneto-optical media such as computer disks or tapes. Computer files may be erased by accidental exposure to a magnetic field or a surge in electric current. Exposure to oxidation and humidity can cause the substrate material of the disk to degrade. Even with proper storage, digital media degrade over time. According to National Archives, a CD will last from five to fifty years, depending on the quality of its manufacture.<sup>2</sup> The lifespan of magnetic tape, under the best of conditions, is measured in decades<sup>3</sup>. Unless the data is periodically “refreshed” by copying it from

one disk to another, it will become unreadable. And when digital data fails, it fails completely. In contrast, archival quality paper and microform can last up to 500 years.<sup>4</sup>

An even greater problem is that digital data is created to work within a particular software and hardware environment. As the software and hardware become obsolete, the data becomes less accessible, and finally, inaccessible. Two possible solutions exist to rescue older documents from a premature digital demise: migration and emulation.

These methods are discussed in detail in the section on digital preservation below.

For paper documents, decisions about preservation are usually made years after the document's creation when an archivist or records manager appraises its long-term value. If the document is deemed worthy of preservation, steps are taken to ensure its continued existence. Digital preservation requires a more pro-active approach. Unless planning is done at a much earlier point, ideally at the time the document is created, it will be costly and perhaps impossible to preserve the document years later. This report recommends steps the State of New Jersey can take to preserve vital government information for generations to come.

A final consideration in moving government information to the Internet is the problem of the "digital divide." The digital divide is a term used to refer to the differences, based on race, gender, geography, economic status, and physical ability, in access to information, the Internet, and other information technologies.<sup>5</sup> This term also includes differences in the knowledge and ability to use information and the technological skills needed to access digital or electronic information. The difficulties experienced by users and potential users of digital information complicate access and confound their efforts to keep informed. Areas affected by the digital divide include

individuals who rely on information to make decisions and improve their quality of life; businesses and economies that rely on informed customers and employees; government agencies that seek to communicate current information to citizens; and the democratic process that relies on informed participants.

By disseminating information via the Internet, the State of New Jersey has reached a greater audience and provided citizens with quick and easy access to basic information about state government. The Internet, however, does not reach everyone. Providing alternate access through depository libraries serves two purposes: it makes government information accessible to those who cannot or will not use the Internet, and it provides a reliable means of preserving that information for generations to come.

### **New Jersey Government Documents**

New Jersey has a vital and effective depository program that distributes government documents to fifty-one libraries in every corner of the state. This system ensures that New Jersey taxpayers can find important government information close to home. In the past several years, most New Jersey governmental bodies have undertaken efforts to publish documents directly on the Internet. This allows some New Jerseyans to download government information directly to their computers at home or work. Depository libraries offer additional access by providing the public with computers and by cataloging digital documents to aid patrons in locating specific documents.

While computer savvy researchers may turn to the Internet as the first stop, the life of a digital document can be frustratingly brief. The URL that worked last week turns up an error message today. Agencies may redesign their websites and remove

content without notice. Some agencies may leave reports on their servers for years, others for only a few months.

To combat the problem of shifting URLs and broken links, the State Library of New Jersey has established a web site entitled *NJ Gov't Publications on the Web*, located at <http://www.njstatelib.org/cyberdesk/gbgday2.htm>. This site has proved to be a reliable and comprehensive resource for New Jersey librarians searching for elusive state electronic documents. Despite the excellent work of the State Library in maintaining this site, it is not a long-term solution to the problem of disappearing documents. The source files for the various government documents remain (or do not remain) on the agency servers. The links or the file names may be changed at any time. Older documents may be overwritten by newer documents with the same file name. There does not appear to be any systematic policy for how long documents will remain available on the agency servers and what will become of the documents when the agency decides to remove them.

As state agencies move toward posting documents on the Internet, depository libraries receive fewer documents in tangible format. The number of documents shipped in 2000 dropped fifteen percent over 1998.<sup>6</sup> This decrease is concentrated in a few state agencies. Documents shipped to depositories by the Environmental Protection Agency dropped twenty seven percent in the same period, legislative documents by forty one percent and documents from the Treasury Department declined by thirty eight percent.<sup>7</sup>

This shift, toward electronic formats and away from tangible documents, raises serious concerns for the long-term survival of New Jersey government information. Without a comprehensive plan to preserve electronic information, the State risks the loss

of vast amounts of information by the end of the decade. The next section examines the topic of digital preservation and recommends specific actions the State can take to preserve government information for the long-term.

### **Electronic Preservation**

Digital documents are coded documents that consist of, at the most basic level, zeroes and ones. These digital bits convey information about the content and format of the document and also about the software environment necessary to resurrect the bits and bytes into a meaningful representation of the original document. To view the coded document it must be viewed within the software and hardware environment in which it was created, or a good emulation of the original environment. Otherwise we are left with a meaningless string of zeroes and ones.

The coded bits and bytes of a digital document are stored on magnetic media such as floppy disks, hard disks, tape, zip drives, CD-ROMs and DVDs. The average shelf life of a floppy disk is two to three years. The average shelf life of a standard CD-ROM is five to fifty years, although some high quality optical disks may last for up to 100 years.<sup>8</sup> A document stored on a floppy disk can be preserved by copying it or “refreshing” it to a new disk. The solution to digital preservation, however, is not as simple as storing documents on high quality optical disks, or faithfully refreshing the data. In one hundred years (or even in ten years) the software and hardware necessary to decode a digital document created today will no longer exist. To preserve the document for periods beyond a decade, additional steps are necessary.

The method most widely used to keep digital documents accessible is migration. Migration is defined as “the periodic transfer of digital materials from one

hardware/software configuration to another, or from one generation of computer technology to a subsequent generation.”<sup>9</sup> Migration is a more costly and labor intensive method of preservation of data than simply refreshing the data, but it is necessary if the document is to remain usable beyond the short lifecycle of the software in which it is embedded.

A simple example of migration is moving a document created in one version of word processing software to a newer version. Most word processing software programs will read the last few generations of the same program without problems or any loss of content of formatting. Converting a document created in Wordstar on a CP/M operating system twenty years ago is a far greater challenge. It is possible (assuming that the data has been periodically refreshed) to retrieve most of the content, if not the formatting, if one is willing to invest time and money in the recovery, but heroic recovery is not the ideal document management plan.

Documents that are created in simpler, standardized formats are usually easier to migrate than documents embedded in complex formats of word processing software. Two widely used coding standards are the American Standard Code for Information Interchange or ASCII, and Unicode, a standard that embraces a much larger character set than ASCII. Most word processors will save a document in ASCII format when the author selects the “txt” file extension. A universal format for graphic images, Tagged Image File Format or TIFF was developed in the 1980’s. Many documents, however, are created in proprietary software formats that become obsolete after a few years. Saving a master copy of an electronic document in open standard formats such as ASCII, Unicode

and TIFF will increase a document's longevity and make migration to newer formats easier.

Another migration strategy is to move the information from less stable to more stable media. The simplest expression of this is copying less stable digital files to paper or microform. Archival quality paper and microform can last up to 500 years. The analog copies are also not dependent on hardware or software for interpretation. This strategy may be acceptable for straight text documents, but more complex data is often lost through this method. For example, a digital spreadsheet may contain embedded formulas and have the functionality to manipulate the data in various ways. A printout of the spreadsheet preserves only one static view of the data.

Through the use of basic file formats and careful refreshing and migrating of data, text documents can be preserved for many years. More complex Internet documents, which combine text with graphic images, sound files and videos, or contain embedded programs (e.g. Java applets), are more difficult to preserve. In migrating multi-media digital documents, crucial elements of the formatting may be lost. To maintain complex digital files, some researchers advocate another preservation method known as emulation.

Emulation seeks to recreate a digital document's original functionality, look, and feel.<sup>10</sup> Emulation would create software that would mimic the functionality of older software and hardware, allowing legacy digital documents created in obsolete software and hardware environments to be viewed on current computer systems. Some researchers have been critical of this strategy, arguing that there is no "magic bullet solution" to the problem of digital preservation and that migration remains the only proven method for preserving digital documents.<sup>11</sup> Emulation offers the promise of

preserving complex digital files, but finding a practical method of emulation will require much more research.

To facilitate both migration and emulations strategies, digital archivists advocate applying descriptive metadata to a document at or near the time of its creation.

Traditionally, librarians have used metadata to describe the intellectual content of documents, both analog and digital, in their library catalogs. An example of a metadata standard used by most libraries is US MARC. Other metadata standards include Dublin Core, SGML and XML. For preservation purposes, metadata can be used to describe the software and hardware environment in which a document was created, as well as information about its appearance and functionality. Metadata can also be used to embed a “digital signature” that can serve to verify a document’s provenance and authenticity. While much research is being done to develop universal standards, as of the writing of this report, there is no universally accepted standard for descriptive metadata. Research to develop a uniform standard for preservation metadata is being conducted by OCLC and the Research Libraries Group<sup>12</sup> and the National Archives.<sup>13</sup>

Digital preservation is a labor-intensive effort, and therefore costly. A report done for the British Library in 1998 enumerates some of the costs of creating, managing and preserving digital information.<sup>14</sup> The report argues that the costs of digital preservation can be minimized by careful planning from the point of creation of the data and throughout its lifecycle. The next section offers recommendations for the State that will reduce the costs of preservation efforts to the taxpayers while ensuring permanent public access to vital government information.

## **Recommendations**

1. Documents that have enduring legal, historical or cultural value should be preserved to ensure permanent public access.
2. Some government documents are ephemeral and do not merit long-term preservation. By identifying those documents that must be preserved and those that may be deleted after a period of time, the State can maximize its preservation resources.
3. State digital documents of enduring value should be stored in a central repository. Placing source files on the servers of the State Library or State Archives would protect information from deletion and allow for the assignment of a permanent uniform resource locator or PURL.
4. Documents that must be preserved for periods of greater than a decade should be distributed in tangible format to the State's depository libraries. Paper and microform are stable media that will ensure cost-effective preservation and access to the intellectual content of valuable State documents.
5. Where possible, master copies of documents should be stored in open standard formats, such as ASCII, Unicode, HTML, XML or TIFF.
6. The State should explore embedding digital documents with descriptive metadata that will enhance the chances for successful migration of the documents to new software and hardware environments.
7. The State Library should catalog digital documents of enduring value. This will increase the likelihood that libraries will include state documents in their online catalogs and increase public access to these resources. This committee recognizes

- that additional staff resources will be necessary to catalog state electronic documents. While there is still debate about the best methods of cataloging electronic documents, the work of the federal Government Printing Office offers a good model for cataloging electronic government documents.
8. Cooperation among all state agencies and officials responsible for the creation and dissemination of electronic documents is essential. State agencies should work closely with the State Library and State Archives to develop preservation strategies for digital information.
  9. The State should develop a master plan to preserve its ever-growing store of digital information. Unless planning and resources are directed at preserving the state's digital information base, the state faces a catastrophic loss of vital information in the next decade.

### **Conclusion**

The State of New Jersey has entered the digital age and now provides both information and services to many of its citizens directly over the Internet. For the State to remain digital it must take immediate steps to preserve its digital heritage. New Jersey is fortunate to have a pro-active State Library that has already taken steps to provide greater access through a central web site with links to many agency documents, but more is needed. Preservation of state government information over the long-term will require careful planning, clear policies and standards for those who create information and resources. These efforts, if undertaken now, will not only save money for New Jersey taxpayers, they will preserve the heritage of vital state information for generations to come.

## Notes

---

<sup>1</sup> Waters, Donald and John Garret, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*, 1996, 2-3. [<http://www.rlg.org/ArchTF/>]

<sup>2</sup> National Archives and Records Administration, <http://webgopher.nara.gov/0/managers/archival/papers/optical/critiss.txt>. See also the Stanford Conservation Online (CoOL) web site on electronic storage media for additional reports and links on the longevity of electronic media at: <http://palimpsest.Stanford.edu/electronic-records/electronic-storage-media/>

<sup>3</sup> Van Bogart, John W. C., *Magnetic Tape Storage and Handling*, Commission on Preservation and Access and the National Media Laboratory, 1995. [<http://www.clir.org/pubs/reports/pub54/>]

<sup>4</sup> Puglia, Steven, *Creating Permanent and Durable Information: Physical Media and Storage Standards*, Cultural Resource Management, No. 2 – 1999, 25. Available online at [<http://www.cr.nps.gov/crm/>] and [<http://purl.access.gpo.gov/GPO/LPS1447>]

<sup>5</sup> *What is the Digital Divide?* Washington, D.C.: American Library Association, Office for Information Technology Policy. Available at [<http://www.ala.org/oitp/digitaldivided.what.html>]

<sup>6</sup> Myer, Shelley, *Availability of New Jersey State Government Documents Online*. Graduate research paper, Rutgers University School of Communication, Information and Library Studies, 2000, 3.

<sup>7</sup> *Ibid.*, 4.

<sup>8</sup> National Archives and Records Administration, <http://webgopher.nara.gov/0/managers/archival/papers/optical/critiss.txt>

<sup>9</sup> Waters and Garret, *Preserving Digital Information*, 6.

<sup>10</sup> Rothenberg, Jeff, *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. The Council on Library and Information Resources, 1999, 17.

<sup>11</sup> Bearman, David, *Reality and Chimeras in the Preservation of Electronic Records*. D-Lib Magazine, April 1999. [<http://www.dlib.org/dlib/april99/bearman/04bearman.html>]

<sup>12</sup> OCLC/RLG Working Group on Preservation Metadata. *Preservation Metadata for Digital Objects: A Review of the State of the Art*. January 31, 2001. [[http://www.oclc.org/digitalpreservation/presmeta\\_wp.pdf](http://www.oclc.org/digitalpreservation/presmeta_wp.pdf)]

<sup>13</sup> Thibodeau, Kenneth, *Building the Archives of the Future*. D-Lib Magazine, February 2001. [<http://www.dlib.org/february01/thibodeau/02thibodeau.html>]

<sup>14</sup> Hendley, Tony. *Comparison of Methods & Costs of Digital Preservation*. British Library Innovation Report No. 106, 1998. [<http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html>]