**Provenance, Journal of the Society of Georgia Archivists**

Volume 25 | Number 1 Article 3

January 2007

# Designing a Preservation Survey: The Digital Library of Georgia

Sheila McAlister
*Digital Library of Georgia*

# Designing a Preservation Survey: The Digital Library of Georgia

**Sheila McAlister**

Since the mid-1990s, libraries have been digitizing cultural-heritage resource materials for access purposes. The digital medium provides additional opportunities for innovative approaches to scholarship and the creation of new collections through the aggregation of geographically distributed materials of similar provenance or theme. According to Donald Waters, formerly head of the Digital Library Federation, "the promise of digital technology is for libraries to extend the reach of research and education, improve the quality of learning, and reshape scholarly communication."[1] Accordingly, the cultural-heritage community has widely embraced digitization. In 2002, Clifford Lynch pointed to this widespread acceptance:

> We're getting pretty good at digitizing material at scale. We have a wealth of experience and a large number of successful projects (not to mention some highly educational failures) to build upon.... [T]he research

[1] Donald Waters quoted in Abby Smith, "Why Digitize?," Washington, D.C.: Council of Library and Information Resources, February 1999, http://www.clir.org/pubs/reports/pub80-smith/pub80.html (accessed November 30, 2007).

> questions are less about how to do it at all and more about how to optimize—how to do it more efficiently or effectively, how to be sure that you've chosen the most appropriate strategies and technologies. We are training a large cadre of people qualified to plan, manage, and execute digitization projects through vehicles like the Schools for Scanning. Best practices are becoming well established—consider the work that IMLS [the Institute of Museum and Library Services] has done in this area, or the Digital Library Federation, or the forthcoming *Guide to Good Practice* in preparation by the National Coalition for a Networked Cultural Heritage (NINCH). Costs are becoming more predictable for these projects. There are commercial and non-commercial mass production operations that are becoming well established to support organizations that want to do large-scale digitization; one no longer has to do it in house as part of a research and development effort.[2]

Consequently, digital files are now counted among an institution's assets and must be considered as part of its strategic preservation planning.

As Paul Conway says, "[t]he essence of preservation management is resource allocation. People, money, and materials must be acquired, organized, and put to work to ensure that information sources are given adequate protection."[3] In an era during which libraries and other cultural-heritage institutions are increasingly building digital collections, the question of resource allocation for preservation becomes increasingly complicated. Preservation of digital objects is an ongoing and potentially labor-intensive endeavor that is centered around short "preservation

---

[2] Clifford Lynch, "Digital Collections, Digital Libraries and the Digitization of Cultural Heritage Information," *First Monday* 7, no. 5 (May 2002), <http://firstmonday.org/issues/issue7_5/lynch/index.html> (accessed November 30, 2007).

[3] Paul Conway, "Preservation in the Digital World" (Washington, D.C.: Council of Library and Information Resources, March 1996), <http://www.clir.org/pubs/abstract/pub63.html> (accessed November 30, 2007).

cycles." Currently, cost models for such endeavors are few.[4] As such, the incorporation of digital preservation needs into an institution's preservation-management plan is necessary for balancing resource allocation.

As a first step in the re-examination of preservation priorities, the needs-assessment survey provides the raw data necessary for creation of a strategic vision for preservation. Sherelyn Ogden explains:

> A survey must evaluate the policies, practices, and conditions in an institution that affect the preservation of all the collections. It must address the general state of all the collections, what is needed to improve that state, and how to preserve the collections long-term. It must identify specific preservation needs, recommend actions to meet those needs, and prioritize the recommended actions.[5]

Most survey instruments currently available are geared towards more traditional collections. For example, Beth Patkus's 2003 self-survey guide addresses paper-based materials both bound and unbound, photographs and negatives, oversized and framed materials, newsprint, scrapbooks and ephemera,

---

[4] The Florida Center for Library Automation received an IMLS grant to develop a working digital preservation archive to be used by the Florida public universities. See their final report at <http://www.fcla.edu/digitalArchive/pdfs/FinalReport.pdf> (accessed May 28, 2008). See also Tony Hendley, "Comparison of Methods & Costs of Digital Preservation," 1998, at <http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html> (accessed May 28, 2008); Steve Chapman, "Counting the Costs of Digital Preservation: Is Repository Storage Affordable?," *Journal of Digital Information* 4, no. 2, <http://jodi.tamu.edu/Articles/v04/i02/Chapman/> (accessed May 28, 2008); Shelby Sanett, "The Cost to Preserve Authentic Electronic Records in Perpetuity: Comparing Costs across Cost Models and Cost Frameworks" *RLG DigiNews* 7, no. 4 (August 15, 2003), at <http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?fileid=00000070511:000006283731&reqid=92451#feature2> (accessed April 8, 2008). In July 2005, the Digital Preservation Coalition held a workshop on cost modeling the preservation of digital assets.

[5] Sherelyn Ogden, "What is Preservation Planning" in *Preservation of Library and Archival Materials: A Manual*, ed. Sherelyn Ogden, 3rd ed., rev. and expanded (Andover, Mass.: Northeast Document Conservation Center, c1999), <http://www.nedcc.org/resources/leaflets/1Planning_and_Prioritizing/01WhatIsPreservationPlanning.php> (accessed November 30, 2007).

audiovisual materials as well as reformatted objects.[6] Yet Patkus's treatment of reformatting through digitization is very general, and the volume as a whole does not consider some of the special requirements for digital collections. Furthermore, the survey does not address many specific needs, such as those of a state-wide digital project, which may be charged with safeguarding the digital assets of distributed institutions.

Therefore, I propose to use Patkus's preservation needs-assessment survey as a framework for use by digital projects, with special reference to the digital collections of the Digital Library of Georgia (DLG). The digital-preservation needs-assessment survey is intended to be used over a series of years, so it will contain questions that do not apply to the current state of the DLG. In order to adapt the survey effectively, it is important to survey both the institutional context of the Digital Library of Georgia and the current digital-preservation landscape. Issues such as the barriers to digital preservation, requirements of digital-preservation systems, the current preservation strategies employed, and best practices with regards to metadata and digital object creation must be considered. A thorough understanding of these aspects of the problem is necessary also for the eventual evaluation of survey responses.

## I. THE DLG'S INSTITUTIONAL CONTEXT

Based at the University of Georgia Libraries under the auspices of GALILEO, Georgia's Virtual Library, the DLG is a collaborative digital-library program that assists Georgia libraries, archives, and cultural-heritage organizations in digitizing and publishing online resources related to life in the state. The DLG actively develops, maintains, and preserves digital-library content and provides access to Georgia-related, digitized resources. With the help of Georgia HomePLACE (Providing Libraries and Archives Electronically), the Digital Library has recently reached out to public libraries to assist them in making their local-history resources available online. The Digital Library's infrastructure includes a state-wide metadata catalog and archival storage for the master files of the HomePLACE partner institutions and other

---

[6] Beth Patkus, "Assessing Preservation Needs: A Self-Survey Guide" (Andover, Mass.: Northeast Document Conservation Center, 2003), <http://www.nedcc.org/resources/downloads/apnssg.pdf> (accessed November 30, 2007).

grant-funded collaborative projects. As of November 2007, DLG is responsible for the stewardship of thirty-five digital collections and approximately eleven terabytes of master files.

## II. THE DIGITAL PRESERVATION LANDSCAPE

### *Barriers to Digital Preservation*

When considering the technological barriers to digital preservation, many experts identify three aspects of the problem: media longevity, and software and hardware obsolescence. Media longevity deals with the lifespan of the digital information's carrier. Over time, the device will deteriorate. Because of the nature of digital storage, one small flaw or scratch can be catastrophic. If a sector of the media is damaged, one may be unable to access any information from it. The proper care and handling of digital media has a direct effect on its longevity. In 1996, a National Media Lab study said the average digital media device had a lifespan of less than five years.[7]

The commercial and changing nature of technology also affects hardware and software. In 1976, 10,000 records of the 1960 Census were lost during the migration process because the data was stored on an obsolete tape drive. Many of the Vietnam War-era electronic documents are unusable because they can only be accessed by obsolete hardware.[8] It is neither feasible nor cost effective to attempt to maintain museums of antiquated computer equipment for preservation purposes.[9] Software, too, poses similar challenges. Popular desktop applications are only engineered to be backward compatible by a few versions. Software encryption can also be a preservation barrier.

The easy mutability of digital objects or lack of fixity also may be problematic. In order to demonstrate that a digital object has not changed over time, checksums and digital signatures

---

[7] Jeff Rothenberg, "Avoiding Technological Quicksand" (Washington, D.C.: Council of Library and Information Resources, 1999), 7, <http://www.clir.org/pubs/reports/rothenberg/pub77.pdf> (accessed November 30, 2007).

[8] Susan S. Lazinger, *Digital Preservation and Metadata: History, Theory, Practice* (Englewood, Colo.: Libraries Unlimited, 2001), 9.

[9] Rothenberg, "Avoiding Technological Quicksand," 12-13.

may be used as a means of verification.[10] Additionally, one must be able to ensure that a digital object is authentic or, as Peter Graham says, one must ensure "intellectual preservation."[11] In discussing the authenticity issues related to electronic records, Anne Gilliland-Swetland and Philip B. Eppard describe the base-level requirements for establishing authenticity: "[They] may be very similar to the heuristics that information literacy programs seek to inculcate in end users working with any type of information—that is, establishing the who, what, when, where, how, and why associated with that information."[12] Additionally, the reliability of a digital object can be demonstrated through systems controls during its life-cycle.

*Requirements For Digital Preservation Systems*

In 1990, the Consultative Committee for Space Data Systems (CCSDS) began to create a reference model for developing archives of digital data. The model, known as the Open Archival Information System (OAIS), delineates the basic functions and responsibilities of an archive dedicated to the long-term storage of digital data. The five functions of the system are to ingest data or accept submission information packages (SIP), archive data objects known as archival information packages (AIP), manage data including descriptive data as well as handling day-to-day management of the archive, and provide users access to the repository's data objects sent in the form of dissemination

---

[10] Because it is easy to change digital objects, digital preservation must demonstrate that an object has *fixity*, i.e., that it has remained unchanged from the original. *Checksums* are values created by adding up the bytes of a message. They are used to ensure that a file has not been altered or corrupted.

[11] Peter S. Graham, "Issues in Digital Archiving" in *Preservation: Issues and Planning*, eds. Paul N. Banks and Roberta Pilette (Chicago, Ill.: American Library Association, 2000), 101.

[12] Anne J. Gilliland-Swetland, and Philip B. Eppard, "Preserving the Authenticity of Contingent Digital Objects: The InterPARES Project," *D-Lib Magazine* 6, no. 7/8 (July/August 2000), <ttp://www.dlib.org/dlib/july00/eppard/07eppard.html> (accessed November 30, 2007).

information packages (DIP).[13] In discussing the AIP in further detail, the standard describes the necessary components to preserve a digital object over time. The AIP consists of the digital object itself as well as any representation data (in the case of emulation[14] this would include emulators and their own suite of metadata), preservation description information (PDI), packaging information (PI), and descriptive information (DI).

The impact of OAIS was deepened through the development of the concept of trusted digital repositories. These repositories are committed to providing reliable, long-term access to digital resources for a specific community of users. In order for a repository to be "trusted," system requirements include financial security and sustainability; standards-based methods for the ongoing management, access, and security of deposited materials; and auditability and procedures for systems evaluation. Responsibilities of such archives include ingesting, controlling, and maintaining data and their accompanying metadata; following well-documented policies and procedures for collections development, access control, storage, and updating of procedures over time; providing access to the community of users; and encouraging content providers to follow current best practices for digital object creation.[15]

*Preservation Strategies*

A wide variety of digital-preservation strategies exist currently, and most repositories employ a combination of

---

[13] For a fuller discussion of OAIS, see Brian Lavoie's "The Open Archival Information System Reference Model: Introductory Guide," <http://www.dpconline.org/docs/lavoie_OAIS.pdf> and the standard itself, the most current version of which may be found at <http://public.ccsds.org/publications/archive/650x0b1.pdf>.

[14] *Emulation* is a digital-preservation strategy that employs programs to translate another computer environment into a newer one. Emulation attempts to imitate the original functionality and look-and-feel of a system. For a fuller discussion, see Rothenberg, "Avoiding Technological Quicksand."

[15] RLG/OCLC Working Group on Digital Archive Attributes, "Trusted Digital Repositories: Attributes and Responsibilities: An RLG-OCLC Report" (Mountain View, Calif.: Research Libraries Group, 2002), 55-56, <http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf> (accessed November 30, 2007).

them. Each method has varying success addressing viability, renderability, and the understandability of digital objects. At the most basic level of preservation is *redundancy*. Primarily used as a disaster mitigation strategy, redundancy or *bitstream copying* is the creation of an exact copy of the object. Often accompanied by remote storage, bitstream copying is also employed by the consortial project LOCKSS (Lots of Copies Keeps Stuff Safe). Redundancy does not ensure that a digital object can be rendered properly or that it can be understood. It provides only a back-up copy.

By contrast, *refreshing* addresses issues of media decay and obsolescence. During refreshing, one moves the data from one durable or persistent storage medium to another without altering the bitstream. However, refreshing alone is not a viable approach as it does not address hardware or software obsolescence. Even though the media is not decayed, it may be impossible for the digital object to be understood by humans or computers.

Several other strategies have been proposed to combat technological obsolescence of hardware or software. While altering the digital object to transfer it from one technological environment to another, *migration* attempts to ensure that the object continues to possess its essential characteristics. For example, one performs migration when one updates a file that utilizes an obsolete version of Word Star to the current incarnation of Microsoft Word. During the transfer process, there may be some loss of data, and it may be difficult to identify these losses. Moreover, critics point out that it can be not only a time-consuming and complex proposition, but that because of the speed at which technology advances, it is difficult to predict how often migration may need to be performed. A corollary to migration is *canonicalization*, a strategy designed to test migration integrity through the comparison of a migrated object to a "canonical" version that describes its key features.[16]

Digital programs may also rely on the use of file formats that are standards. It is thought that widely adopted standards-

---

[16] For more information of canonicalization, see Clifford Lynch, "Canonicalization: A Fundamental Tool to Facilitate Preservation and Management of Digital Information," *D-Lib Magazine* 5, no. 9 (September 1999), <http://www.dlib.org/dlib/september99/09lynch.html> (accessed November 30, 2007).

compliant file formats are more likely to be viable over the long term. The sheer mass of users will push the market to address such a file format in new technologies. Repositories may choose to rely on a handful of standard file formats and convert all other formats to these preferred standard ones. This strategy is known as *normalization*.

A final strategy is *emulation*. It seeks to mimic the original technological environment of a digital object and to allow it to behave as it did with its original platform, software, and hardware. It employs programs to translate one computer environment into a newer one. Emulation attempts to imitate the original functionality and look-and-feel of a system.

*Metadata*

Metadata (commonly known as "data about data") aids in the discovery, longevity, and interoperability of digital objects. Commonly divided into three categories—descriptive, structural, and administrative metadata—it plays an integral role in any digital-preservation strategy.[17] Administrative metadata, the broad type within which preservation metadata falls, governs the data needed to manage a digital object over its entire life-cycle. Preservation metadata provides "the information necessary to maintain the viability, renderability, and understandability of digital resources over the long-term."[18] It may document the digital object's source, content, and structure and elucidate the relationships of the various parts of a digital object as well as technical information about its creation and life cycle. It uniquely identifies the object, documents its history and context, and creates an audit trail to demonstrate fixity. The data assists

---

[17] According to scholars, the categories of metadata vary. Some relegate technical, preservation, and administrative metadata to separate categories. See, for example, Cornell University's Moving Theory Into Practice tutorial. Others add usage metadata as a separate category. See Anne Gilliand-Swetland's "Setting the Stage" in the Getty Research Institute's "Introduction to Metadata," <http://www.getty.edu/research/conducting_research/standards/intrometadata/setting.html> (accessed July 7, 2008).

[18] OCLC/RLG Working Group on Preservation Metadata, *A Metadata Framework to Support the Preservation of Digital Objects* (Dublin, Ohio: OCLC, 2002), 1, <http://www.oclc.org/research/pmwg/pm_framework.pdf> (accessed November 30, 2007).

managers in making appropriate preservation decisions and supports the rendering and interpretation of a digital object despite technological changes. The metadata may encapsulate the digital object.

In 2000, Online Computer Library Center, Inc. (OCLC) and the Research Library Group (RLG) drew together an international team to compare the preservation metadata elements employed by a variety of digital-preservation projects from around the world. Using OAIS as the basis for their enquiry, the team enumerated an extensive list of elements; however, the project did not provide the practical tools and methods for data capture and management. Since the development of the OCLC/RLG framework, several projects have begun to explore the practical side of preservation metadata including the PREMIS (PREservation Metadata: Implementation Strategies) Working Group and the National Library of New Zealand. The PREMIS Working Group identified the core elements necessary for digital-preservation activities along with examples of the data dictionary's use in its May 2005 final report.[19] Free tools for capturing technical and other preservation metadata include DROID, JHOVE, and the National Library of New Zealand's Metadata Extractor.[20]

*Digital Object Creation*

One of the responsibilities outlined for trusted digital repositories is advocacy for creation of digital content that follows best practices and standards, for "the preservation and archiving process is made more efficient when attention is paid to issues of consistency, format, standardization and metadata

---

[19] PREMIS Working Group, "Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group," http://www.oclc.org/research/projects/pmwg/premis-final.pdf (accessed November 30, 2007).

[20] DROID, created by the National Archives in the United Kingdom, identifies file formats through a batch process (see <http://droid.sourceforge.net/wiki/index.php/Introduction> accessed May 28, 2008). JHOVE identifies, validates, and characterizes file formats (see <http://hul.harvard.edu/jhove/index.html> accessed May 28, 2008). The Metadata Extraction Tool extracts preservation-related metadata from digital files and outputs it in XML (see <http://www.natlib.govt.nz/about-us/current-initiatives/metadata-extraction-tool> accessed May 28, 2008).

description in the very beginning of the information life cycle."[21] A variety of standards and guidelines exist, including *Moving Theory into Practice*, the NINCH Guidelines, and the Northeast Document Conservation Center *Handbook*. At creation, the digital-preservation cycle begins and thus the context of creation should be captured through appropriate metadata.[22]

## III. SURVEY DESIGN

Now that both the DLG's institutional context and the overarching issues of the preservation of digital objects have been examined, it is time to consider the survey itself. Patkus's survey examines the institution and its collections, the building plant, environmental control and conditions, and disaster planning and security, all of which must be considered for both analog and digital collections.

### *Institutional and Collections Overview*

When beginning a preservation survey, one considers the institutional context and the holdings of the institution. In the case of digital library projects, particularly those with issues of distributed ownership, a careful analysis of the relationships between repositories may be necessary. The DLG, for example, digitizes materials held at other repositories and, save the microfilm for the Georgia Newspaper Project, has no analog collections. What licensing agreements for the digital content exist and what do they allow? Who has chief responsibility for these digital assets and to whom do the assets belong? Are preservation responsibilities spread across institutions and departments? Will

---

[21] Gail M. Hodge, "Best Practices for Digital Archiving: An Information Life Cycle Approach" *D-Lib Magazine* 6, no. 1, <http://www.dlib.org/dlib/january00/01hodge.html> (accessed November 30, 2007).

[22] Anne Kenney and Oya Rieger, *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, (Mountain View, Cal.: Research Libraries Group, 2000); *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials* (Washington, D.C.: National Initiative for a Networked Cultural Heritage, c2002), <http://www.nyu.edu/its/humanities/ninchguide/> (accessed November 30, 2007); and Maxine K. Sitts, ed., *Handbook for Digital Projects: A Management Tool for Preservation and Access* (Andover, Mass.: Northeast Document Conservation Center, 2000), <http://nedcc.org/oldnedccsite/digital/dighome.htm> (accessed November 30, 2007).

any activities be outsourced? Have these tasks been delineated? Is the service fee-based or will other revenue strands provide funding? The Florida Center for Library Automation, for example, developed a model contract between the libraries and the Florida Digital Archive to clarify such issues.[23]

Issues of ownership and intellectual property rights do not extend only to the content of the objects. Some methods of digital preservation, such as emulation, require knowledge of proprietary information. If using emulation, a project may need to identify such rights holders and secure their permission to copy, alter, and emulate. Also, accessing copy-protected materials may be problematic. For example, the Digital Millenium Copyright Act prohibits the "circumvention of technological access controls" and the distribution of programs that do so.[24] These rights holders may include not only the content creators, but also software, hardware, and platform developers. In response to such issues, the Library of Congress's National Digital Information Infrastructure and Preservation Program and the U.S. Copyright Office convened a group of copyright experts to recommend how Section 108 of the copyright law might be altered for the digital age. At this writing, the Section 108 Study Group has held three public roundtables to gather comments.[25]

In considering the basic composition of collections for digital-library projects, recording information on the types of materials, quantity, and units of measurement may not be enough. Digital objects may be composed of many individual files and file types. For example, the digital object for a digitized book may include several hundred master tiff files, derivative jpgs and thumbnails, and a full-text searchable XML file encoded using the Text Encoding Initiative (TEI) schema or DTD. For the purposes of considering the scope and volume of the DLG's collection, one would want to consider "material" types (i.e.,

---

[23] Florida Digital Archive, "Interim Report 2," Florida Center for Library Automation, 2003, <http://www.fcla.edu/digitalArchive/pdfs/interimReport2.pdf> (accessed November 30, 2007).

[24] June M. Besek, "Copyright Issues Relevant to the Creation of a Digital Archive: A Preliminary Assessment" (Washington, D.C.: Council of Library and Information Resources, January 2003), 13, <http://www.clir.org/pubs/reports/pub112/pub112.pdf> (accessed November 30, 2007).

image, text, sound, moving image, or multimedia), file formats, numbers of digital objects and files, and the total volume of data. In addressing selection, the format and purpose of files as well as institution of origin also should be considered. In the current version of the DLG's "Archival Master Data Storage Policy," for example, priority for preservation is given to master files of Georgia HomePLACE-funded projects.

*Surveying the Building: The Physical Plant*

Digital libraries may need to consider more structures than just their own buildings. As redundancy of data is a hallmark of digital preservation, one may also want to consider off-site storage facilities as well. The University of Michigan's Digital Library Production Services, for example, stores three copies of any file: one on a production server, one in offline storage, and a third on magnetic tape.[26] Other than consideration of the redundancy issue, no changes would be made to Patkus's building survey.

*Environmental Conditions, Storage, and Handling*

As with more traditional library collections, digital-library media longevity is dependent on environmental factors including climate and light exposure. For optical media such as CD-ROMs and DVDs, stable relative humidity and temperature is necessary. ISO 18925 recommends that for both types of media temperatures range between 14°F and 73°F with a relative humidity of 20-50 percent that cycles no more than ±10 percent.[27] The Association for Moving Image Archivists (AMIA) recommends that polyester-based magnetic tape be stored at either 20°C (68°F) and 20-30%

---

[25] Section 108 Study Group Web site, <http://www.loc.gov/section108> (accessed November 30, 2007).

[26] Maria Bonn, "University of Michigan Polices and Practice for the Long Term Retention of Locally Produced Digital Projects and Materials: A Report Prepared for the Joint RLG/TASK Force on Digital Preservation" (Ann Arbor: University of Michigan, Digital Library Production Services, 1998), <http://www.lib.umich.edu/lit/dlps/pubs/um-rlg.html> (accessed November 30, 2007).

[27] Fred R. Byers, "Care and Handling of CDs and DVDs" (Washington, D.C.: Council on Library and Information Resources and the National Institute of Standards and Technology, October 2003), 16, <http://www.clir.org/pubs/reports/pub121/pub121.pdf> (accessed November 30, 2007).

RH; 15°C (59°F) and 20-40% RH; or 10°C (50°F) and 20-50% RH. For optimum long-term storage, tapes should be stored at approximately 8°C ±2°C (46°F ±4°F) and 25% ±5% RH.[28] The Digital Preservation Coalition also provides guidelines for environmental conditions based on the British Standards Institution's BS4783 that takes into account the level of access required for the media.[29] Servers and on-, off-, and near-line storage also require stable, cool temperatures.

CD-Rs' longevity is compromised by prolonged exposure to both ultraviolet (UV) and infrared light. Sunlight increases the rate of degradation of CD-Rs' dye layer; whereas DVDs and CDs-RW are more prone to damage through heat build-up from infrared light. Likewise, magnetic tape is damaged by UV light so it should not be exposed to direct sunlight or other sources of UV light.

While optical media are immune to the effects of magnetism, magnetic tape may suffer from exposure to strong magnetic fields. AMIA recommends "that a tape can be stored safely in a magnetic field with a maximum strength of 1/10 of the tape's coercivity. A more conservative figure of 1/20 provides a safer margin of error. To determine a tape's coercivity, refer to the product's specification sheet available from the manufacturer."[30] Nonetheless, Cornell University's tutorial "Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems" recommends avoiding such exposure.[31] Storage cabinets should be electrically grounded.

---

[28] Association of Moving Image Archivists, "Fact Sheet 8—Environmental Conditions," 2003, <http://www.amianet.org/resources/guides/fact_sheets.pdf> (accessed November 30, 2007).

[29] Maggie Jones, and Neil Beagrie, eds.,"Environmental Conditions" in *Preservation Management of Digital Materials: A Handbook* (Digital Preservation Coalition, 2001), <http://www.dpconline.org/graphics/orgact/storage.html#enviro1> (accessed November 30, 2007).

[30] Association of Moving Image Archivists, "Fact Sheet 6—Common Tape Problems," 2003, http://www.amianet.org/resources/guides/fact_sheets.pdf (accessed November 30, 2007).

[31] Cornell University Library, Instruction, Research, and Information Services, "Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems," 2003, <http://www.icpsr.umich.edu/dpm/dpm-eng/oldmedia/mediathreats.html> (accessed November 30, 2007).

When storing media, one should control contaminants and pests by avoiding exposure to dust and fumes (including cigarette smoke). Additionally, there should be no food or drink in the storage areas. The media should be stored vertically, and hardware must be maintained. One should use lint-free gloves or clean, dry hands when handling media, and the exposed media should not be handled. Optical media should not be labeled using pens, pencils, or adhesive labels.

*Disaster Planning and Security*

Digital libraries need to consider threats to their collections, including natural or man-made disasters. Through adequate planning and consideration of security and other external threats, one may more successfully mitigate emergencies. Staff members should be trained to respond appropriately, and off-site storage and redundancy of data is essential. Likewise, security procedures safeguard the digital resources from unauthorized changes, deter hacking and other security invasions, protect authenticity, and provide for accountability through audit trails or random checking. Physical access should be limited by storage in a protected area, and virtual access should be protected through passwords and other network security procedures such as write-once policies.[32]

## CONCLUSION

While many of the elements of preservation planning for digital objects mirror those of more traditional library materials (i.e., security, disaster planning, environmental controls, etc.), issues related to ownership, mutability, and the speed of technological change make planning all the more important. Institutions must balance not only resources and technological capacity, but also an adequate policy framework to adequately address long-term stewardship of digital objects.[33] A preservation

---

[32] Jones and Beagrie, "Security," <http://www.dpconline.org/graphics/orgact/storage.html#secur2> (accessed November 30, 2007).

[33] On Cornell's "three-legged stool," see Cornell University Library, Instruction, Research, and Information Services, "Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems," 2003, <http://www.icpsr.umich.edu/dpm/dpm-eng/conclusion.html> (accessed November 30, 2007).

needs-assessment is a critical piece in benchmarking a repository's readiness for such activities and its areas of concern. A modified version of Beth Patkus's preservation needs-assessment survey, as suggested by the adapted questionnaire in the Appendix, can serve as a basis for such activities. Self-assessment is key in the iterative process of digital preservation. An institution must understand not only its own context, but also the critical issues facing digital content. Thus, an institution must look internally and to current and future developments in the technological landscape.

**Sheila McAlister** is the assistant director of the Digital Library of Georgia. Prior to her arrival in DLG, she worked as the electronic access coordinator for the Richard B. Russell Library for Political Research and Studies.

## APPENDIX

QUESTIONNAIRE
Adapted from Beth Patkus, "Assessing Preservation Needs: A Self-Survey Guide" (Andover, Mass.: Northeast Document Conservation Center, 2003).

GENERAL INSTITUTIONAL, COLLECTIONS, AND PRESERVATION MANAGEMENT

*Overview*

• Describe the institution conducting the survey including its history and significant collections. Also include its mission.
• What are the staffing and professional levels? Which staff members are responsible for which collections? What percentage of their time is devoted to each of these?
• What is the institution's overall budget for all of its activities? What part of the budget is devoted to preservation activities? Is funding ongoing or one-time? Will cost-sharing assist in preservation activities?
• What is the long-term strategic vision and how does preservation fit into it?
• Does the institution have plans for expansion or renovation in the foreseeable future?
• Who are the partner organizations and how may they be categorized?

*Collections*

Describe the collection(s) being surveyed. For each category of material, estimate and use the unit of measurement that is most convenient (exact counts are not necessary).

| Type of material | File formats | Numbers of digital objects | Storage Media | Proprietary/ Encrypted | Number of files | Total volume |
|---|---|---|---|---|---|---|
| Full text | | | | | | |
| Image | | | | | | |
| Numeric | | | | | | |
| Sound | | | | | | |
| Moving Image | | | | | | |
| Multimedia | | | | | | |
| Other | | | | | | |

• What does the institution consider the most important areas of these collections?
• What types of formats or collections are prioritized for preservation?
• Do policies for selection and acceptance of digital objects exist? Who has chief responsibility for these digital assets and to whom do they belong? Are there format requirements? Is normalization to be used?
• Are re-appraisal guidelines available? Do all collections fit within the collection-development policy?
• What is the expected rate of growth for collections by media type, etc.? by type of donor?
• What are the types and levels of usage?
• Are systems in place to evaluate rights issues which may be barriers to preservation? Do appropriate workflows already exist? What licensing agreements for the digital content exist and what do they allow? Are there costs associated with securing these rights? Can they be sustained?

*Preservation Management Issues*
• Have preservation priorities been established? Is there a preservation plan?
• What preservation activities are already taking place? What strategies are being employed?
• What are the staffing levels devoted to preservation? Are preservation responsibilities spread across institutions and departments? Will any activities be outsourced? Have these tasks been delineated?
• How will preservation activities be managed? Do regular procedures and timetables exist?
• Does staff possess adequate preservation-related training? If not, is such training available?
• Is there an institutional commitment to preservation activities? Fiscally? Sustainable?

*Building Survey*
Use Patkus's survey and consider applying it to off-site storage areas as well.

*External Threat And Water Protection Worksheet*

*Fire Protection Worksheet*
Use these worksheets without change.

*Disaster Planning*
Use the questions outlined by Patkus and add the following:
• If using third-party services for off-site storage, can the institution be considered a "trusted digital" repository? Is it bonded?
• What is recovery turn-around time?
• How often are systems backed up? By whom?

*Security and Access Worksheet*
• What methods are currently in use to ensure authenticity and integrity? Checksums? Other methods? Is this validation information stored in the preservation metadata? What is the schedule for such verification?
• Is there an audit trail? Is the change history and technological context recorded?
• Is there write protection?
• How is virtual access protected?
*File formats*
• Are the file formats proprietary? Are they encrypted?
• Are the file formats well defined by file format viability services?[34]
• What versions are they?
• Is the format acceptable according to archive specifications? Do they fit in with best practices in the community?

*Media*
• Is media suitably durable and persistent?

---

[34] Some file format registries include PRONOM <http://www.nationalarchives.gov.uk/pronom> (accessed May 28, 2008), sponsored by the National Archives in the U.K.: the Global Digital Format Registry <http://hul.harvard.edu/formatregistry> (accessed May 28, 2008), and the Library of Congress's "Sustainability of Digital Formats: Planning for Library of Congress Collections" <http://www.digitalpreservation.gov/formats/index.shtml> (accessed May 28, 2008). The Florida Center of Library Automation's Digital Archive maintains a preferred format list <http://www.fcla.edu/digitalArchive/formatInfo.htm> (accessed May 28, 2008).

• Is media stored under appropriate environmental controls? In appropriate housing?
• Do policies for handling media exist? Are they followed?
• Is equipment clean and maintained?
• What is the general condition of the media?

*Creation of the digital objects*
• Were the files created following best practices and guidelines? Which set of guidelines?
• Who was responsible for the creation of the files?
• Was enough detail captured to warrant long-term retention?
• Were longevity issues considered during the course of creation?

*Metadata*
• What types of metadata are available for the digital library objects? Descriptive, technical, administrative, etc.? Does the metadata follow best practices and guidelines?
• Is there a metadata specification and agreed-upon implementation?
• Do the objects have unique, persistent identifiers? Locally? Globally? What type?
• Is metadata accessible through encapsulation[35] or by linking? Is it easy to identify, extract, and associate with digital objects? Is it extractable? Is it easily associable with the digital object?
• How is it managed?
• What metadata is included for preservation purposes?
• Is adequate information recorded?

*Strategies*
• What preservation strategies are currently employed? For what type of objects? Does documentation for these decisions exist?
• Is outsourcing an option?
• What are the significant properties of the objects? What must they retain for appropriate preservation?
• Is staff monitoring changes in the field to adapt to new preservation strategies?

---

[35] *Encapsulation* is the "wrapping" or "bundling" of a digital object with all the information or tools needed for its access. See "Encapsulation" in *Preserving Access to Digital Information*, <http://www.nla.gov.au/padi/topics/20.html> (accessed November 30, 2007).