

NEDCC Survey and Colloquium

Explore Digital Preservation Policies and Practices

By Tom Clareson
Program Director for New Initiatives, PALINET

The Northeast Document Conservation Center (NEDCC), with funding from a National Leadership Grant from the Institute of Museum and Library Services (IMLS), conducted an online survey in May 2005 on digital collection policies and practices as the first component of a project to develop a methodology for assessing the digital preservation readiness of cultural heritage institutions. While the findings clearly illustrate the growing presence of digitization in libraries, archives, and museums, there is a distinct lack of policy to deal with the preservation of these items once they are created. Utilizing the results of the e-mail survey, NEDCC convened a colloquium of digitization and digital preservation experts in July 2005 to discuss the digital preservation needs of cultural heritage institutions, and how to begin addressing those needs. The next step in this project, beginning in early 2006, will be to conduct on-site digital preservation readiness assessment at selected institutions to test a new digital preservation needs assessment methodology.

Survey Development and Methodology

The project, with cooperating partners including the American Institute for Conservation (AIC), the Center for Research Libraries (CRL), Heritage Preservation, and the Museum Computer Network (MCN), began in 2004. An Advisory Committee developed a prototype questionnaire, to gather data on a wide variety of topics related to digitization and digital preservation. Sections of the survey focused on topics including:

- Preservation readiness, with questions regarding traditional and digital preservation activities;
- Information technology infrastructure;
- Creation and acquisition of digital collections;
- Delivery methods for digital collections;
- Administration and management practices for digital collections;
- Rights and licensing issues for digital materials.

NEDCC engaged TDC, a Boston firm that performs marketing research and analysis, to format the survey for the Web, administer the survey instrument, and tabulate survey data. On April 25, 2005, the survey was posted on the Web, and a link was e-mailed to a total of 1350 individual e-mail addresses divided among the library, archive, and museum communities. Additionally, the survey link was posted to four popular e-mail listservs in the cultural heritage community:

- American Library Association – Preservation Administration Discussion Group (PADG)

- Conservation of Archive, Library, and Museum Materials (ConsDistList)
- Museum Listserv (Museum-L)
- Southeastern Museums Conference – Registrar’s Committee listserv

A reminder e-mail was sent to the individuals and posted on the listservs on May 6, 2005. The survey process ended on May 16, 2005 with 174 total responses received and 169 determined to be valid for a response rate of 12.5%, according to TDC data.

Survey Results: Institutional Data

The majority of the institutions answering the on-line survey instrument were college and university libraries (55 responses or 33.1%), archives (24 responses or 14.5%), art museums (15 responses or 9%) and public or municipal libraries (13 responses or 7.8%). Other cultural heritage institutions which answered the survey were anthropology/ethnology, natural history and science/technology museums; historical societies; and state libraries. In addition, there were 41 responses (or 24.7% of the total) which were categorized as “other” types of institutions; most of these were additional types of museums, including private and historic estate museums, historic site/museums, and regional library/museum/institutions. Preservation Officers, Archivists, Librarians, and Administrators were the majority of staff members who answered the survey.

The survey respondents included institutions with a wide variety of staff sizes and operating budgets. Many smaller institutions—especially archives—with 1-20 full-time equivalent (FTE) employees responded (64, or 38.8% of all responses to the survey). Those with 21-100 employees (42 or 25.5%) and 101-500 FTE (46 or 27.9%) were also well-represented, and this group mainly included college and university libraries. The very largest institutions polled (501 staff or greater) provided 13 responses or 7.9% of the total. Annual operating budgets of institutions responding to the survey ranged from 32 institutions (20%) with annual operating budgets of \$250,000 or below; 37 respondents (23%) with budgets ranging from \$250,000-\$1,000,000, and the majority of the respondents (66, or 40%) with budgets of \$1,000,000-\$20,000,000. In addition, there were 28 institutions (17%) which reported budgets of \$20,000,000 and above—most of these were college and university libraries.

A few questions dealt with overall preservation practices, including analog and digital materials. Most interesting were the results of questions dealing with written institutional policies, plans, or procedures. A strong majority of responding institutions had, or was developing, documents including Mission and Goals, Collection Development, and Emergency Preparedness policies. The traditional preservation activities most likely to be carried out by responding institutions were environmental monitoring, item-level condition assessment of objects, and emergency preparedness activities. By far the weakest area of policy development among respondents was in the crafting of written plans and procedures focused on the creation of digital resources. Only 29% of respondents had such a policy. Archives, as a group, were especially weak in development of this type of policy; institutions with larger budgets were leaders in this type of policy development. A positive sign was that policy development for digital

resource creation was the area in which the largest group of respondents by far, 41% said a policy was being developed.

In findings similar to those found in recent statewide preservation needs assessment surveys, most respondents (102 or 63%) said that 5% of their budget or less was devoted to any type of preservation activities. There were 14 institutions, or 9% of the response pool, which had no funds whatsoever allocated for preservation activities.

Information Technology Issues and Infrastructure

Information Technology (I.T.) staff are key in the success of cultural heritage digitization projects. Almost all of the respondents have information technology services provided at their institution, including such activities as network support, desktop support, security and protocols, back-up and disaster recovery, centralized hardware and software acquisition and maintenance, and file management and storage. The survey asked if institutions supported information technology applications for digital collection management, and a clear majority of respondents were active in digital imaging, web design and development, and building collection information management databases. Over half of the institutions polled provided the public with the capability to search the institution's collection database online.

The management of information technology services and policymaking at the respondents' institutions provided some surprises. While 115 (or 68.9%) respondents had an I.T. Department, 52 or 31.1% did not. In those facilities with I.T. as a separate department, staffing levels were relatively low, with 139 (or 82.2%) having 0-4 FTE assigned responsibility for some information technology activities, and 29 or 17.2% with 4-8 staff members. In 71 institutions (43.8% of respondents) there was not a specific person in the institution with primary responsibility for information policy. This was especially true in the archival community. Information policy was a full-time responsibility of a staff member at 52 institutions (32.1%) and part-time at 39 facilities (24.1%). In the majority of cases where there was a staff person with information policymaking responsibility, it was the Library Director/Librarian (18 or 20% of respondents); Director, Vice President, or Manager of Information Technology (11 or 12.2%); or Information Systems Manager (also 12.2% of respondents). A larger number of responding institutions had a staff member with full- or part-time primary responsibility for systems administration and/or database administration.

A key finding in this area is that, while 36 respondents (24%) allot over 10% of their annual budget to information technology, there are a number of cases (25 responses or 17% of the total) where no or very low funding levels exist for I.T. support. Historical society and archives respondents were among those groups with a high percentage indicating no funds expended in this area. A positive trend was that 44 percent of respondents (72 cases) noted that their institution's I.T. budget has increased over the past five years, while 28.4% said this budget area has remained stable. It is likely that some of this funding went to systematic updating of software, which 142 respondents (86.6%) said they did, and/or for systematic upgrading of hardware (132 or 80%). An

interesting finding in this area is that mid-sized institutions – those with budgets between \$1-20,000,000 – were less apt to upgrade their hardware and software routinely.

Digital Collection Creation and Acquisition

A vast majority of the respondents to the online survey (151 respondents or 92%) had created digital assets from physical source materials. This portion of the survey instrument looked at the types of digital collections currently being created and/or acquired.

The variety of formats being digitized was wide-ranging, from flat works on paper/photographic prints, which 142 or 82% of the respondents had digitized, to books and other multi-page items, analog audio and video, three-dimensional objects and motion picture film, which were other top-rated answers in a question that asked respondents to select all of the answer categories that applied to their collections.

If so, from which of the following types of source materials (Select ALL that apply)		
Flat works on paper/photographic prints	142	84%
Books and other multi-page items	83	49%
Analog audio and video	67	40%
Three-dimensional objects	66	39%
Film	53	31%
Microfilm	39	23%
Other	12	7%

In addition, the vast majority of respondents stated that their institutions were already engaged in collecting, acquiring, or creating digital assets, most of which were in the categories of simple text; photography and other still image materials; and music, spoken word, or other audio materials.

One of the key findings of this survey was that 59 respondents or 39% said that the majority of the items they consider to belong to digital collections are unique, single-copy works. As one might expect, archival respondents reported a majority of these unique materials. This finding supports the good practice of acting to preserve both the digital and original copies of works which are thought to be one-of-a-kind.

Of all the items you consider to belong to digital collections, are the majority (Select ONE):

Unique (single copy works)	59	39%
Replicated in both digital and analog versions	43	29%
Replicated in other analog versions	28	18%
Replicated in multiple digital versions	21	14%

In the area of knowledge about formats held, a majority of responding institutions knew what type of file formats (TIFF, JPEG, etc.) existed in their collections, and most said they held three or fewer formats of original items, including text, encoded text, still and moving images, sound, and geo-spatial materials.

Digital Collection Delivery Methods

A somewhat surprising result to the question of who may access digital collections was that internal staff (147 respondents, or 87%) and on-site visitors (99 or 58.6%) were the leading responses. At 92 (54.4%) of the institutions, there was general availability of digital collections via the Internet or another network to users worldwide. Additionally, 61 (36.1%) of the institutions polled offered controlled availability, via the Internet or another network, to selected external users. The control of access to collections was particularly evident among archival respondents. The leading method of providing access to digital collections was through a website associated with the institution (113 or 66.9%); ROM or other removable media, and stand-alone computers were utilized by about one third of the responding institutions. Evidence of collaborative digital activities was also discovered in the answers to this question, as 41 institutions, or nearly one-quarter of the responding organizations, offered at least a portion of their digital collections through a website associated with another organization.

Another positive finding of the survey was that a large number of the respondents (135 or 83%) have created descriptive metadata for their digital assets in order to facilitate discovery and use of their digital collections. Later in the survey, respondents were asked to enumerate the types of metadata they create to manage their assets, and descriptive metadata was again the leading result. Other types of metadata that 50% or more of the respondents said they were creating include technical and administrative metadata. Metadata, which plays a key role in the identification of digital items, was rated, on a five-point scale, as the most important goal of the institution's digital program by 98 or 61.3% of facilities polled. The next important goal was study and use by local users (70 or 44% of respondents), which far outweighed the importance placed on study and use by remote users, which ranked 5th, identified as most important by 44, or 27.7% of those responding to the survey. And, although it is always a discussion topic in the digitization community, generating revenue was seen as the least important goal by the respondents.

Finally, the survey results indicated that responding institutions used a mix of methods to manage their metadata, including integrated databases, stand-alone databases, file-

directory conventions, and file structure. No single methodology was a clear leader in this area.

Administration and Management of Digital Collections

No or low levels of institutional funds are allocated for creation, acquisition, management or sustainability of digital collections in a high number of locations. While 25% of the institutions responding do not assign any portion of their budget to create digital collections, an even greater number (almost 42%) do not have budget lines for acquiring digital collections. The lack of budget for acquisition and maintenance of digital materials was most clearly evident among the archive, public library, and ethnological/anthropological museum respondents. Most of the institutions that are directing some funding into digital activities allocate only between 1-5% of their institutional budget for these activities. For example, a majority of college and university libraries allot less than 1% of their budgets to digital content creation.

Digital Collections: Administration and Management					
	No funds allocated	Less than 1% (but not zero)	1% to 5%	6% to 10%	Over 10%
If your institution <u>creates</u> digital collections, approximately what percentage of your institution's annual budget is allotted to this activity?	39 25%	58 37%	41 27%	12 8%	5 3%
If your institution <u>acquires</u> digital collections, approximately what percentage of your institution's annual budget is allotted to this activity?	62 42%	33 22%	33 22%	11 8%	9 6%
Approximately what percentage of your institution's annual budget is allotted to manage and sustain digital collections you have created or acquired?	46 30.3%	46 30.3%	47 30.9%	8 5.3%	5 3.3%

According to respondents, current or future funding specifically targeted toward digital preservation will be derived from institutions' regular budgets (43 respondents or 25%) through grant funding (37 respondents or 22%) or in an approach that is a hybrid of grants and regular budget funds.

In the area of staff responsibility/primary activity for digital preservation, 96 respondents (60%) did not have a specific person in the institution responsible for this role. In the cases where there was a staff person with responsibility for maintaining digital

collections, it was more often preservation staff (104 or 64.6%) than I.T. staff (74 or 47.4%) with this set of tasks as part of their activities.

A bright spot in the survey results was that 136 institutions, or almost 84% of those surveyed, supported staff development and professional education and training in the area of digital preservation. However, it seems as though this commitment to education has not translated into policy development as yet. An indication of the shortfall in digital collections management is that a pronounced minority of written policies for activities such as collection development, emergency preparedness, preservation, and especially exhibitions, specifically address the institutions' digital holdings. The lack of this type of policies was most noticeable in the art museum and archival communities.

With regard to digital preservation, participants were asked to select all strategies their institution had discussed implementing. Answers included regular data backup (utilized by a vast majority of the respondents), migration, and refreshing the data. Other strategies used are the maintenance of legacy equipment (hardware to read obsolete or less-often-held formats such as floppy disks), outsourcing digital preservation to an externally-managed repository, and emulation. The types of media used for storing digital collections are online magnetic media such as network hard drives (132 or 78%), or removable magnetic media (109 or 65%). Digital collections are most often stored in-house in systems managed by the institution, or by a combination of in-house, partner organization, and storage vendor (outsourced) methods.

In two other strategic areas, there is cause for concern. Eighteen respondents (11%) are not utilizing a backup strategy at all, and 30 (19%) back materials up once, which means that 30% of collections are not adequately protected by a backup strategy. Also, when asked if their digital assets are insured, 84 respondents (52.8%) said no, 58 (36.5%) did not know, and only 17 institutions, or just over 10%, said they insured these holdings. Clearly, these are areas for further digital policy and practice development.

The majority of respondents noted that the maximum lifespan required to both maintain digital materials and to retain their deliverability and usability so the materials can serve their intended purposes was 25 years or more, which was the longest option offered by the questionnaire.

Finally, the survey touched on some specific rights and licensing issues. A majority of respondents consider copyright and intellectual property concerns in their digitization activities. Most also attempt to acquire digital rights to materials when they are acquiring or collecting digital materials.

Conclusions from the Survey

Clearly, digitization is fast becoming a routine activity in many museums, libraries, and archives. More than 90% of the responding institutions are creating digital assets from physical source materials; over 88% are either collecting, acquiring, or creating digital assets. In addition, institutions are considering digital preservation issues – supporting

staff development and education on the topic, and establishing 25 years or more as the targeted maximum lifespan for deliverability, use and maintenance of digital materials. A majority is creating metadata for their digital collections, and is taking copyright issues into consideration in their creation workflow.

However, the creation of policies addressing the preservation and management of digital assets are lagging far behind other areas of policy development in cultural heritage collections. Only 29% of the institutions responding to the survey have policy, planning, or procedure documents on the creation of digital resources, although 41% say policies are in development. A gap between digitization and digital preservation practice is further suggested by the fact that, except for inclusion in rights and licensing policies, digital holdings are not included in the majority of policy statements for many areas of institutional operation, from Mission and Goals to Emergency Preparedness, to Exhibit Policies.

Moving Forward

NEDCC convened a colloquium of fourteen digital preservation experts in Boston on July 11-12, 2005, who addressed this lack of policymaking for digital preservation. The group reviewed the e-mail survey responses, and discussed digital preservation needs and proposed solutions. The group considered preliminary data from the Heritage Preservation “Heritage Health Index,” which looks at the preservation needs of all formats of cultural heritage collections. It reviewed a variety of existing assessment models, especially the IMLS-funded Conservation Assessment Program (administered by Heritage Preservation) which enables museums to work with outside conservation consultants to review their preservation needs and write a report identifying priorities.

The key conclusion from the colloquium was that small and medium-sized institutions will need the assistance of experts to assess the preservation status and needs of their expanding digital collections. The group suggested development of expert-facilitated on-site assessments, supported by an online assessment tool on digital preservation. The participants discussed methods to expand this type of technical assistance into a national program.

As a result of the colloquium discussions, a subgroup of the colloquium attendees, along with experts from the collaborative digitization arena, are refining the existing questionnaire on digital readiness to use as a diagnostic instrument, and testing it through a limited number of site visits conducted by expert teams. The first of these site visits will take place in January 2006, and information about the effort will be released to the cultural heritage community as it becomes available.

Major Conclusions of the Survey and Colloquium

The survey data indicated widespread needs for improved planning for digital readiness:

- The vast majority of institutions (92%) are creating digital assets from original source materials.
- 71% of institutions, however, have no written plans or procedures for the creation or management of digital resources.

Given that 92% of institutions are creating digital assets, and 71% have no written digital management plans, the colloquium recommended:

- Additional training to make institutions more aware of the fragility of digital assets and the need to be concerned about preserving them.
- Helping smaller institutions determine the internal guidelines they should follow for managing digital files, in the absence of a cohesive body of best practices.
- Focusing on onsite, expert-facilitated assessments of digital preservation needs, with written recommendations, as the most promising means of providing technical assistance on digital preservation readiness to smaller institutions.

NEDCC is grateful to the members of the Colloquium Committee for their expertise and dedication to this project.

Participants in the colloquium were:

Steve Chapman, Harvard University
 Paul Conway, Duke University
 Kevin Glick, Yale University
 Ken Hama, J. Paul Getty Trust
 Kristen Laise, Heritage Preservation
 Sam Quigley, Harvard University, MCN
 Bernard Reilly, CRL

Tom Clareson, OCLC
 Steve Dalton, Boston College
 David Green, Knowledge Culture
 Peter Hirtle, Cornell University
 Paul Messier, Conservator
 Rebecca Hatcher, NEDCC
 Ann Russell, NEDCC