



[CONTENTS](#)

[SEARCH](#)

[HOME](#)

Table of Contents

- **Feature Article**
 -  **[Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress](#) by Caroline R. Arms**
 - **Technical Feature**
 -  **[Risk Management of Digital Information: a File Format Investigation](#)**
 - **[Highlighted Web Site - TechFest](#)**
 - **[FAQ - SCSI, UDMA, USB, and Firewire Technology](#)**
 - **[Calendar of Events](#)**
 - **[Announcements](#)**
 - **[RLG News](#)**
 - **[Hotlinks Included in This Issue](#)**
-

Feature Article

Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress

by Caroline R. Arms, National Digital Library Program & Information Technology Services, Library of Congress

caar@loc.gov

Introduction

The National Digital Library Program (NDLP) provides remote public access to unique collections of Americana held by the Library of Congress through *American Memory*. During the 1990s, the program digitized materials from a wide variety of original sources, including pictorial and textual materials, audio,

video, maps, atlases, and sheet music. Monographs and serials comprise a very small proportion of the textual materials converted. The program's emphasis has been on enhancing access. No digitization has been conducted with the intent of replacing the original materials; indeed, great care is taken over handling fragile and unique resources and conservation steps often precede scanning. (1) Nevertheless, these digital resources represent a significant investment, and the Library is concerned that they continue to be usable over the long term.

The practices described here should not be seen as policies of the Library of Congress; nor are they suggested as best practices in any absolute sense. NDLP regards them as appropriate practices based on real experience, the nature and content of the originals, the primary purposes of the digitization, the state of technology, the availability of resources, the scale of the *American Memory* digital collection, and the goals of the program. They cover not just the storage of content and associated metadata, but also aspects of initial capture and quality review that support the long-term retention of content digitized from analog sources.

Institutional Context

The National Digital Library Program was established as a five-year program for fiscal years 1996-2000, building on the pilot *American Memory* project that ran from 1990 to 1995. Although managed as a special program, NDLP has worked closely with the divisions that have custodial responsibility for the materials it has digitized. NDLP established a core "production" team, but also funded positions within custodial divisions, recognizing that when the program came to an end or entered a new phase, responsibility for the digitized materials would likely fall to those units. In some divisions, NDLP staff positions have been used to jumpstart a mainstream digitization activity steered by the division. In others, the NDLP coordinators still largely drive the activity. Conversion of most materials has been carried out by contractors under multi-year contracts. Administrative responsibility for digital content is likely to remain aligned with custodianship of non-digital resources. The Library recognizes that digital information resources, whether born digital or converted from analog forms, should be acquired, used, and served alongside traditional resources in the same format or subject area. (2) Such responsibility will include ensuring that effective access is maintained to the digital content through *American Memory* and via the Library's main catalog and, in coordination with the units responsible for the technical infrastructure, planning migration to new technology when needed.

The computing support infrastructure at the Library of Congress is centralized. Responsibility for computer systems, disk- and tape-storage units, and all aspects of server operations lies with Information Technology Services (ITS). At the physical level (digital "bits" on disk or tape), responsibility for the online digital content created by NDLP (and all other units) falls to the ITS Systems Engineering Group. The Automation Planning and Liaison Office (APLO) coordinates most activities relating to automation and information technology in Library Services, the umbrella unit for all services related to the collections. Responsibility for coordinating technical standards related to digital resources has recently been assigned to the Network Development and MARC Standards Office (NDMSO). In practice, the coordination activities build heavily on the expertise of staff who have been involved in building *American Memory*.

The scale and heterogeneity of the collections and the Library's commitment to full public access provide a

challenge for ensuring ongoing access.

Scale

Table 1 provides a general picture of the volume of digital content generated for *American Memory* and the current rate of scanning through examples from selected custodial units. Two divisions, Geography and Map Division and the Prints and Photographs Division, dominate the total volume. The Law Library example comes from a single multi-year project with high production rate but lower space requirements for textual content.

Table 1: Examples of storage requirements for raw content for selected custodial units

Custodial Unit (selected)	Content converted by April 2000	Current rate of scanning	Notes on daily volume
Geography & Map (G&M) Division	.8 Tb	6 Gb/day	20 scans at ~300 Mb per scan. 24-bit color, 300 dpi, TIFF, uncompressed.
Prints & Photographs (P&P) Division	3.8 Tb	8 Gb/day	400 b&w negatives at ~20 Mb per scan. 8-bit grayscale, TIFF, uncompressed.
Law Library of Congress	65 Gb	50 Mb/day	400 scans at average of 125 Kb. 300 dpi, bitonal, Group 4 compression (lossless).

Mission to Serve the Nation

For materials identified as no longer protected by copyright, the Library strives to provide public networked access to the highest-resolution files except when individual files are too large for use on today's computers and transmission over the Internet. In practice, access is currently provided to all versions of pictorial materials other than maps, with the largest being around 60 Mb (for an uncompressed 24-bit color image about 5,000 pixels on the long side). The uncompressed TIFFs of maps (averaging 300 megabytes in size) are stored online but public access is not facilitated. Access within the Library to all files is intended to support future delivery on paper or offline formats such as CD-ROM when permitted by rights-holders or copyright law.

Heterogeneity

The Library wishes to establish institution-wide practices that will ensure long-term retention of digital resources from all sources, whether created by the Library or acquired for its collection by purchase or through copyright deposit, whether born-digital or converted from analog originals. Challenges to developing a comprehensive supporting system include the following:

- the variety of materials (in both original and digital form, extant today and expected in the future)
- the variety of routes and workflow patterns that may lead to ingestion of digital content into the Library's collection

- the variety of modes in which access and output may be required

In NDLP's experience, enforcement of standard practices in a program of this scale proves infeasible unless there are automated systems to support and facilitate those practices.

Conceptual Framework

In discussing preservation of digital materials at the Library of Congress, it has proved useful to use the conceptual framework of the "repertoire of preservation methods" shown in Table 2. This list of five methods was developed by Bill Arms at the Spring 1998 meeting of the Coalition for Networked Information (CNI) during a panel discussion on "Digital Preservation." (3) Don Waters, one of the other panelists and co-author of a seminal report (4) that introduced the distinction between refreshing and migration, later used the list (with minor variations) in many presentations in his role as Director of the Digital Library Federation. The annotations in the table were derived during brainstorming sessions at the Library of Congress during 1999.

Table 2: Repertoire of Five Preservation Methods

Better media	Longevity and technology-independence for storage media. May be more important for offline storage than online storage because of rate of change in computing technology. Library community reliant on industry to develop and deploy better media.
Refreshing bits	Includes any operation that simply involves copying a stream of bits from one location to another, whether the physical medium is the same or not. Any process that can be applied to any stream of bits regardless of characteristics of the information those bits represent is regarded as refreshing in this context.
Migration of content	Includes any transformation operation from one data representation (digital format) to another. Examples: conversion from one digital format for images (say TIFF) to a future standard format that provides enhanced functionality; upgrading or transformation from one SGML Document Type Definition to another. Migration to an enhanced format can improve functionality of a digital reproduction.
Emulation of technical environment	Using the power of a new generation of technology to function as if it were the technology of a previous generation. For materials digitized by the Library from analog sources, the aim should be to reduce the need for future emulation through effective choice of digital formats and planning for migration.

Digital archaeology

If all else fails... figuring out how to read stored bits and work out what they mean when systematic archiving has not been performed.

The five methods listed above can be seen as strategic methods. At an operational level, additional practices protect any digital resources: appropriate and secure environments for storage and computer systems; careful selection of storage system technology for reliability and data integrity; replication in all its variations (mirrors, routine incremental and full backups, duplicates); geographically separate storage areas for duplicates; procedures to validate refreshing actions and migrations; monitoring systems to validate integrity of files; and systems to support decision-making and transformation process management.

Refreshing can be carried out in a largely automated fashion on an ongoing basis. Migration, however, will require substantial resources, in a combination of processing time, out-sourced contracts, and staff time. Choice of appropriate formats for digital masters will defer the need for large-scale migration. Integrity checks and appropriate capture of metadata during the initial capture and production process will reduce the resource requirements for future migration steps. We can be certain that migration of content to new data formats will be necessary at some point. The future will see industrywide adoption of new data formats with functional advantages over current standards. However, it will be difficult to predict exactly which metadata will be useful to support migration, when migration of master formats will be needed, and the nature and extent of resource needs. Human experts will need to decide when to undertake migration and develop tools for each migration step.

The challenge of ensuring long-term access to digital material demands a new preservation model for archival institutions. In the past, preservation practice has typically involved a one-time reformatting step taken many years (usually at least fifty) after the creation of the original physical artifact. Effective preservation of resources in digital form requires (a) attention early in the life-cycle, at the moment of creation, publication, or acquisition and (b) ongoing management (with attendant costs) to ensure continuing usability. The Library's current practice for retention of locally digitized materials is based primarily on ongoing refreshing of bits. Many choices are made or steps taken during capture in the hope of reducing the need for migration, retaining metadata that will support effective migration when necessary, and avoiding the future need for emulation or archaeology.

Practices for Storage and Management of Digital Content

Responsibility for the digital materials at the physical level of files and bit streams lies with the Systems Engineering Group within Information Technology Services. All digital versions are stored online in a consistent logical structure on the primary storage system supporting the Library's central servers. This includes files of the highest resolution captured; typically 20-60 Mb for photographic reproductions and segments of early movies and 200-400 Mb for maps. Files delivered by vendors on other media are uploaded as soon as feasible. CDs delivered by contractors are labeled according to strict specifications and are organized and stored in archival boxes to serve as an additional level of backup once files are uploaded. However, the online files are considered as the masters.

Files on the central servers are protected by standard backup and restore procedures, providing replication,

and frequent refreshing. These operations can be performed unattended and show great economies of scale in administrative costs. ITS is establishing an Enterprise Storage Area Network, a network of storage devices independent of the Library's primary network based on TCP/IP. The separate storage network will allow backup, restore, and tape archive functions to be performed without degrading network performance for regular use. The Storage Area Network is currently based on a combination of SCSI and Fibre Channel links; a full transition to Fibre Channel is planned. All file-systems for the Library's servers are currently on magnetic disk, with built-in mirroring or RAID for reliability. From 1996 to 1998, a hierarchical storage management system (combining magnetic disk with tape storage) was used. However, it was determined that the administrative load and software shortcomings were not justified, given the falling price of storage and increasing data density for disk units of the same physical size. The Library acquired mass storage units occupying the same footprint in the computer center in 1995 (1 Tb raw capacity), 1996 (3 Tb each), 1998 (6 Tb each), and 1999 (14 Tb each). The 6Tb units in 1998 cost less than the 3 Tb units in 1996. Costs for processor capacity and storage media are expected to continue to drop (halving every 18 months at least, according to Moore's Law) for several years to come.

Full backups of the entire storage-system are made to magnetic tape cartridges in a robotic "library" on a two-week cycle; incremental backups (of any files that have been modified or created) are made each night. As the Enterprise Storage Area Network is expanded, two new capabilities are planned for further protection of digital data: snapshot archives of digital collections (on demand or on a schedule as determined by collection custodians); and an offsite location for an additional robotic backup facility that is accessible over the storage network.

ITS has also developed procedures for generating and comparing checksums for all files in a directory tree. (5) These procedures can be used for periodic checks against data corruption (inadvertent or malicious) and to verify copy operations. This procedure was introduced after a problem with the hierarchical storage management system had the effect of corrupting files in some directories. Those with administrative responsibility for the content decide which files to check and how often to check them. NDLP plans to introduce procedures that use checksums earlier in the production process (e.g., to verify correct transfer of files to delivery media by contractors and to verify correct uploading of files to the central storage system).

Storage and Management of Metadata

The National Digital Library Program has identified several categories of metadata needed to support access and management for digital content. Descriptive metadata supports discovery through search and browse functions. Structural metadata supports presentation of complex objects by representing relationships between components, such as sequences of images. In addition, administrative metadata is needed to support management tasks, such as access control, archiving, and migration. Individual metadata elements may support more than one function, but the categorization of elements by function has proved useful.

Since *American Memory* is aimed at access, the current system is designed to handle metadata that supports discovery and presentation of the resources. To facilitate indexing and to avoid the need to maintain two sets of catalog records, descriptive bibliographic metadata is stored separately from the digital reproductions. Structural metadata that relates page-images to parent documents or several related

photographs to a single bibliographic description for the group is stored with the digital reproductions in ancillary files. For such sequences, a "page-turning" dataset is prepared as a comma-separated ASCII file. (6) This dataset has a row for each page, indicating the sequence number, page number from the original (optional), and information to allow linking to all sizes/resolutions of the page and to a bibliographic record if available. For documents where only images of pages are produced, the datasets are typically generated automatically by scanning directories of image files. If transcriptions are produced and marked up in SGML, a special extract of the page information from the marked-up files is delivered by the contractors as a separate file accompanying the full file.

Other metadata elements are recorded in file headers or in documentation associated with the conversion of each collection. The need to build a more comprehensive system to store and manage digital collections for the long term has been recognized and several prototype activities have been undertaken by NDLP and more are planned within the Library. None has been wholly satisfactory; the process has illustrated the challenge of designing large-scale repositories that will support effective production, maintenance, access, and archival management for large-scale heterogeneous collections of digital materials. However, each exercise has improved understanding of the requirements and practical constraints. It has been recognized that metadata representations appropriate for manipulation and long-term retention may not always be appropriate for real-time delivery. For example, it has been necessary to "compile" the ASCII page-turning datasets into a more efficient binary structure for generating presentations dynamically. It has also been realized that metadata embedded in image-file headers may be too hard (processor-intensive) to get at to be useful in generating presentations. It would be helpful to have pixel dimensions for images (particularly for thumbnails) more easily accessible to presentation software. Similarly, sequencing information and page characteristics should be explicitly recorded as metadata elements and not merely embedded in filenames (although NDLP continues to find such semantics in filenames useful during production).

It has also been realized that some basic descriptive metadata (at the very least a title or brief description) should be associated with the structural and administrative metadata (even if full item-level MARC records exist in the Library's catalog or the items are pointed to from a finding aid marked up using the Encoded Archival Description DTD). For quality review and collection maintenance, it often proves useful to be able to identify which intellectual work (and possibly which physical original) the object represents.

The need to store and manipulate metadata of various types has led the Library of Congress to plan for a system that will allow the capture of metadata during production (interfacing with a variety of workflow patterns) and the export of selected metadata in formats appropriate to different functions. During 1999, an internal working group reviewed past experience and prototype exercises and compiled a [core set of metadata elements](#) that will serve the different functions identified. This set will be tested and refined as part of pilot activities during 2000.

Important aims of the current pilot activities in repository development are (1) to develop tools and procedures to capture core metadata during production in a more comprehensive, coordinated, and validated form than at present and (2) to import metadata already captured for the *American Memory* materials, for the benefit of the staff who will take on custodial responsibility in the future. The core set of elements is not intended to incorporate explicitly all technical metadata elements that might be needed to

support future migration of particular classes of material (particularly audiovisual materials) or all items of information that are recorded at capture time under certain contracts (e.g. pictorial scanning, where as many as 70 facts about a scanning operation and the derivation of the image versions delivered to the Library may be recorded for a single picture). It is envisaged that the system will incorporate other relevant information through links to supporting documents or data structures.

At the same time, planning is under way for a National Audio-Visual Conservation Center in Culpeper, Virginia, which will be used to house and process audiovisual materials (planned for occupancy in 2004). The Center is expected to incorporate a repository for audiovisual resources in digital form, for which [prototype development activities](#) are proceeding in parallel and in conjunction with more general repository development. During planning discussions, it has been hypothesized that the long-term archival/migration function will be best served by a data representation that binds certain metadata closely to the bits representing the content (encapsulation), but that the sheer size of the resulting files makes such a representation far too unwieldy to support access. (7) In other words, the form of an "archival digital object" may need to differ from the form of a digital object used to support access. This corresponds to the concept of an "access repository" as distinct from an "archival repository" described by Bernie Hurley in conjunction with the Making of America II project. (8)

Practices Associated with Capture and Quality Review

Certain NDLP production practices enhance preservability. The choice of formats for master versions of digital reproductions takes into account suitability for long-term retention. Resources are named in consistent ways designed to support persistent identification. Quality review procedures check data integrity as well as faithfulness of reproduction. Capture practices are documented for future reference.

Choice of digital formats for master versions of digital reproductions

Master formats are well documented and widely deployed, preferably formal standards and preferably non-proprietary. Such choices should minimize the need for future migration or ensure that appropriate and affordable tools for migration will be developed by the industry.

Consistent identification scheme and persistent identifiers

The Library has used a two-tiered naming approach for all *American Memory* materials. The digital object representing an item is identified by the "aggregate" to which it belongs and an item identifier that is unique within that aggregate. Aggregate names are registered centrally as part of the procedure for requesting storage space; this ensures that each two-part identifier is unique. These local identifiers can be used as the basis for identifiers that are globally unique and externally resolvable, for example using the Library's handle server (developed by the Corporation for National Research Initiatives).

Data integrity checks as part of quality review

Quality review procedures involve checking not only the faithfulness of a digital reproduction but also the technical integrity of the digital files that comprise it. Checks for technical integrity are important because effective automated migration will rely on consistency in its data source. NDLP bases its current practices on experience with particular contractors and workflow patterns.

Example 1: Integrity checks for page-images for text

NDLP commissioned development of a software utility to check TIFF and JPEG headers from one of its scanning contractors. NDView is used as a first pass review on receipt of a batch of files before they are reviewed for image quality. Each file is opened and values within the headers of image files are validated against a contract specification profile. Header elements that can be validated include pixel dimensions, resolution (in dpi), compression type, identifiers, and "Library of Congress" as creator. NDView also flags files that are unexpectedly small or large for the particular material.

Example 2: Integrity checks for text marked up in SGML

For text markup, 100 percent compliance is required on three important classes of integrity check. First, all SGML files must parse. Second, all supporting files expected must be present. Third, all identifiers and data elements that relate text files to the images for pages, tables, or illustrations (which have previously been checked for completeness) must be present and syntactically correct, including page-numbering and page-sequencing elements. A compliance rate of 99.95% is required for specified file-naming conventions and treatment of special characters.

A detailed list of quality control checks for marked-up text has been developed over several years. Early NDLP practice proved inadequate for fully automated transformation to an enhanced Document Type Definition because errors had not been caught during the initial production phase. Adequate checks for technical integrity will be an important prerequisite for effective migration.

Example 3: Integrity checks for images under pictorial contract

Since large numbers of similar images (e.g., black-and-white negatives) are captured under fully automated control by the same contractor, and experience has proved the contractor reliable, a sampling approach is used for checking both integrity and image quality. Files are delivered on CD-ROM in weekly batches of about 50 CDs. For initial quality review, at least one image is opened on every CD and every image is opened on two CDs per batch. A CD with three errors or more is returned for complete redelivery; otherwise, errors are corrected by delivery of individual files. Any other image files that do not open are discovered when derivative images for Web-delivery are generated.

Documentation of conversion process

Most materials are captured under the three multi-year contracts, with specifications for each project developed during a test phase. No formal, centralized system or location currently exists for holding this information, but details of equipment used and scanning/keying instructions are maintained in a paper trail. Summaries are provided online. (9) Some characteristics will be recorded as part of the core metadata set. Automated links to more detailed supporting documentation are envisioned.

Example 4: Targets to check calibration of scanning equipment

The current pictorial imaging contract requires that a target image be scanned with each batch of materials. Separate targets exist for reflective and transmissive capture. Each target contains gray patches (a version of a traditional photographer's step wedge) and sinusoidal tonal representation (dark and light zones in regular frequencies) whose reflective or transmissive densities can be plotted as a sine wave of varying amplitudes. The manufacturer individually calibrates the targets. The image densities of the gray patches

can be measured to determine the capture system's ability to render the range of tones desired. For this contract, the image files of the targets are measured only to verify that system integrity has been maintained from batch to batch. The files containing the target images are named to associate them with the production batch and delivered to the Library where they are stored for future reference.

Conclusion

Developing long-term strategies for preserving digital resources presents challenges associated with the uncertainties of technological change. There is currently little experience on which to base predictions of how often migration to new formats will be necessary or desirable or whether emulation will prove cost-effective for certain categories of resources. In the medium term, the National Digital Library Program is focusing on two operational approaches. First, steps are taken during conversion that are likely to make migration or emulation less costly when they are needed. Second, the bit streams generated by the conversion process are kept alive through replication and routine refreshing supported by integrity checks. The practices described here provide examples of how those steps are implemented to keep the content of *American Memory* alive. Technological advances, while sure to present new challenges, will also provide new solutions for preserving digital content.

Footnotes

- (1) Library of Congress, National Digital Library Program and the Conservation Division. Conservation Implications of Digitization Projects <<http://memory.loc.gov/ammem/techdocs/conservation.html>>.
- (2) Library of Congress. The Library of Congress Collection Policy for Electronic Resources, June 1999 <<http://lcweb.loc.gov/acq/devpol/electron.html>>. Includes the paragraph: "Selection of works for the collection depends on the subject of the item as defined by the collections policy statement for the subject of the work, regardless of its format. Formats include home pages, Web sites, or Internet sites required to support research in the subject covered. The Recommending Officer responsible for the subject, language, or geographic area of the electronic resources is responsible for recommending these materials. Electronic editions of audio-visual materials, prints, photographs, maps, or related items are also covered by the Collections Policy Statements for their appropriate formats."
- (3) Coalition for Networked Information. CNI Spring 1998 Task Force Meeting Final Report <http://www.cni.org/tfms/1998a.spring/final_report98Stf.html>.
- (4) Garrett, John and Don Waters. Preserving Digital Information: Final Report and Recommendations <<http://www.rlg.org/ArchTF/index.html>> The task force responsible for the report was sponsored by RLG and the Commission on Preservation and Access (before it merged into the Council on Library and Information Resources <<http://www.clir.org>>).
- (5) These checksums are digests or fingerprints of files that can be compared to detect changes. The checksum calculations can be performed on any set of bits and hence applied to any type of file. The MD5 message digest algorithm is currently used by ITS for this function.

- (6) This approach was documented in May 1998 for LC/Ameritech awardees at <<http://memory.loc.gov/ammem/award/docs/page-turning.html>>. Since then the same technique has been used to support presentation of groups of photographs in a grid of thumbnails.
- (7) Library of Congress. Preservation Issues <<http://lcweb.loc.gov/rr/mopic/avprot/rfq4.html>>, July 1999. Issued in relation to a limited competition to select a vendor to develop a prototype digital repository for the National Audio-Visual Conservation Center.
- (8) Hurley, Bernie. System Architecture Considerations for Digital Archival Repository Services <<http://sunsite.berkeley.edu/MOA2/papers/dump3.html>>, November 4, 1997.
- (9) Library of Congress. Building Digital Collections: Technical Information about *American Memory* Collections <<http://memory.loc.gov/ammem/techdocs/digcols.html>>.
-

Technical Feature

Risk Management of Digital Information: A File Format Investigation

The Council on Library and Information Resources (CLIR) sponsored a risk assessment study conducted by Cornell University Library (CUL) in 1999 that focused on the file format risks inherent in migration as a preservation strategy for digital materials. Project participants included: Gregory Lawrence (PI), William Kehoe, Anne R. Kenney, Oya Y. Rieger, and William Walters. The following summary has been prepared from the [full report](#) of this study, which will be available from CLIR in late June 2000.

Introduction

Digital information is produced in a wide variety of standard and proprietary formats, including ASCII and EBCDIC, common image formats, and as word processing, spreadsheet, and database documents. Each of these formats continues to evolve, becoming more complex as revised software versions add new features or functionality. It is not uncommon for software enhancements to "orphan," or leave unreadable, files generated by earlier versions. The most pressing problems confronting managers of digital collections are not unstable media or obsolete hardware, but data format and software obsolescence.

Migration is one of the major strategies for managing the later period of a digital lifecycle. [The Task Force on Archiving of Digital Information](#) described migration 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." With the exception of simple data streams, most files contain two basic components: structural elements and data elements. A file format represents the arrangement of the structural and data elements in a unique and specific manner. In this context, migration is the process of rearranging the original sequence of structural and data elements (the source format) to conform to another configuration (the target format).

In practice, migration is prone to generating errors. An obvious error occurs when the set of structural elements in the source format does not fully match the structural elements of the target format. For instance, in a spreadsheet file a structural element defines a cell containing a numeric value. If a comparable element is missing from the format specifications of the target format, data will be lost. A more subtle error might occur if the data itself does not convert properly. Floating point numbers (numbers with fractions) are found in many numeric files. Some formats might allow a floating-point number of 16 digits while others might only allow 8 digits. For certain applications, such as vector calculations in GIS programs, small but significant errors could creep into calculations. In other situations, migration might be able to preserve the content of the file, but may lose the internal relationships or context of the information. A spreadsheet file migrated to ASCII may save the current values of all the cells, but would lose any formulas embedded within those cells that are used to create those numbers.

Cornell's decision to investigate migration as a preservation strategy was partially determined by the resources at its disposal. With locally developed and commercial off-the-shelf data migration software available, migration could be tested, measured, and evaluated based on certain common criteria from which we could design a suite of risk assessment tools. In addition, file migration was appealing since it could encompass different preservation scenarios:

- Routine refreshing of digital files
- Changes in digital formats when files are converted from one application to another
- Changes in digital formats when files are converted from one format to another, such as the conversion of numeric files from proprietary formats to ASCII
- Creation of derivative access copies, e.g., the conversion of TIFF images into a PDF file

Risk Assessment as a Migration Analysis Method

In its present state, migration can be characterized as an uncertain process generating uncertain outcomes. One way to minimize the risk associated with such uncertainty is to develop a risk management scheme that deconstructs the migration process into discrete steps that can be described and quantified. A risk assessment is simply a means of structuring the process of analyzing risks. If the risk assessment methodology is well specified, different individuals, supplied with the same information about a digital file, should estimate similar risk values.

Cornell identified three major categories of risk that must be measured when considering migration as a digital preservation strategy:

- *Infrastructure risks* - including the presence (or lack) of institutional support, funding, system hardware and software, and the staff to manage the digital collections. All these, and the stakeholders who use the collections, will be affected to some degree by a migration of data. Legal and policy issues associated with digital information will introduce additional risks.
- *File format risks* - these will include the internal structural elements of the file that may be subject to modification during migration.
- *Conversion process risks* - the conversion software may or may not produce the intended result, in gross or very subtle ways.

Different methods or tools had to be developed to help quantify risk probability and impact in each of these risk categories. Over the course of the project, Cornell developed three assessment tools:

- A risk assessment workbook for the general collection. The workbook is intended to provide a broad review of risks associated with migration at the collection level.
- Reader software to examine specific files, or collections of files, for high-risk format elements.
- Test file for .wk1 format of known structural and data elements to test conversion software.

Individually, these three tools provide meaningful information. Used together, they provide a means to gauge institutional readiness to successfully migrate information from one format to another.

Risk Assessment of General Collections

In an ideal situation, the risk assessment would involve a team of experts, each a specialist in a specific area, who also would have a good understanding of digital preservation as a whole. In lieu of such a team, the Risk Assessment Workbook provides a systematic approach to assessing risks and problems inherent in file formats, technical infrastructure, metadata, and administrative procedures. It also moves the reader through the pros and cons of migration as a preservation strategy, and is designed to be used not only to identify potential risks, but to measure risk in terms of impact. Impact can be assessed along two scales: one measuring the probability a hazard would occur, and another measuring the impact of that occurrence. The Risk Assessment Workbook is included in the full report in Appendix A.

Assessing Risk in Conversion Software

Conversion process risks can be accomplished by examining a file before and after the migration process. A test file can be passed through the conversion software, migrating from source to target format. If the fields and field values of the original source file are properly reproduced in the target file, the risks incurred in migration are significantly reduced. If the fields or their values are not properly converted, the risk of migration is significantly increased. When the field tags and values in the test file are known, data changes associated with file conversion can be independently verified and characterized along a risk matrix.

To test this process, Cornell created a test file for the Lotus 1-2-3 ".wk1" format. Using public domain specifications and reference manuals published with the original application software, a large file was generated that exercised all the field tags and field values. This file was used to test various conversion software, and to compare results in the before and after states. This manual process was labor intensive (requiring three hours to complete) but quite accurate for the formats tests. Conversion of different structural elements and data elements is not always a complete "hit or miss." Some conversions resulted in target files that were almost, but not quite, identical to the source file. Based on this analysis, staff developed a rough scale of conversion risk (ranging from 1 - minor risk, to 5 - high risk) to characterize the impact of migration changes. Documentation for the test file is presented in Appendix B in the report.

Assessing Repeatable Risk Inherent in a Heterogeneous File Collection

Manual identification of risk associated with file structures is possible for a small number of files. For

large digital collections, manual methods are expensive and inefficient. To quantitatively measure the collection for files that contain at-risk elements, Cornell prepared a file reader program that can read structured ASCII and binary files. Given the project name Examiner, the program reads a file and detects the presence and frequency of specific file format elements. It then writes a report that identifies the file, its location in the collection and the type and number of at-risk elements associated with that file. The program also describes the risk level for the structural element based on the initial source/target analysis conducted on the test file. The program can be set to report at-risk tags only if the risk value equals or exceeds a certain threshold (e.g., values >2-5). The program does not read or evaluate the data value, although this feature could be implemented.

One strong feature of the Examiner program is that it is non-destructive. It simply reads a file from end to end and declares what is found. Also, Examiner can be set to read a single file, all the files in a directory, or all the files on a drive. The program is reasonably efficient and scans approximately 10,000 .wk1 files per hour. Finally, Examiner is written in Java, a modern programming language designed to be easily compiled on different operating systems. The program has been fully tested in the Unix and Windows 95/NT environments. General documentation for Examiner is described in the full report in Appendix C.

Assessing Risk Associated with the File Conversion Process

Finally, there are risks associated with different conversion software. The project examined two commercial off-the-shelf programs and quickly scanned the advertisements or published reviews of six others. In any mix of possible conversion programs available, each will provide some or all "core" functions, as well as optional features. General performance benchmarks, which can be tailored for specific migration scenarios, provide some uniformity of measurement and help highlight obvious defects. A Software Assessment Sheet was developed as a result of this analysis, which can be used to compare and assess potential conversion programs prior to initiating actual tests. This worksheet is included in the full report.

Case Studies

The risk assessment tools developed were tested on two digital collections at the Cornell University Library: the Ezra Cornell Papers and the USDA Economics and Statistics System. Each collection contains a dominant file format: TIFF or .wk1. The assessments of these two different collections are included as Appendixes D and E in the report.

Findings and Recommendations

Migration Risk Can Be Quantified

Migration, or the conversion of data from one format to another, has measurable risk, and the quantity of risk will vary, sometimes significantly, given the context of the migration project. One form of risk depends on the nature of the source and target formats. This project demonstrated that it is possible to compare formats in a number of ways and to identify the level of risk for different format attributes. The format analysis techniques and software may be technical, but the results can be described in general terms. Since basic file structure concepts are common to many file formats, experience with one format can be used to understand other formats. The greatest challenge is interpretation: when is a risk acceptable?

This study provides examples to illustrate the evaluation process. In practice, the risk assessment tools are not fully developed. Further refinement of these tools will provide more reliable results, but none will replace experience and good judgment.

Conversion Software Should Meet Basic Performance Requirements

Based on the review of conversion programs, Cornell determined that migration software should perform the following functions:

- Read the source file and analyze the differences between it and the target format
- Identify and report the degree of risk if a mismatch occurs
- Accurately convert the source file(s) to target specifications
- Work on single files and large collections
- Provide a record of its conversions for inclusion in the migration project documentation

The two programs fully analyzed in this project failed to achieve each of these basic criteria, although the results suggest that commercial conversion programs have the potential to meet them with further development. Considering the cost of writing conversion software for a wide range of file formats, a commercially developed solution for migration software will ultimately be cheaper and more flexible than locally developing conversion software. Cultural institutions should work with vendors to help them develop products that promote safer file migration.

Ready Access to Complete and Reliable File Format Specifications is Necessary

The most difficult aspect of this project was the acquisition of complete and reliable file format specifications. Format-specific information was difficult to acquire from a single source; ultimately, format information for this study was acquired from four general sources: software developers, public ftp archives, monographs, and Internet discussion lists.

Software developers of applications that utilize a specific proprietary file format should be the best source for file format information. This was not the case for Lotus .wk1 format information. Lotus, like other large software companies, treats file format information as a business product to sell to software developers. Lotus business products evolved, responding to revisions in 1-2-3 as well as changes in the DOS/Windows operating system. With the introduction of Windows 3.1, developer interest in earlier DOS specifications disappeared. Since the specifications for the .wk1 format were integrated into the format specifications for later releases (i.e., .wk3, .wk4), the specifications and documentation for the earlier .wk1 format quietly disappeared. Lotus as a company also evolved, and key members of the early development staff - often the corporate memory in software companies - moved on to establish their own companies. Lotus maintains an ftp archive that contains 1-2-3 .wk1 format specifications. These specifications have been described in great detail by outside parties, one of whom provides a sample .wk1 file analyzed byte by byte. Unfortunately, these specifications are incomplete and describe the .wks file format, the format of 1-2-3 release 1A. In the last months of this project, Cornell staff located Lotus employee who had been with the company since the middle 1980s. This individual provided a copy of Lotus File Formats for 1-2-3, Symphony, and Jazz. This work, authored by Lotus, is the only surviving documentation from the company for that period. Fortunately, it describes the .wk1 format in complete detail.

TIFF specifications are accessible from two Internet locations. The official specifications for TIFF 6.0 are available from the [Adobe developers support site](#). Adobe's site does not list the specifications for TIFF 4.0 and 5.0. These can be located at the [Unofficial TIFF Home Page](#). Manual examination of the specifications showed them to be consistent with each other, but they are incomplete. For years developers have been adding their own proprietary tags to the TIFF specification that they register with Adobe. Special tags do not appear in either the official or unofficial specifications. Several books on TIFF file format specifications survey large numbers of file formats, but no single work presents a clear, comprehensive description of either the TIFF file format specification or information about proprietary tags.

It seems likely that these difficulties extend to other formats. Conceptually, the solution is to adopt "open" format specifications, where complete, authoritative specifications are available for anyone to access and analyze. Cornell's experience with TIFF and .wk1 suggests that with file formats there are two specifications at work. One is the public document, which describes the basic or "core" elements of any format. The other is a private, non-standard set of file elements, usually developed to extend the original functionality of a file format. These private file elements provide the competitive edge for third-party software, and rarely are openly published. Over time, new format elements are often integrated into format revisions. For example, TIFF grew from 37 tags in version 4 to 74 tags in version 6. New proprietary tags for TIFF version 6 are registered with Adobe, which does not make them public. It is uncertain if all or some of these difficult-to-identify tags will be integrated into the anticipated TIFF version 7. Cornell endorses the concept of open specifications, but more thought must be directed at coordinating access to the more static, public domain specifications and the dynamic, non-standard elements. There is a real and pressing need to establish reliable, sustained repositories of file format specifications, documentation, implementation guides, and related software. Cornell recommends the establishment of such depositories as a prerequisite to the development of an effective national preservation strategy.

Highlighted Web Site

TechFest

This site provides an impressive collection of links on a wide range of topics related to networking and computer hardware. In the words of its creator, TechFest "is targeted towards professionals, researchers, educators, students, and hobbyists who need an initial reference that will help them find the detailed information they need." If you need to know more about some of the many acronyms and other technical terms that appear in computing system or networking specifications, this is a good place to start. TechFest's links for [bus and i/o \(input/output\) standards](#) cover all the standards discussed in this month's FAQ as well as many others. The site also provides a full, relevance-ranked internal searching capability.

FAQ

I'm used to using SCSI devices for mass storage, scanners and other peripherals, and have quite a substantial investment in them. Lately, though, I see more and more devices advertised as UDMA, USB, and Firewire. What are these and are they going to replace SCSI?

Before trying to answer, we need to decode some of the jargon in this question.

SCSI (Small Computer Systems Interface) is an old (1984) standard for connecting such things as storage devices and scanners to computers. The original SCSI specification has been updated many times and exists in many forms such as Wide SCSI, Fast SCSI, Ultra SCSI and various combinations thereof. SCSI has long been considered the choice for high-performance, reliable peripherals such as fast, magnetic disk arrays and high-resolution scanners. Its advantages include superior throughput (currently as high as 1280 Megabits/second), and multiple device support (up to 16 devices on one chain). The downside includes cost, complexity (device IDs and termination) and thick, bulky cables. SCSI is promoted by its own [trade association](#) and extensive information about SCSI can be found in its [FAQ](#).

UDMA (Ultra Direct Memory Access) is but one name for IDE (Integrated Drive Electronics) storage devices. Officially, these drives are known as ATA (AT Attachment) but they are also known under several variant names such as EIDE, Ultra-ATA and ATA/66. IDE drives dominate the consumer market because they are inexpensive and provide good performance for most home uses. IDE is designed only for hard disks, but a related standard, ATAPI (ATA Packet Interface) supports other storage devices such as CD and tape drives. IDE has limited expandability (normally, four devices maximum, all within the computer case) and still lags SCSI in terms of throughput and durability. IDE also puts more of a load on the CPU than other storage technologies. The IDE standard is managed by [Technical Committee T13](#) of the National Committee on Information Technology Standards. Detailed information about IDE can be found in its [FAQ](#).

USB (Universal Serial Bus) was developed by Intel to replace the myriad of ports used on PCs, and to handle every kind of peripheral, including storage devices, printers, monitors, modems, speakers, etc. Early adopters encountered problems with compatibility between devices, driver availability and operating system support. Introduction of the iMac (which sported only USB ports) helped spur development of USB peripherals, and Microsoft included full support of USB in Windows 98. Intel provides hardware support for USB in its chipset, and most PCs now come standard with at least one USB port. The current implementation of USB (v.1.1) supports up to 127 devices either by daisy-chaining or through multi-port hubs. Maximum available bandwidth is 12 Megabits/second, suitable for low-speed devices such as keyboards, mice and low-resolution scanners. USB devices can be plugged or unplugged without turning off the computer or rebooting (i.e. they're "hot-swappable") and are "self-declaring" so they become available immediately after being attached. Also, there's some power on the bus, so USB devices can be truly portable. USB uses thin, flexible cables with snap-in connectors. Large numbers of USB peripherals have appeared recently, and it is strongly supported by Intel, Microsoft, Hewlett-Packard, and many other major industry players who belong to the [USB Implementers Forum](#), whose Web site contains substantial

information about USB.

Firewire was originally conceived by Apple Computer in 1986 as a high-speed replacement for ADB (Apple Desktop Bus). However, it languished for many years, until Sony began producing digital camcorders with built-in Firewire ports (under the name i.Link) in 1995. Firewire became industry standard IEEE-1394 later in 1995 (hereafter "1394"). Apple started providing Firewire ports on its machines with the G3 desktop line. General computing support for 1394 has been slow in coming, though it is now widely used for digital video. Full Windows support finally came with Win98 SE but Intel still doesn't support 1394 in its chipset. The current implementation of 1394 (v.a) supports up to 63 devices. Maximum bandwidth is 400 Megabits/second, making it suitable for external mass storage devices and high-resolution scanners. Like USB, 1394 devices are hot-swappable, self-declaring, portable and use thin, flexible cables. Isochronous data transfer provides dedicated bandwidth for applications, like multimedia, that cannot be interrupted (USB does this, too). However, 1394 has features unavailable with USB. It supports "peer to peer" connections, so two 1394 devices can be directly interconnected, bypassing a computer. For example, a scanner or digital camera can send its output directly to a printer. Also, a 1394 device can be shared by multiple computers simultaneously. 1394 peripherals have started to appear in greater number and variety. Additional information about 1394 is available from the [1394 Trade Association](#) and [Apple](#).

Until now, each of these technologies has occupied a fairly distinct niche in the computing ecosystem, and none has been perceived as a great threat to the others. SCSI has catered primarily to the market for high-end scanner and storage devices for servers, IDE to the consumer market for high-speed storage devices, USB to the consumer market for low-speed devices and 1394 to the high-end consumer and professional market for specialized digital video devices. Don't look now, but this tranquil situation is about to be replaced by an all out battle for survival that may lead to the extinction of one or more of these technologies.

By the end of year 2000, devices supporting the next versions of USB (v.2.0) and 1394 (v.b) should become available. USB will get a major speed boost to 480 Megabits/second, while 1394 will jump to 800 Megabits/second. This puts the two standards on a relative par with each other, and with SCSI, in terms of their ability to support high-speed peripherals. Some pundits are predicting the gradual death of SCSI, citing limitations in the number of devices supported and poor ease of use compared to USB and 1394. SCSI supporters would differ, and have offered a road map showing continued improvements in SCSI's throughput during the next few years. Some predict that 1394 will eventually replace ATA for consumer hard drives. Members of the USB and 1394 communities, who until very recently talked about how the two technologies would peacefully co-exist and were "complementary," have now broken into predictable partisan factions.

Intel and its allies in the personal computer industry now see only a very limited role for 1394, with USB being used for virtually all peripherals. Contrarily, 1394 boosters claim that USB will not be able to keep up with increasing bandwidth demands, while pointing out that 1394b will eventually support up to 1.6 Gigabits/second throughput on plastic/glass fiber or shielded category 5 wire up to 100 meters in length.

What this feud really boils down to is what the future of personal computing will look like. USB backers are betting on the continued centrality of the PC, a condition that USB's basic design just happens to require. 1394 backers see a different future in which the network itself comes to dominate, and devices such as set top boxes, digital televisions, and camcorders are on a par with PCs, and capable of acting independently. Experiments with complete home networks based entirely on 1394 are already underway.

Which finally brings us back to the original question - should you start worrying about your current investment in SCSI peripherals? For the short term, the answer is no. As of now, SCSI continues to have an upper hand in terms of speed, reliability, and availability. But within the next two years, things will start to get a little fuzzy. Both USB and 1394 backers undoubtedly have designs on SCSI's slice of the high-end peripherals market. It is certainly not too early to start paying attention to this issue and asking your equipment suppliers and distributors about their intentions. However, it is inadvisable to rely on any one source, since unbiased information on this subject is extremely difficult to find. We'll keep you posted on developments.

Calendar of Events

Planning and Implementing a Digitisation Project

June 29, 2000, London, England

To be held at the British Library, this conference will address the main issues involved in digitization project design and management. There will be presentations by practitioners who have run successful imaging projects.

Preserving Photographs in a Digital World

August 12-17, 2000, Rochester, New York

Sponsored by the George Eastman House, the Rochester Institute of Technology, and the Image Permanence Institute, this week-long program will focus on traditional photo collection preservation techniques and on the basics of digital imaging.

Creating Electronic Texts and Images

August 20 - 25, 2000, Fredericton, New Brunswick, Canada

This is the fourth institute at the University of New Brunswick, and the course will focus on the creation of a set of electronic texts and digital images. Topics to be covered will include: SGML tagging and conversion; using the Text Encoding Initiative Guidelines; basics of archival imaging; and the preservation of electronic texts and images in the humanities.

DRH 2000: Digital Resources for the Humanities

September 10 - 13, 2000, Sheffield, England

The Digital Resources for the Humanities conference provides an international forum that brings together scholars, librarians, archivists, curators, and information scientists, to share ideas and information about the creation, exploitation, management, and preservation of digital resources in the arts and humanities.

Written in Light: Photographic Collections in a Digital Age

September 12-14, 2000, London, England

The Public Record Office will be hosting this conference as part of the project Safeguarding European Photographic Images for Access (SEPIA) funded under the European Union's Framework Programme in Support of Culture. The focus will be on issues that must be addressed to increase access to collections of photographic materials, while ensuring the preservation of those same materials for future generations.

The 8th Dublin Core Metadata Workshop

October 4-6, 2000, Ottawa, Canada

The Dublin Core Metadata Initiative will sponsor this workshop that will be held at the National Library of Canada. Participants are expected to be familiar with Dublin Core basics and should have expertise and interest in advancing the state of Dublin Core standards or deployment.

Announcements

New Imaging Publication Now Available

Moving Theory into Practice: Digital Imaging for Libraries and Archives (Anne R. Kenney and Oya Y. Rieger, editors and principal authors) focuses on an interdependent circle of considerations associated with digital imaging programs in cultural institutions -- from selection to access to preservation - with a heavy emphasis on the intersection of institutional objectives and practical digital applications. This new publication from RLG is a self-help reference for any institution that chooses to reformat cultural resources to digital image form. More information and an online order form can be found from North American and other world sites <http://www.rlg.org/preserv/mtip2000.html>, from UK Janet sites <http://www.rlg.ac.uk/preserv/mtip2000.html>, and from most European sites <http://www.ohio.rlg.org/preserv/mtip2000.html>.

CLIR and DLF Initiate Program for Distinguished Fellows

The Council on Library and Information Resources (CLIR) and the Digital Library Federation (DLF) are offering a new opportunity for librarians, archivists, information technologists, and scholars to pursue their professional development and research interests as Distinguished Fellows. The CLIR/DLF Program for Distinguished Fellows is open to individuals who have achieved a high level of professional distinction in their fields and are working in areas of interest to CLIR or the DLF.

The NEDLIB Report: An Experiment in using Emulation to Preserve Digital Publications

Authored by Jeff Rothenberg, this report presents the results of a preliminary investigation into the feasibility of using emulation as a means of preserving digital publications in accessible, authentic, and usable form within a deposit library.

New Reports from the Council on Library and Information Resources

CLIR has recently published the following, *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*, and *Authenticity in a Digital Environment*.

The Library of Congress National Digital Library Program (NDLP): Building Digital Collections: Technical Information and Background Papers

This part of the NDLP American Memory Web site documents technical activities relating to the procedures and practices employed by the NDLP over the past decade of digital library efforts. The page now includes direct links to the sections of American Memory framing materials which include RFPs/contracts, papers covering aspects of text mark-up, repository development, Web interface design, workflow and production (NDLP Writer's Handbook, quality review guidelines, project planning, the role of conservation) and also rights and restrictions statements.

RLG News

RLG Creates Alliance to Improve Access to Cultural Resources

Cultural materials include published and unpublished texts, images of many types, artifacts and other objects. Improving access to this range of "documents" is essential to the advancement of research and learning, especially as the definition of "data" expands in many disciplines. Better access is equally important to the sustained health of libraries, archives, museums and other cultural repositories. Although historians, cultural anthropologists, folklorists, historical archaeologists, historic preservationists, and a host of other researchers rely on cultural resources, there is currently no adequate capability for searching across the significant collections that reside in dispersed institutions around the world. Only a small amount of this kind of information is currently available in electronic form and building a sufficiently large resource to support this kind of inquiry is an imperative for many institutions. Teaching increasingly demands access to a large corpus of surrogates for cultural materials. Repositories increasingly seek revenue associated with the off-site use of collection surrogates. To address these issues, RLG member institutions are working together in the new [Cultural Materials Initiative](#).

Overview of the Initiative

An alliance of RLG institutions has agreed to work together to address the barriers and opportunities described above. RLG is committed to work with these institutions to develop consensus on consistent best practice in digital capture, resource description, and collective licensing. Together, RLG and the alliance participants will develop a collective digital resource, and an infrastructure to enable access to the collective resource around the world. A business model in which the use of the resource generates revenue to offset the continuing cost of digital capture and description, is a cornerstone of the plan.

The initiative is a multi-year project expected to unfold over a minimum three-year period. Planning began in mid-1999, supported by a grant from the Ford Foundation. In December 1999, Yale, Cornell and Oxford Universities, the New York State Department of Culture, the Chicago Historical Society and the International Institute of Social History (Amsterdam) agreed to play a lead role in the future development of the initiative. During the first half of 2000, an additional 30 institutions have joined in this effort. Working groups are being formed, and key elements of the service have been prototyped.

Cultural materials will be represented by digital surrogates in formats most effective for learning and research. In many cases these will be digital images accompanied by descriptive text; but other digital media types—for example audio, video, animation, or 3-D models—will also be used where appropriate.

The Alliance participants and the Alliance's policy advisory group have met to discuss issues related to content development and the development of standard approaches to the representation of three-dimensional artifacts and the mapping of descriptive information.

Development of Standard Approaches

Working groups populated by experts from Alliance participating institutions will develop standard approaches for:

Representation of Three-Dimensional Artifacts - Some of the most significant primary sources for the study of man and his works are physical objects: machinery, tools, musical instruments, furniture, clothing, tableware, weapons, scientific instruments and memorabilia, for example. RLG and the Alliance will examine existing technical presentation options used by other communities, identify the most promising technical solutions, and apply them to selected collections in participating institutions.

Mapping of Descriptive Information - The service will provide integrated access to materials from a broad range of institutions that use a number of different descriptive levels and standards such as Categories for the Description of Works of Art; Dublin Core; Visual Resources Association's Core Categories; Encoded Archival Description ; Machine-Readable Cataloging; Museum Educational Site Licensing; Record Export for Art and Cultural Heritage. Common access points will be provided through mapping, at the same time preserving forms of access that are particular to the originating discipline. It is anticipated that the CIDOC Conceptual Reference Model will prove especially valuable in this work.

This year's [RLG Forum](#) at the [American Libraries Association Chicago conference](#) will focus on the Cultural Materials Initiative. Presentations will include an overview of the goals and objectives of this project to build an integrated resource of digital surrogates of rare, and unique cultural materials. In addition, attendees will see a preview of a working interface to manage and display the disparate kinds of resources to be included in this service. The meeting is open to anyone and is not limited to ALA registrants.

For more information about the [Cultural Materials Initiative](#), contact [Anne Van Camp](#), Manager of Member Initiatives, RLG.

Hotlinks Included in This Issue

Feature Article

[American Memory](http://memory.loc.gov/): <http://memory.loc.gov/>

[core set of metadata elements](http://lcweb.loc.gov/standards/metadata.html): <http://lcweb.loc.gov/standards/metadata.html>

[prototype development activities](http://lcweb.loc.gov/rr/mopic/avprot/avprhome.html): <http://lcweb.loc.gov/rr/mopic/avprot/avprhome.html>

Technical Feature

[Adobe developers support site](http://partners.adobe.com/asn/developer/technotes.html): <http://partners.adobe.com/asn/developer/technotes.html>

[bus and i/o \(input/output\) standards](http://www.techfest.com/hardware/bus.htm): <http://www.techfest.com/hardware/bus.htm>

[full report](http://www.clir.org/pubs/reports/reports.html): <http://www.clir.org/pubs/reports/reports.html>

[The Task Force on Archiving of Digital Information](http://www.rlg.org/ArchTF/): <http://www.rlg.org/ArchTF/>

[Unofficial TIFF Home Page](http://home.earthlink.net/~ritter/tiff/): <http://home.earthlink.net/~ritter/tiff/>

Highlighted Web Site

[TechFest](http://www.techfest.com/): <http://www.techfest.com/>

FAQs

[1394 Trade Association](http://www.1394ta.org/): <http://www.1394ta.org/>

[Apple](http://www.apple.com/firewire/index.html): <http://www.apple.com/firewire/index.html>

[SCSI trade association](http://www.scsita.org/): <http://www.scsita.org/>

[SCSI trade association FAQ](http://www.scsifaq.org/): <http://www.scsifaq.org/>

[Technical Committee T13](http://www.t13.org/): <http://www.t13.org/>

[Technical Committee T13 FAQ](http://www.faqs.org/faqs/pc-hardware-faq/enhanced-IDE/): <http://www.faqs.org/faqs/pc-hardware-faq/enhanced-IDE/>

[USB Implementers Forum](http://www.usb.org/): <http://www.usb.org/>

Calendar of Events

[The 8th Dublin Core Metadata Workshop](http://www.ifla.org/udt/dc8/call.htm): <http://www.ifla.org/udt/dc8/call.htm>

[Creating Electronic Texts and Images](http://www.hil.unb.ca/Texts/SGML_course/Aug2000/): http://www.hil.unb.ca/Texts/SGML_course/Aug2000/

[DRH 2000: Digital Resources for the Humanities](http://www.shef.ac.uk/~drh2000/): <http://www.shef.ac.uk/~drh2000/>

[Planning and Implementing a Digitisation Project](http://strc.herts.ac.uk/it/heds/conf-details.html): <http://strc.herts.ac.uk/it/heds/conf-details.html>

[Preserving Photographs in a Digital World](http://www.rit.edu/~661www1/sub_pages/frameset2.html): http://www.rit.edu/~661www1/sub_pages/frameset2.html

[Written in Light: Photographic Collections in a Digital Age](http://www.knaw.nl/ecpa/sepia/events/conference1.html): <http://www.knaw.nl/ecpa/sepia/events/conference1.html>

Announcements

[CLIR and DLF Initiate Program for Distinguished Fellows](http://www.clir.org): <http://www.clir.org>

[The Library of Congress National Digital Library Program \(NDLP\): Building Digital Collections:](#)

[Technical Information and Background Papers](http://memory.loc.gov/ammem/ftpfiles.html): <http://memory.loc.gov/ammem/ftpfiles.html>

[The NEDLIB Report: An Experiment in using Emulation to Preserve Digital Publications](http://www.kb.nl/nedlib/results/emulationpreservationreport.pdf): <http://www.kb.nl/nedlib/results/emulationpreservationreport.pdf>

[New Imaging Publication Now Available](http://www.rlg.org/preserv/mtip2000.html): <http://www.rlg.org/preserv/mtip2000.html>

[New Reports from the Council on Library and Information Resources](http://www.clir.org/pubs/reports/reports.html): <http://www.clir.org/pubs/reports/reports.html>

RLG News

[American Libraries Association Chicago Conference](http://www.ala.org/events/ac2000/): <http://www.ala.org/events/ac2000/>

[Cultural Materials Initiative](http://www.rlg.org/culturalres/index.html): <http://www.rlg.org/culturalres/index.html>

[RLG Forum](http://www.rlg.org/events.html): <http://www.rlg.org/events.html>

Publishing Information

RLG DigiNews (ISSN 1093-5371) is a newsletter conceived by the members of the Research Libraries Group's PRESERV community. Funded in part by the Council on Library and Information Resources (CLIR), it is available internationally via the [RLG PRESERV](http://www.rlg.org/preserv/) Web site (<http://www.rlg.org/preserv/>). It will be published six times in 2000. Materials contained in *RLG DigiNews* are subject to copyright and other proprietary rights. Permission is hereby given for the material in *RLG DigiNews* to be used for research purposes or private study. RLG asks that you observe the following conditions: Please cite the individual author and *RLG DigiNews* (please cite URL of the article) when using the material; please contact Jennifer Hartzell at jlh@notes.rlg.org, RLG Corporate Communications, when citing *RLG DigiNews*.

Any use other than for research or private study of these materials requires prior written authorization from RLG, Inc. and/or the author of the article.

RLG DigiNews is produced for the Research Libraries Group, Inc. (RLG) by the staff of the Department of Preservation and Conservation, Cornell University Library. Co-Editors, Anne R. Kenney and Oya Y. Rieger; Production Editor, Barbara Berger Eden; Associate Editor, Robin Dale (RLG); Technical Researcher, Richard Entlich; Technical Assistant, Allen Quirk.

All links in this issue were confirmed accurate as of June 13, 2000.

Please send your comments and questions to preservation@cornell.edu .

[CONTENTS](#)

[SEARCH](#)

[HOME](#)

[Trademarks, Copyright, & Permissions](#)