

Implementing a Digital Imaging and Archiving Program: Technology Meets Reality

Oya Y. Rieger
Cornell University Library
Ithaca, New York, USA

Abstract

Cornell University Library's Digital Consulting and Production Services (DCAPS) was established in January 2003 to provide a framework for the image creation and management process, including requirements analysis, implementation, assessment, and archiving. Its service infrastructure is considered one of the important elements of the Library's institutional preservation policy for digital image collections. DCAPS aims to meet the needs of a diverse university community with varying needs and expectations. The service framework is based on a business model that brings efficiencies and accountability to the Library's digital content creation operation. Although a systematic and integrated approach to digital collection creation is an effective strategy, it also highlights the challenges in putting digital imaging and archiving standards into practice in a heterogeneous production environment.

Introduction

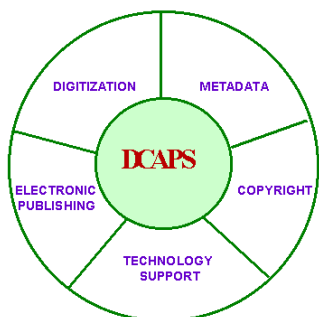
Similar to many other cultural institutions, Cornell University Library embarked on the creation of digital collections in the early 1990s. An important lesson learned during the early implementation stage was that long-term management of digital collections requires a substantial commitment of institutional resources. After a decade of experimentation, the Library felt the need to shift from a project mode to a program approach. A key impetus to this move was the recognition of the value of digital image collections as important assets that should be incorporated into the overall organizational goals, strategies, and systems. The following principles are important components of an institutional preservation policy to assess and manage preservation-related risks.

- Establish digital preservation as an institutional responsibility with committed financial and staff support
- Select and create digital content with long-lasting value
- Store digital media with care, and implement preservation strategies such as refreshing and migration
- Assess risks by monitoring technological changes

DCAPS tries to address these issues while focusing on digital image collections created at the Library. However, the main challenge is implementing such principles in a fast-paced environment with a variety of small- and large-scale projects determined by varying user needs.

Goals of DCAPS

Digital Consulting and Production Services (DCAPS) was established in January 2003 to address the need to develop an organizational infrastructure to ensure sustainability of digital collections. It is comprised of associated services necessary to ensure cost-effective planning, creation, management, use, and preservation of digital collections (Figure 1). It offers a suite of digital asset management services supporting content development from feasibility assessment to full-scale production, while it provides a single point of entry to distributed services because the components of the model are based on collaborations with various Cornell University Library service providers. It supports the effective and efficient creation of collections with a common technical architecture. The DCAPS management team is involved in decision making for several technical and organizational issues, including high-quality capture, indexing and retrieval, digital-image archiving, standards implementation, digital collection delivery-platform selection, user needs and outcomes assessment, and storage requirements. The integrated service model allows the staff to be informed about many stages and steps involved in the life cycle of digital content, in addition to promoting financial accountability. The unit is built on a holistic approach that recognizes the interdependency among the processes and resources involved in creating digital image collections. Managing image and metadata standards, user expectations, and resource requirements is a balancing act that demands a solid business model.



DCAPS: Digital Consulting & Production Services

Figure 1. DCAPS Units



DCAPS: Digital Consulting & Production Services

Figure 2. DCAPS Project Assessment and Management Services

DCAPS Users

The target audience includes Cornell library staff, faculty members, instructors, and researchers. The goal is to provide customers with a comprehensive suite of digital services to ensure the best investment of funds in creating digital collections. The assistance offered by DCAPS forms a continuum from short-term guidance and consultation to more-involved actual production and delivery of digital products. Consultancy levels include project or proposal evaluation; recommendations for database and system standards, models, and structures; suggestions for software, hardware, information standards, etc.; and advice for planning or implementation stages of system-development projects. Although the initial focus was on understanding and meeting the needs of the Cornell University community, the service model is currently being expanded to offer services for external cultural institutions. The unit is involved not only in production and archiving, but also in feasibility assessment and decision making (Figure 2). The DCAPS specialists assist library staff, faculty, and the members of the Cornell community in exploring digitization goals, funding sources, and methods. The unit is in charge of both managing the in-house production operation and outsourcing when that is a more-efficient method. An important role of the unit is to act as a liaison with Cornell groups as needed. For example, the Library's conservation or preservation specialists are consulted when the materials that will be digitized are rare or special. When digital content is created to support learning, the University's instructional technologies are brought in to assist in pedagogical issues.

As shown in Figure 1, DCAPS is composed of five distinct units that closely collaborate to reach a common set of goals. The rest of this article will focus on the imaging component of the unit. More information about the other DCAPS components can be found at <http://dcaps.library.cornell.edu>.

Digital Imaging Unit

The Digital Imaging Unit is composed of a part-time director, four imaging technicians, a structuring specialist, a part-time business manager, and several student assistants. The unit operates a variety of digitization equipment, including digital cameras and a flatbed and overhead book scanner, in addition to some equipment with micro-format digitization capacity. The main goals of the unit include:

- Identifying technical requirements for conversion, including capture method, resolution, bit depth, image processing, file format, compression, and image-quality inspection.
- Preparing documents for digitization and participating in collation and tagging of pages to identify special imaging and metadata requirements
- Imaging—high-quality or consumer-grade digital camera capture, bound-volume digitization, flatbed scanning, etc.
- Image enhancement and color management, including image sharpening, correction, de-skewing, reduction of curvature for bound-volume scans, etc.
- File naming and directory structuring depending on the purpose of digitization and digital collection delivery platform used.
- Creating structural metadata to support the access requirements of the collection, including tagging of bibliographic or structural units of content (articles, indexes, illustrations, bibliographies, etc.)
- Implementing standards to match project requirements, e.g., file naming, persistent identifiers, structural metadata recorded in a specific framework, etc.
- Collecting and recording technical metadata to document how images were created and stored.

Technology Meets Reality

During the first year of implementation, DCAPS has proven itself to be an effective approach to increasing capacity and accountability. Offering a one-stop-shopping unit with a suite of complementary services attracts a wide range of projects -- from digitizing materials for a course Web site to creating a collection of 120,000 images from one of the Library's unique collections to share it on the Web with the scholars of the world. The diverse production environment is an inevitable consequence of the changing academic models and integration of information technologies into learning, teaching, and research. The challenge for the Library is to implement a homogenous set of best practices and standards in such a production environment. It is undisputable that institutional strategies that support the creation of homogenous collections are ideal. However, the new academic models and demands on the digitization activities of cultural institutions may require built-in flexibility to accommodate an assortment of collections. Focusing on the digital-imaging component of the unit, this section will illustrate the challenge with two examples.

Image Quality and Preservation

Cornell University Library has been an advocate of creating complete digital master image files that reflect all significant informational content of the original source for preservation, access, and economic reasons.¹ Although this continues to be one of the guiding principles in DCAPS, moving from project-based imaging for preservation to a wide range of initiatives to support learning, teaching, and research has motivated the staff to be more flexible. The institutional guidelines for selection, digitization, metadata creation, and storage of digital images were prepared in 2001 through an Institute of Museum and Library Services (IMLS) grant.² The goal then was to support the creation of homogenous collections on the assumption that the preservation of collections via consistent techniques is more practical and cost-effective in the long run.

The DCAPS digitization unit is facing the challenge of being a part of the Library's preservation strategy in a hybrid production operation. Some digital image collections are created with a long-term commitment to maintaining them to support scholarship. On the other hand, some collections are created to support pedagogical needs that may not require high imaging standards. For example, imaging requests from faculty are on the rise, whether it is to create digital content for Web sites or to share their research results with their colleagues. One good example is the pilot implementation of DSpace at Cornell. DSpace is a digital library system to store, organize, preserve, and distribute the intellectual product of faculty in digital formats.³ The preservation issues in regard to content deposited in DSpace have not been fully addressed yet. It will certainly be tricky to set and enforce standards for faculty-created digital content. The key challenge is to balance quality and demand cost-effectively and to create digital collections that are worthy of

preservation. As the projects diversify, the staff feels an increasing need to revise the guidelines created in 2001.

Preservation Metadata and Z39.87 Implementation

One of the key requirements of a preservation policy is to have a framework for collecting and recording technical metadata to safeguard information that is essential in the face of changing technologies. Many argue that the continued viability of the digital image collections depends heavily on the availability of information on the technical characteristics of collections, technological dependencies, change history, and rights management. Preservation metadata encompasses a range of information that is required for the short- and long-term management of digital image files. This category of metadata includes both micro-information that describes the technical specifications of an image collection (resolution, bit depth, file format, etc.) as well as administrative information (copyright and rights-management information, storage and backup requirements, types of metadata generated, standards implemented, etc.) that will support future preservation decision-making and action. Technical metadata serves several purposes. In a managerial context, it supports image-quality assessment and image enhancement and processing, and facilitates workflow management. Although there is limited evidence at this point, technical metadata is also seen as an important source for long-term collection management.

It is easy to convince people about the importance of collecting and storing preservation metadata. Nevertheless, the task becomes quite a complicated mission in a hybrid production environment such as the DCAPS digitization unit. The unit needs to comply with the Library's overall preservation metadata practices in addition to adapting these guidelines to the specific genre at hand, image collections. In 2003 the Library's digital preservation officer and the head of the Metadata Unit were charged with examining the OCLC-RLG Working Group on Preservation Metadata⁴ guidelines to decide which Open Archival Information System fields to implement at Cornell. The OCLC-RLG document includes a comprehensive metadata framework describing the information necessary to carry out, document, and evaluate digital preservation processes. After general digital content guidelines were established, a brainstorming session was held to test the decision on a specific format—digital image collections. The focus of the discussion was not only on the data elements and their roles but also on the practical aspects of implementing preservation metadata in the DCAPS digital imaging unit. It is certainly challenging to try to map standards in a real-life production environment. The preservation metadata team continues to investigate ways to integrate preservation metadata into the production cycle. The essential component of a successful implementation is seamless integrating of metadata harvesting into the production environment.

Currently, the unit is assessing how to implement NISO Z39.87-2002, Data Dictionary—Technical Metadata for Digital Still Images.⁵ The standard (at this time a draft

standard for trial) presents a comprehensive list of technical data elements required to manage digital image collections. The data elements cover an important subset of preservation metadata, including image-quality assessment, image enhancement and processing, and long-term collection management for ensuring the longevity of digital images. The most-perplexing part of the test implementation is agreeing on where and how to record this information in a systematic and persistent manner. In other words, developing best practices for harvesting, encoding, storing, and managing the metadata set. The digitization team is excited about the recent RLG effort in “Automatic Exposure—Technical Metadata for Digital Still Images,” a development in the right direction, as the main challenge at this point is the economics of capturing technical metadata.⁶ The initiative seeks to minimize the cost of acquiring and managing technical metadata. The project seeks to determine how digital repositories and asset-management systems can be supplied with technical metadata that is automatically captured by high-end scanners and digital cameras. This approach seems to dovetail nicely with automated digitization workflow software such as METAe.⁷

Conclusions

DCAPS is instrumental in supporting life-cycle management of digital image collections at Cornell University Library in several ways, including

- Supporting effective and efficient creation of digital image collections with a common service framework and technical architecture
- Enabling implementation of best practices and standards to create digital collections with long-lasting value
- Promoting financial accountability and business-plan development

Trying to accomplish these goals in a fast-paced production environment with diverse clients and collections requires an adaptable service framework. The ultimate objective is to balance image quality, user needs, resources, and preservation concerns in a cost-effective manner.

References

1. Anne R. Kenney, “Digital Benchmarking for Conversion and Access,” in *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, by Anne R. Kenney and Oya Y. Rieger, Mountain View, CA: Research Libraries Group, 2000.
2. “Preserving Cornell’s Digital Image Collections: Implementing an Archival Strategy,” <http://www.library.cornell.edu/imls/index.htm>.
3. DSpace Federation, <http://www.dspace.org/>.
4. “Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects, OCLC/RLG Working Group on Preservation Metadata, June 2002 <http://www.oclc.org/research/pmwg/>.
5. NISO Z39.87–2002 Data Dictionary—Technical Metadata for Digital Still Images http://www.niso.org/committees/committee_au.html.
6. “Automatic Exposure—Technical Metadata for Digital Still Images,” <http://www.rlg.org/longterm/autotechmetadata.html>.
7. The METAe engine is a comprehensive software package for digitizing books and journals in an automated and efficient way. It detects the structural elements of printed material automatically without any training. It also facilitates image-quality control and the OCR process. More information about the system can be found at <http://meta-e.aib.uni-linz.ac.at/>.

Biography

Oya Y. Rieger is the Assistant Director for Services in the Digital Library and Information Technologies division of the Cornell University Library. She manages the Library’s Digitization Services and coordinates the Digital Consultancy and Production Services. Rieger also serves as the Coordinator of Distributed Learning and facilitates the development of new policies and programs in support of technology-mediated instruction throughout the Library system. She serves on several national and international task forces, including co-chairing a NISO committee on technical metadata for image collections.