Winter 2003

Digitization in an Archival Environment

Sally McKay

Getty Research Institute

Follow this and additional works at: https://digitalcommons.unl.edu/ejasljournal

Part of the Archival Science Commons, Communication Technology and New Media Commons, Databases and Information Systems Commons, and the Scholarly Communication Commons

https://digitalcommons.unl.edu/ejasljournal/31

This Article is brought to you for free and open access by the E-JASL: The Electronic Journal of Academic and Special Librarianship at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in E-JASL 1999-2009 (volumes 1-10) by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Digitization in an Archival Environment

Sally McKay  
Research Library, Getty Research Institute  
smckay@getty.edu

Introduction

Cultural institutions such as museums, libraries, archives, and historical societies house remarkable collections of cultural artifacts. It is the responsibility of the staff working for those institutions to preserve, protect and provide responsible stewardship for the materials, and to the best of their ability, provide continued long-term access (Russell, 2000).

Advances in technology allow institutions to provide expanded access and education; however, there are important priorities that must be addressed prior to embarking on a digital conversion project.

Digitization in an archival environment includes taking a physical object or analog item, such as an art object, a tape recording, a map, or correspondence, from a collection that is rare or unique, often extremely fragile, and taking photographs of the item, and transferring the photographs to a digital medium. The negatives or prints are scanned into digital format such as a JPEG (1,400 pixels) and even larger, TIFF (Tagged Image File Format, 2000 pixels) files (Library of Congress, 2000). Digital files are imported into, and managed with the use of software programs. Digital files may be read, compressed, transferred and retrieved over computer networks then made accessible and viewed on computer monitors. The end product is determined by how well these functions are performed (Beamsley, 1999, p. 364).

Governmental agencies, institutions of higher learning, and the commercial and entertainment industries are fast developing technological infrastructures to accommodate the needed access on the Internet. Kenney and Rieger (2000) state, “the Internet will become the agora for research, teaching, expression, publication, and communication” (p. 1). Many, especially the younger generation, consult libraries and archives as a last resort. This must change if libraries and archives want to continue as primary information providers (Kenney & Rieger, p. 1).
Cultural institutions are investing in digital projects for several reasons including: to provide access, to reduce over-handling of material in order to preserve it, and “public relations” to assist in promoting the collections and the institution. By creating digital surrogates of their collections, institutions continue to support the notion that there is value in the materials they house (Kenney & Rieger, 2000, p. 1).

Most digital conversion projects are driven in part by the institutions strategic goals. Unfortunately institutional goals are often in conflict with the necessary structure of an ideal digital conversion project (Smith, 2000, p. 3). Resources are useless unless they are accessible. Therefore, if an institution is to embark on a digital conversion project, sufficient thought, planning, risk management, and correct infrastructure, both professionally and technologically, must go into the process or the project will fall short of the intended goals (Kenney & Rieger, 2000, p. 3).

**Advantages of digitization**

Digital imaging projects offer unique advantages. Information and content may be delivered directly to end-users, and can be retrieved remotely. Image quality can be quite good, and is often enhanced, with capabilities continuously improving (Conway, 2000). There is added advantage with the possibility of full-text searching, cross-collection indexing and newly designed user interfaces that allow for new uses of the material and content (Conway, 2000). Flexibility of the digital material is another advantage. Since the data is not “fixed”, as with paper or printed text, it is easy to reformat, edit and print (Smith, 2000, p. 3).

Moreover, the ability to provide a large number of users’ access to unique or special collections’ material (normally viewed only on-site) may be the most attractive feature of digital conversion projects. Online resources serve local, national and international needs. Increasing access by any means, specifically remotely, makes historical or literary research much easier (de Stefano, p. 13). Allowing for a wider audience to view digital surrogates of primary material provides a great service and increased utility to the collection (de Stefano, p. 13). There are no travel costs involved and this interaction may allow for the creation of new knowledge (Ingram, 2000, p. 19).

Providing access to primary material can help to “publicize” the material to other departments and peers, and to demonstrate the importance of the collections. The Special Collections departments may present the “jewels in the crown” from the research library (Ingram, 2000, p. 19).

Profound changes in professional attitudes, private and public funding, availability of image reproductions, and electronic communication technologies have resulted in museums and archival institutions re-evaluating their target audience. The general
education market is the new target audience, and the new method of providing information is through electronic media, most often through the World Wide Web (Beamsley, 1999, p. 362). Digitization projects allow for extended data recovery, enabling scholarship that was previously not possible with analog material. Computer enhancements, such as enhanced optical character recognition (OCR), allow for more in-depth analysis (de Stefano, 2000, p. 14).

But institutions need to realize that digital resources are institutional assets in their own right, and not merely surrogates of an analog object; they must be managed, preserved and migrated over time (Kenney & Rieger, 2000, p. 6).

Participating in digitization projects, allow for professional development as staff gain new skills, knowledge and expertise while completing the project. An institution and its staff also become “assets” and may share expertise and lessons learned with other institutions (Smith, 2000, p. 3). Not only does digitization provide “added value” to the resources; it may also breathe new life into older institutions (Kenney & Rieger, 2000, p. 2).

Another advantage of creating digital surrogates is, use of the surrogate reduces handling of the old or fragile material, hopefully extending the life of the original (de Stefano, 2000, p. 21).

**Disadvantages of digitization**

Required staff expertise and additional resources are often the greatest costs in digitization projects. Not only are large budget allocations needed to fund research and intellectual selection, but also time must be spent for feasibility assessments, training, and methodical prioritization of items or collections to be digitized. These requirements pull staff away from their regular workloads. Cataloging the new material adds additional base costs to the budget (Ingram, 2000, p. 19). Digital conversion projects require added levels of work not needed in traditional reformatting projects (Gertz, 2000, p. 100). Many institutions lack expertise and preparation must be well-planned (Ingram, p. 18).

Digital conversion is not a yet a form of preservation; which relies on long-term, stable media, which cannot be expected with today’s technology. The only accepted long-term preservation media are durable acid-free paper or preservation microfilm (Gertz, 2000, p. 97).

Access to successful digital surrogates often encourages people to wish to consult the original. This impacts staff in other ways with more calls, letters, and requests for publication or reproduction of the materials, and added reference service is necessary
(de Stefano, 2000, p. 13). High-quality surrogates must be created in order to satisfy the users’ needs, or they will need to go back and consult the original (de Stefano, p. 21-22).

Financial costs are extremely high and cultural institutions usually operate with either flat or marginally increasing budgets. Operational environments must have fundraising and accountability. With such great costs of staff time and funding, the “risk of loss”, is very high (Conway, 2000).

Another disadvantage of creating digital surrogates is that users are completely reliant on computers and stable Internet connections to view and retrieve the digital information (Smith, 1999, p. 2). Depending on users’ hardware and software capabilities access may be frustrating because of the large variety of computer models, platforms, software, and hardware around the world.

Ease of access to a digital collection leads to high expectations of end-users. There is a tendency to believe that everything is available online, that every piece of information is true and accurate, and that everything available online is free. Rarely do users understand or appreciate the scope of the collection and its relationship to other parts of the collection. (Ingram, 2000, p. 19).

**Legal Issues**

Selection of material to digitize should first be based on a clear and comprehensive understanding of ownership rights and copyright (de Stefano, 2000, p. 11). Tennant (2000) also states that copyright is the first issue that must be reviewed (p. 26).

Physical ownership does not automatically mean that an institution owns the rights to reproduce it; this is a mistake that some institutions make. In the past it was thought that when an object was transferred to an institution, so too were the legal rights to reproduce the object. Institutions can no longer count on the fact that legal rights are transferred (Beamsley, p. 372).

Technology is usually ahead of the law, and the Internet creates added pressures for new legislation to be created in order to protect digital material. There are many proactive ways institutions may protect their digital collections. Listing full copyright information with the images on Websites is one method. Institutions may also include a full overview of the United States Copyright Law and instructions for acquiring rights and reproductions (de Stefano, 2000, p.13). Controlled access may be used with required passwords, or unlimited access to the collections may be provided when digital images are marked correctly with ownership (Beamsley, 1999, p. 373).
Authentication of an object is one of the most important issues in museum administration. Special collection’s departments and archives house a large variety of materials, often including artistic works. Traditionally, cultural institutions such as museums have kept all administrative information protected (Beamsley, 1999, p. 361-62). With the expanded use of digital resources, how do individuals and institutions know that the digital image or reproduction is what is says it is? With the added ease of electronic manipulation, this is an extremely pressing issue (Beamsley, 1999, p. 371).

Adopting the use of metadata standards (or data about data), cultural institutions could adhere to the most current standards and established “best practices”, in an effort to keep digital files intact and authentic (Gilliland-Swetland, 2000).

Providing a highly structured information object with the use of metadata, allows for added searching, preservation, record keeping and specifically authentication (Gilliland-Swetland, 2000).

**Ethical Issues**

As professionals in the United States, librarians and archivists act in accordance with the ethical guidelines outlined by the American Library Association and the Society of American Archivists. Ethical guidelines outline professional conduct and stipulate that continued “public trust” is essential. This trust maintains that Special Collections librarians continue to promote scholarship, preserve material, provide access, and carefully handle material (Association of College and Research Libraries, 1992).

Professionals must act with integrity, avoiding any activities that may compromise themselves, their institutions or the collections (Association of College and Research Libraries, 1992). Conflicts of interest should be addressed and special collections’ librarians should not suppress professional judgement in order to conform to their institution (Association of College and Research Libraries).

Institutions are not only responsible for the care and preservation of their original assets, but also for the digital assets derived from originals. Specific to cultural institutions and non-profits is the 501(c)3 charter dictating ethical responsibility, that it is not enough to own and care for objects but that the legal definition of a museum stipulates that objects must exhibited to the public on a regular basis. In the last two decades this has been interpreted as a proactive mandate to provide useful mediated access to the collections with electronic media (Beamsley, 1999, p. 362). It is the job of the museum to provide access to its collections, and to facilitate and encourage their use. By controlling this aspect of access, museums can continue to control effectively the educational experience (Beamsley, p. 362).
Institutional administrative security must also be kept in mind. How much of the administrative information routinely kept by cultural institutions should be provided online? Do the institutions digitize all of the administrative data that accompanies objects? Insurance, security, and valuation information need to be kept secure (Hoopes, 1997, p.89).

**Technical Issues**

Technical feasibility, and a long-term commitment to technological infrastructure, must be evaluated prior to embarking on a digital conversion project. Systems managers must not only understand the constantly changing software and hardware available, but they must now learn the nuances of museology and related disciplines (Hoopes, 1997, p. 89).

Constantly changing software and hardware needs, along with changing product development, create even greater pressures on the cultural institutions, as the “risk of loss” is very high. Preservation of digital resources centers around the interim mechanism for storing the digital information, migrating it to a new form, and providing continued long-term access (Conway, 2000).

Technical issues in digital conversion projects also include establishing proper workflow, planning, and training of staff with continuous review of the project. Workflow must be tested throughout the project in order to evaluate and make any necessary changes. It is necessary to document decisions made and why, and costs associated with each step. This is important for all projects, in order to allow for continuation, but also to allow for future projects. Proper reporting needs to be documented, especially if funds were provided through a grant (Macklin, L & S. Lockmiller, 1999, p.13).

One of the greatest issues facing the longevity of digital information is not only the storage media deterioration but the problem of “rapidly changing storage devices” (Besser, 1999). Unlike analog information, where concentration is placed on preservation of the physical artifact, we must make a “conceptual leap” in order to preserve information in the digital age. It is the informational content that must be preserved. The problem lies in the fact that the content may now be completely removed from the physical artifact (Besser). It will take a conscious effort to make sure that the digital information survives. Analog formats (for example, papyrus, tablets, or books) have remained intact for long periods unless destroyed by natural disaster or personal intent. Continuously changing hardware and software creates real headaches for staff working on digital longevity (Besser).

Short-term solutions may be used to protect online resources. Digital assets may be protected by providing image files at low-resolution, protecting them from misuse, by
application software, protecting from unauthorized use, and by encryption or selective access to the file content (Beamsley, p. 364-374).

Encryption may be a short-term solution for securing use of digital files, but it will create even more difficulty for those in the future (Besser, 1999). Increasingly, Web-pages consist of several distinct files. Web designers are encouraged to use “good practice”, to properly manage long-term interoperability of files, which may assist with the long-term preservation of complex files (Besser).

Extrapolating out what future technology might be is futile. There are some short-term and long-term concepts, but there are no “cut and dry” solutions to digital longevity. It is difficult to imagine what technology might be available. Several long-term preservation concepts are emulation and migration, which may help in the longevity of digital resources. Migration is the process of moving files from one encoding environment to another, updating the information to a more “modern” computing environment, for example, moving information from WordPerfect to Microsoft Word95, then to Microsoft Word97 (Besser, 1999). The concept of emulation actually addresses the software application. Emulation focuses on the functionality of software allowing any file written in any format to run on whatever computing environment the end-user is running (Besser).

“Refreshing” is necessary with both emulation and migration. Digital files need to be transferred periodically to new physical storage media, in order to “refresh” the material and hopefully to keep it from physical decay, and obsolescence of the medium (Besser, 1999). If not, the material will be inaccessible, and therefore lost. Standards need to be agreed upon, as to who will preserve the material and professionals need to set guidelines and best practices (Besser).

Loss of format is a troubling issue, as information is transferred from program to program. Information is lost when analog material is digitized, and information may be lost as digital resources are “refreshed” or migrated to modern computing environments. Although identical digital copies may be made from digital files, functionality from every software program, ever made cannot possibly be emulated (Besser, 1999).

Besser (1999) states that metadata is the first line of defense to protect digital information and content. By providing detailed metadata, institutions may minimize the risks of digital resources becoming inaccessible in the future. Important unique technical information may be captured including: scanning specifications, operating systems, software versions, and decompression schemes. In addition to the institutional administrative data, it is important to maintain the digital integrity of the files (Beamsley, p. 371).
Decision-making is different for each institution. What may be “right” for the Library of Congress may not be feasible for a local archive or historical society. It is recommended that institutions, create the richest digital master initially, more than what may be required for the immediate use (Kenney & Rieger, 2000, p. 4,9).

Users needs should also be taken into account. Large-format items such as maps are difficult to view. Users prefer tools for viewing, such as zooming and panning capabilities. Navigation and manipulation capabilities are also important, with minimal scrolling and jumping (Sandore, 2000, p. 4).

Collaboration is a crucial element of all digitization projects. The many “players” must work together: librarians, archivists, curators, administrators, systems analysts, photographers and scanning technicians, scholars, programmers, catalogers and end-users (Kenney & Rieger, 2000, p. 6).

The importance of maintaining intellectual control of digitized collections cannot be overstated. This is particularly important due to the fact that powerful technology allows electronic material to be easily manipulated and the end user must be educated to the importance of authenticating the object, through the intellectual integrity of the digital material. Every effort must be made to educate end-users regarding the inherent problems associated with the digital medium and problems with authentication and accuracy of information (Ingram, 2000, p. 19).

Digital projects almost always take more time and money than originally expected; thus, loss of staff time and missed deadlines are almost inevitable outcomes of digitization projects (Ingram, 2000, p. 19).

Professionals working together should take into account that at the center of any digital program are both the users and the collections. Both provide a touchstone to which the other structures must be measured (Kenney & Rieger, 2000, p. 6).

Methodologies such as digital benchmarking may be followed with digital conversion projects. There is no one correct way to do things (Kenney & Rieger, 2000, p. 4). But it is important to try to “do it once” and create an enduring digital copy of the original item (Beamsley, 1999, p. 362). Benchmarking consolidates the decision making process based on complex issues regarding institutional goals, objectives, nature of the source document, institutional resources, and technological infrastructure. Benchmarking assists Institutions’ in making knowledgeable overall decisions, instead of individual decisions based item by item. It outlines consequences of decisions, and may assist in negotiating with vendors for services or products (Kenney, 2000, p. 24).
Benchmarking also allows institutions to focus on its needs and careful resource management, instead of having to format to vendors’ specifications. By outlining requirements an institution has a better idea of what it can deliver, and may make clear budget requests, and make sure that it buys the right equipment (Kenney, 2000, p. 24).

Prior to embarking on a digital conversion project, librarians (working with other departments) need to review and outline what “your” specific institution wishes to accomplish. Criteria may be discussed based on several questions; How well will the image be able to be captured? Will the digital image be a quality representation of the original with the full content of the original? Will adequate digital versions be made without damaging the original (Gertz, 1999, p. 99)?

**Selection of collections to be digitized**

Institutions select collections for digitization based on specific goals. Each decision should be weighed against a cost and benefit analysis for the end users and the institution. Subject specialists such as curators (familiar with both the collection and how it is used) must be key decision-makers (de Stefano, 2000, p. 15).

Selection for digital conversion continues to work under the premise of access and ease of use, rather than physical deterioration. Digital conversion projects should be goal driven rather than technologically driven (Gertz, 2000, p. 98 and Kenney & Rieger, 2000, p. 6).

Institutions contemplating digitization projects should research what other institutions have done and review their successes, mistakes, and lessons learned (Kenney & Rieger, 2000, p. 3). Selection of material may be based on a desire to educate and provide access to “hard-to-find” items of great intrinsic value. Selection may also be based on the value of the original item. For some institutions it is most important to spend the time and money to create a perfect reproduction in order to protect the original item (Fleischhauer, 2000, p. 20). The rare or fragile material often requires special handling. Material chosen for digital conversion often require preservation and conservation treatment prior to photography or scanning (Ingram, 2000, p. 18).

Some key issues that most institutions agree upon are: Does your institution have a legal right to digitize the items and make them available online? Does the material or collection have intrinsic value to make it highly used by the target audience? Is there added value by increasing extended access to the collection? Is the material unique? Is there necessary support by the parent organization and technological infrastructure to make the project possible? If the answers are “yes”, the project should be good to go forward (Tennant, 2000, p. 26).
Often microfilm is the only copy of an item after the original was destroyed and may also be scanned. However creating digital files after several generations may not provide the highest possible image quality (Beamsley, 1999, p. 370).

Providing expanded access and exposure to collection material on the Internet may also work as promotional device, by increasing donations, of both funding and “gifts” to the institution (Ingram, 2000, p. 19).

**Determination of items that should not be digitized**

Selecting collections or material to digitize should be based on a clear understanding of the users’ needs, and of what digital material can do, that analog material cannot (Conway, 2000). Often there is a misunderstanding between the stakeholders of what digits can and cannot do (Smith, 2000, p. 2).

It is often difficult for libraries and archives to determine if they have the legal right to scan and display material, even if they own the physical item. Institutions should determine if they have a legal right to digitize the material and make it available to the public (Ingram, p. 19). An institution may choose to digitize a collection that it does not have the copyright to, for “in-house” use only. The material may be fragile, or it may be used a great deal (M. Baca, personal communication, April 4, 2001).

Institutions need to know that digitizing is not a less expensive or safer than microfilming. A digital master is not a “preservation master”. Presently the only way that a digital master contributes to preservation is that is reduces handling and the physical “wear and tear” on an item (Smith, 2000, p. 2).

Collections that are rarely used or in low demand should not be selected for digital conversion, as the costs in time, resources, and money cannot be justified (de Stefano, 2000, p. 17). Mere potential does not increase or add value to underused collections (Conway, 2000, p. 5).

Digitization does not replace collection management. Digital surrogates should never replace the original analog item, even when trying to save shelf space. If an institution decides to deaccession brittle newspapers, it should first obtain microfilm copies to serve its patrons (Smith, 2000, p. 3).

**Conclusion**

Cultural institutions house rare and unique artifacts recording the history of humankind. Providing greater access to collections may bring together vast, disparate collections and may inspire new scholarly work. By prioritizing digital projects, allocating funds, and working together, cultural institutions provide added utility to collections.
Advances in technology create new challenges and workloads, for staff and institutions, not present before. Part of the problem lies in the fact that currently there is no consensus regarding digital conversion or preservation of digital material.

Professionals must work together to address the problems stemming from the fact that there are no set standards for preservation of digital material.

Even after one has addressed the legal, ethical, technical, and professional issues surrounding digital conversion projects, what still needs to be addressed are the needs of the end-users. Providing access to digital collections in an unmediated environment creates continued challenges.

**References**


Back to Contents

http://southernlibrarianship.icaap.org/content/v04n01/Mckay_s01.htm