DOCUMENT RESUME

ED 427 787 IR 057 294

AUTHOR Smith, Abby
TITLE Why Digitize?
INSTITUTION Council on Library Resources, Inc., Washington, DC.; Council on Library and Information Resources, Washington, DC.
ISBN ISBN-1-887334-65-3
PUB DATE 1999-02-00
NOTE 20p.
AVAILABLE FROM Council on Library and Information Resources, 1755 Massachusetts Ave., NW, Suite 500, Washington, DC 20036; Web site: http://www.clir.org ($15).
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Access to Information; Costs; Information Management; Information Services; *Information Storage; Information Technology; Library Materials; Library Technical Processes; Nonprint Media; *Preservation
IDENTIFIERS Council on Library Resources; *Digital Imagery; Digital Information Services; Digital Technology; *Digitizing

ABSTRACT

This paper is a response to discussions of digitization at meetings of the National Humanities Alliance (NHA). NHA asked the Council on Library and Information Resources (CLIR) to evaluate the experiences of cultural institutions with digitization projects to date and to summarize what has been learned about the advantages and disadvantages of digitizing culturally significant materials. Findings revealed that digitization often raises expectations of benefits, cost reductions, and efficiencies that can be illusory and, if not viewed realistically, have the potential to put at risk the collections and services libraries have provided for decades. One such false expectation--that digital conversion has already or will shortly replace microfilming as the preferred medium for preservation reformatting--could result in irreversible losses of information. This paper defines digital information; identifies weaknesses of digitization as a preservation treatment; discusses the benefits and drawbacks of digital technology for access; and highlights issues institutions must consider in contemplating a digital conversion project. (AEF)

# Why Digitize?

by Abby Smith

February 1999



Council on Library and Information Resources

Commission on Preservation and Access

Digital Libraries

Economics of Information

Leadership

# Why Digitize?

by Abby Smith

Council on Library and Information Resources
Washington, D.C.

February 1999

## Commission on Preservation and Access

The Commission on Preservation and Access, a program of the Council on Library and Information Resources, supports the efforts of libraries and archives to save endangered portions of their paper-based collections and to meet the new preservation challenges of the digital environment. Working with institutions around the world, CLIR disseminates knowledge of best preservation practices and promotes a coordinated approach to preservation activity.

## Digital Libraries

The Digital Libraries program of the Council on Library and Information Resources is committed to helping libraries of all types and sizes understand the far-reaching implications of digitization. To that end, CLIR supports projects and publications whose purpose is to build confidence in, and increase understanding of, the digital component that libraries are now adding to their traditional print holdings.

ISBN 1-887334-65-3

Published by:

Council on Library and Information Resources
1755 Massachusetts Avenue, NW, Suite 500
Washington, DC 20036

Web site at http://www.clir.org

Additional copies are available for $15.00 from the above address. Orders must be prepaid, with checks made payable to the Council on Library and Information Resources.

The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials ANSI Z39.48-1984.

## Contents

## Preface

Digital conversion of library materials has advanced rapidly in the past few years. It promises to continue to expand its reach and improve its capabilities with extraordinary speed. Digitization has proven to be possible for nearly every format and medium presently held by libraries, from maps to manuscripts, and moving images to musical recordings. The use of hardware and software for capturing an item and converting it into bits and bytes, matched by a quickly developing set of practices for describing and retrieving digital objects, is giving form to the talk of a "library without walls." But such a virtual library has a very real price. Managers of cultural institutions and those responsible for policy matters related to digitization often find themselves struggling not only to understand the new technologies, but also, and more importantly, to grasp the implications of those technologies and to understand what digitization of their collections means for their institution, its patrons, and the public.

This paper was written in response to discussions of digitization at meetings of the National Humanities Alliance (NHA). NHA asked CLIR to evaluate the experiences of cultural institutions with digitization projects to date and to summarize what has been learned about the advantages and disadvantages of digitizing culturally significant materials. As one might expect from the early years of growth of a popular yet experimental technology, the lessons learned vary greatly from one institution to another. It is risky to generalize, but CLIR has been actively engaged in fostering the development of digital technologies for libraries, and we feel it is important to provide an early assessment of the impacts of new technologies on traditional library roles.

What we have found is that digitization often raises expectations of benefits, cost reductions, and efficiencies that can be illusory and, if not viewed realistically, have the potential to put at risk the collections and services libraries have provided for decades. One such false expectation—that digital conversion has already or will shortly replace microfilming as the preferred medium for preservation reformatting—could result in irreversible losses of information. This paper seeks not to raise false alarms, but to encourage every professional responsible for some aspect of cultural custody to assess this new technology with a hopefulness tempered by patience and informed by experience.

## Author's Acknowledgments

For nearly a decade, first at the Library of Congress and now at CLIR, I have worked closely with librarians and curators charged with converting the riches of their institutions' holdings into digital form. Being able to look over their shoulders as they have gone about the daunting tasks of sorting through tough intellectual issues, while simultaneously meeting technological challenges, has deeply informed my thinking about possibilities and limitations, both human and technological. In their conversations with me, they have been generous with their time and helpful in their candor, and I am grateful to them. I owe a special debt to three who read this paper and offered many insights and improvements: Anne Kenney of Cornell University Library; Barclay Ogden of the Main Library at the University of California, Berkeley; and Donald Waters of the Digital Library Federation. While they may not agree with everything written here, their perspectives on a number of issues were invaluable to me and I thank them.

*Abby Smith*
*Director of Programs*

> The dream of the virtual library comes forward now not because it promises an exciting future, but because it promises a future that will be just like the past, only better and faster.
>
> —JAMES J. O'DONNELL, *Avatars of the Word*

## Introduction

In the digital world, all knowledge is divided into two parts. The binary strings of 0s and 1s that make up the genetic code of data allow information to be fruitful and multiply, and allow people to create, manipulate, and share data in ways that appear to be revolutionary. It is often said that digital information is transforming the way we learn, the way we communicate, even the way we think. It is also changing the way that libraries and archives not only work, but, more fundamentally, the very work that they do. It is easy to overstate—and underestimate—the transformative power of a new technology, especially when we do not yet understand the full implications of its many applications. Nonetheless, people have embraced this technology enthusiastically, often as an answer to questions that had not, in many cases, yet been posed. Librarians everywhere hear the voices of people speaking like evangelicals, urging the conversion of text and visual materials into digital form as if conversion per se were a self-evident good. But because we tend to imagine the future in terms of the present, as O'Donnell points out, such projections of the present onto the future may, at best, be misleading. If this new technology does, indeed, turn out to be revolutionary, then we cannot anticipate its impact in full, and we should be cautious about letting the radiance of the bright future blind us to its limitations.

While we may not yet fully understand the ways in which this technology will and will not change libraries, we can already discern some simple, yet profoundly important, patterns in digital applications that presage their effective and creative use in the traditional library functions of collecting, preserving, and making information accessible. A critical mass of experience is accumulating among libraries and archives active in digitizing parts of their collections, ranging in size from the Library of Congress, the National Archives, and major research libraries in the Digital Library Federation, to smaller institutions such as the Huntington and Denver Public libraries. Their experiences reveal patterns that can help us assess when the technology is able to meet expectations for improvement of

traditional library services, when it cannot, and when it may do so, but not in a cost-effective manner. This paper will address the question of why a library should invest in the conversion of its traditional materials into digital form—in other words, what are the advantages and disadvantages of converting traditional analog materials into digital form.

## What is Digital Information?

Until very recently, all recorded information was analog—that is, a continuous stream of information of varying density and type. Analog information can range from the subtle tones and gradations of the chiaroscuro in a Berenice Abbott photograph of Manhattan in early morning light, to the changes in volume, tone, and pitch recorded on a tape that might, when played back on equipment, turn out to be the basement tapes of Bob Dylan or the Welsh accents of Dylan Thomas reading *Under Milk Wood*. But when such information is fed into a computer, broken up into 0s and 1s and put together in a binary code, its character is changed in quite precise ways.

Digitally encoded data do not represent the infinitely variable nature of information as faithfully as analog forms of recording. Digits are assigned numeric values which are fixed, so that great precision is gained in lieu of the infinitesimal gradations that carry meaning in analog forms. For example, when a photograph is digitized for viewing on a computer screen, the original continuous tone image is divided into dots with assigned values that are mapped against a grid. The pattern of the dots is remembered and reassembled by the computer upon command.

Those bits of data can be recombined for easy manipulation and compressed for storage. Voluminous encyclopedias that take up yards of shelf space in analog form can fit onto a minuscule space on a computer drive, and that same digital encyclopedia can be searched in many ways other than alphabetically, making possible information retrieval that would have been unimaginable if one had only the analog copy, on paper or microfilm.

Data that are not being used are not like books on a shelf or the family correspondence and photos stored in shoe boxes at the back of a closet. They are more like the stacks of LPs or the 8mm family home movies in storage in a basement. That is, digital information is not eye-legible: it is dependent on a machine to decode and represent the bit streams in images on a computer screen. Without that machine, and without active human intervention, those data will not last.

One of the most important qualities of information in digital form is that by its very nature it is not fixed in the way that texts printed on a paper are. Digital texts are neither final nor finite, and are fixed neither in essence nor in form except when a hard copy is printed out, for they can be changed easily and without trace of erasures or emendations. Flexibility is one of the chief assets of digital information and is precisely what we like about text poured into a word processing program. It is easy to edit, to reformat, and to commit to print in a variety of iterations without the effort required to produce hard copy from a typewriter. That is why visual designers like computer-assisted design programs. It is easy to summon up quickly any number of variations of value, hue, shape, and placement to see, rather than to imagine, what different visual options look like. Furthermore, we can create an endless number of identical copies from a digital file, because the file does not decay by virtue of copying.

From the creator's point of view this kind of plasticity may be ideal, but from the perspective of a library or archives that endeavors to collect a text that is final and in one sense or another definitive, it can complicate things considerably. Because the digital text is flexible and easily changed, the matter of preserving digital information becomes conceptually problematic. Which version of the file, or how many versions, should be archived? There are also formidable technical obstacles to ensuring the persistence of digital information.

## Digitization is not Preservation—at Least not Yet

All recorded information, from the paintings on the walls of caves and drawings in the sand, to clay tablets and videotaped speeches, has value, even if temporary, or it would not have been recorded to begin with. That which the creator or transcriber deems to be of enduring value is written on a more or less durable medium and entrusted to the care of responsible custodians. Other bits of recorded information, like laundry lists and tax returns, are created to serve a temporary purpose and are allowed to vanish. Libraries and archives were created to collect and make available that which has long-term value. And libraries and archives serve not only to safeguard that information, but also to provide evidence of one type or another of the work's provenance, which goes towards establishing the authenticity of that work.

Though digitization is sometimes loosely referred to as preservation, it is clear that, so far, digital resources are at their best when facilitating access to information and weakest when assigned the traditional library responsibility of preservation. Regrettably, because

digitization is a type of reformatting, like microfilming, it is often confused with preservation microfilming and seen as a superior, if as yet more expensive, form of preservation reformatting. Digital imaging is not preservation, however. Much is gained by digitizing, but permanence and authenticity, at this juncture of technological development, are not among those gains.

The reasons for the weakness of digitization as a preservation treatment are complex. Microfilm, the preservation reformatting medium of choice, is projected to last several centuries when made on silver halide film and kept in a stable environment. It requires only a lens and a light to read, unlike computer files, which require hardware and software, both of which are developed in often proprietary forms that quickly become obsolete, rendering information on them inaccessible. At present, the retrieval of information encoded in an obsolete file format and stored on an obsolete medium (such as 8-inch floppy diskettes) is extremely expensive and labor-intensive, when at all possible. Often the medium on which digital information is recorded is itself inherently unstable. Magnetic tape is one example of a common digital medium that requires special care and handling and has been known to degrade within a decade, beyond the point where information can be recovered. Magnetic forms of analog recording, such as video and audio tape, are equally fragile and unreliable for long-term storage. In its inherent physical fragility, magnetic tape is not different in essence from the acid paper so widely produced in the last 150 years, but its life span is often dramatically shorter than that of poor quality paper.

More important even than the durability of the medium is the need to keep the data fresh and encoded in readable file formats. Ongoing investigations into two possible ways of ensuring data persistence—the migration of data from one software and hardware configuration to a more current one, and the creation of software that emulates obsolete encoding formats—may develop solutions to this problem. As yet, we have no tested and reliable technique for ensuring continued access to digital data of enduring value, although information stored on nonproprietary formats such as ASCII has been migrated successfully (in the case, for example, of specific government records). Nevertheless, migration from one software to another does not produce a new file exactly identical to the old one. Though data loss may not necessarily mean loss of intellectual content, the file has been changed.

Another reason that preservation goals are in some fundamental way challenged by digital imaging is that it is quite difficult to ascertain the authenticity and integrity of an image, database, or text when it is in digital form. How can one tell if a digital file has been

tampered with and the content changed or falsified? Looked at from the traditional perspective of published or manuscript materials, it is futile even to try: there is no original with which to compare a suspect file. Copies can be deceptively faithful: one cannot tell the difference between the original output of a scan of the Declaration of Independence, and one that is output four months later. In contravention of a core principle of archival authenticity, one can change the bit stream of a file and leave no record of its having been altered. There is much research and development being dedicated to solving the dilemma posed by the stunning fidelity of digital cloning, including methods for marking images and time-stamping them, but as yet there is no solution.

Authenticity may not be important for a digital image of a well-known document like the Declaration of Independence, in which access to either the analog original or a good photographic image is easy enough to obtain for comparison's sake. But anyone who has seen the digitally engineered commercial in which Fred Astaire can be seen dancing with a vacuum cleaner can readily understand the ease with which improbable digital occurrences can become real because we can be made to see them. After all, the evidence is before our eyes, and our eyes cannot detect a falsehood. It is our cognitive reasoning that detects that falsehood, not our eyes. That image of the suave, gliding across the floor with the functional, startles and amuses us because it confounds our expectations.

But what if we arrive at a library Web site, for example, looking for an image that we have never seen and about which we have few expectations. The only reason that we expect that image to be a truthful representative of the original is that we can rely on the integrity of the institution that has mounted the files and makes them available to us. We transfer the confidence we experience in the reading room of that library to our work station, wherever it may be. We go to the New York Public Library Web site with the full expectation that the library "guarantees" the integrity of the images they mount. But it would be very hard indeed for a researcher in Alaska looking at New York Public Library's Digital Schomburg site to verify independently that any given image is indeed a faithful representation of the original.

The problem of authenticity is far from unique to the digital realm. Forgers and impostors have a distinguished history of operating successfully and often long undetected in print and photographic media, although they have had to work harder and smarter than their digital counterparts. The traditional methods for authenticating documents that have served the library and archival professions well until now have relied largely on practices derived from markers car-

ried on the physical medium itself. After a textual examination to look for obvious differences in content, researchers have often then examined the physical carrier itself—the book or manuscript leaf—to see if there are any signs of modification or falsification. From a simple examination of watermarks to a variety of sophisticated chemical, optical, and physical tests that can verify the age of paper, the composition of inks, and the physical traces of erasures and palimpsests, researchers have resorted to a number of strategies to verify the authenticity of a document. Granted, there are few who routinely insist on that level of authentication in doing research, but that is because the pitfalls of using books, manuscripts, and visual materials are familiar to us and we tend to discount them without much conscious thought. We should be wary of reposing the same quality of trust in digital resources that we do in print and photographic media until we are equally familiar with their evidentiary weaknesses.

As in other forms of reformatting, digital scanning has implications for the original item and its physical integrity. Depending on the policy of a library or archival institution, the original of a scanned item may or may not be retained after reformatting. To the extent that a reader can make do without handling the original, the digital preservation surrogate can serve to protect it from wear and tear. If there is concern that the scanning process could damage materials, one would choose to scan a film version of the original.

The advantages of scanning for access purposes may be combined with those of preservation microfilming by using the model of hybrid conversion, that is, creating preservation-standard microfilm and scanning it for digital access purposes, or, conversely, beginning with a high-quality scan of the original and creating computer-output microfilm (COM) for preservation purposes. Work is presently underway to articulate and refine best practices for implementing the hybrid approach to reformatting so that it can be adopted by libraries across the country. Of course COM, unlike microfilm created from the original, is only a recording of digital images on an analog medium. Though it has been fixed on a durable medium, some would argue that the image itself, having been generated digitally, has lost some essential information—or has at least lost its fundamental analog character—and cannot therefore claim to be as desirable for preservation as film made by photographing the original source.

Although this may seem a minor point to those more interested in easy access than in that level of authenticity, it is still important to understand that digital technology transforms analog information radically. There has to be some loss of information when an analog item is made digital, just as there is when one analog copy is made

from another. On the other hand, there is virtually no loss of information from one generation of a digital copy to another. Images will not degrade when copied, in contrast to microfilm, which loses about 10 percent of its information with each copy. Once there is more than one copy of a digital file, it is impossible to pick out the original, and one will never speak of "vintage files" the way that one now speaks of vintage photographs. On the other hand, digital images are less likely to decay in storage if they are refreshed, the images will not degrade when copied, and the digital files will not decay in use, unlike paper, film, and magnetic tape.

## Digitization is Access—Lots of It

Digital files can provide extraordinary access to information. They can make the remote accessible and the hard to see visible. Digital surrogates can bring together research materials that are widely scattered about the globe, allowing viewers to conflate collections and compare items that can be examined side by side solely by virtue of digital representation. The easy access to reference surrogates—images that provide a great deal of the information contained in the original, even if at fairly low resolution—is a boon to researchers when developing efficient and effective research strategies. Through the use of thumbnail images, which do not require high resolution, one can at a minimum acquaint oneself with the source enough to know whether or not one needs to consult the original. Very often one can make do with the digital surrogate because it provides all the information required. An image of the 1612 map of Virginia by John Smith may provide a scholar enough information to determine how far inland Smith actually traveled. The black crosses he laid down on paper to mark the furthest points he reached on various treks are clearly legible even on a low-resolution image.

One must think about the nature of the source materials (color, black and white, or shades of gray) and the use of the images (who will be consulting them and for what) when making decisions about the parameters for image capture. The quality and utility of an image depend upon the technology of capture and display, and the usefulness of an image, even if only for reference, can be severely compromised by a low-resolution monitor on which the image will be displayed. While work is ongoing to address the quality control and variability of computer monitors, as yet the lack of control over display mechanisms constitutes one of the weakest links in the digital chain of transmission.

Image processing—the manipulation of images after initial digital capture—can greatly expand the capacity of the researcher to

compare and contrast details that the human eye cannot see unaided. Images can be enhanced in size, sharpness of detail, and color contrast. Through image processing, a badly faded document can be read more easily, dirty images can be cleaned up, and faint pencil marks can be made legible. The plan of the District of Columbia prepared by Pierre-Charles L'Enfant for George Washington in 1791 is so badly faded, discolored, and brittle that it resembles a potato chip. It cannot be used by researchers and yields little detailed information to the unaided eye. Digitized several years ago, the map now can be displayed to allow us to make out all the subtle contours of the architect's plan and to read the numerous annotations made by Thomas Jefferson. Like successful archaeologists, we have, with our digital picks and brushes, excavated important historical evidence that has changed the way we understand the planning of the nation's capital.

Digital technology can also make available powerful teaching materials for students who would not otherwise have access to them. Among the most valuable types of materials to digitize from a classroom perspective are those from the special collections of research institutions, including rare books, manuscripts, musical scores and performances, photographs and graphic materials, and moving images. Often these items are extremely rare, fragile, or, in fact, unique, and gaining access to them is very difficult. Digitizing these types of primary source materials offers teachers at all levels previously unheard-of opportunities to expose their students to the raw materials of history. The richness of special collections as research tools lies in part in the representation of an event or phenomenon in many different formats. The chance to study the presidential election of 1860 by looking at digital images of daguerreotypes of the candidates, political campaign posters (a recent innovation of the time), cartoons from contemporary newspapers, abolitionist broadsides and notices of slave auctions, and the manuscript of Lincoln's inaugural address in draft form reflecting several different stages of composition—such an opportunity would be possible with a well-developed plan of digital conversion of materials from different repositories normally beyond the reach of students.

While we know, for example, that the daily number of hits at the Library of Congress American Memory site is greater than the number of readers who visit the library's reading rooms each day, we have very little data now as to how much these types of online images are used and for what purposes. Some large libraries are attempting to compile and analyze use statistics, but this labor-intensive task presents quite a challenge. We need more user studies before we can assert confidently what may seem self-evident to us now: that adding digitized special collections to the mass of information available

on the Internet is in the public interest and enhances education. We also need to ensure that libraries are working collaboratively in their efforts to digitize materials so that together they create a critical mass of research sources that are complementary and not duplicative, and that begin to fulfill the promise of coordinated digital collection building. However, at present there is no central source of information about what has been digitized, and with what care in the process, as there is for titles that have been microfilmed for preservation.

Some of the drawbacks of digital technology for access, as for preservation, stem from the technology's uncanny ability to represent the original in a seemingly authentic way. Working with digital surrogates can distort the research experience somewhat by taking research materials out of the context of the reading room. The nature of computer display makes only serial viewing possible, very different indeed, for example, from spreading photographs in their original sizes around a flat surface and looking at them simultaneously and in different groupings. Every object, every page, is mediated by the screen, which automatically flattens and decontextualizes the images. And a digital image, no matter how high the resolution and sensitive the display monitor, is always presented through the relatively low density of information of the computer screen, compromising the high-density nature of analog materials, which can be critical for assessing some visual evidence.

Digital "raw materials" on the Web are not as raw as they might appear to be. Many of the items that may be viewed now on the Web sites of such institutions as the National Archives, the Library of Congress, and the New York Public Library, come from special collections that are large, often cataloged only at the collection level, and often unedited, with few descriptions that aid a scholar. In order to digitize them, curators familiar with the materials sift through collections and make selections from them. The amount of physical preparation and intellectual control work that is needed for every digital project is very large indeed. Scanning is a very expensive process, and most of the cost occurs before the item is laid on the scanner. Part of that cost is the physical preparation of, research into, and description of an item. A collection of daguerreotypes that may have been in reasonably good physical condition but not very well cataloged may undergo extensive conservation review and treatment before it is scanned, and labor-intensive searches into the identities of faces that have been anonymous for decades may precede the cataloging and description of the digitized images. While these searches may be viewed as extraneous, or at least discretionary, editorial expenses, in fact they are more commonly incurred than not. The collections that are on the Web are, in a real sense, publications, accom-

panied as they are by a great deal of descriptive information created in order to make the items understandable in the context of the Internet.

The users of library Web sites need this information. Because they are used to having a reference librarian available to help them in their searches when they are at a library, they often want a library site to provide comparable reference and searching functions. They expect higher levels of functionality of digital objects than they do of library materials, in part because there is no online equivalent to a reference specialist available.

Despite the high cost of digital conversion, many institutions are taking on ambitious projects in order to find out for themselves what the technology can do for them. They are investing large amounts of money in projects to make their collections more accessible and, too often, believing that they are also accomplishing preservation goals at the same time. The impact of digitizing projects on an institution, its way of operating, its traditional audience, and its core functions, is often hard to anticipate. The challenge of selecting the parts of a large collection that will be scanned is, for some, a novel task that calls into question basic principles of collection development and access policies. Many libraries and archives have collections that are intrinsically valuable by virtue of being comprehensive and containing much information that is essentially unpublished. But they also may contain sensitive materials, those that deal with historical events or previously popular attitudes that may be offensive to us now and that must be understood in the larger context, and this is precisely what a comprehensive collection provides—context.

How does one deal with sensitive materials in a networked environment? Making information available on the Internet removes the very barriers from use that we take for granted in physical collections. No one has to travel to a library, nor do they have to present proof of their serious research interest in order to gain access to complex, disturbing, and uninterpreted material. On the other hand, if one makes the difficult decision to edit out materials that are readily served in a reading room, but are too powerful to broadcast on the Internet, what does that do to the integrity of a research collection? There are ways to build in electronic barriers to access for all or portions of a site, using much the same technology that commercial entities use in granting fee-based access. However, constructing these barriers adds a layer of administrative complexity to managing the site that libraries and archives may not be prepared to take on, even if the technology does exist. Only when digitization is viewed specifically as a form of publishing, and not simply as another way to make resources available to researchers, are the thornier issues of selection for conversion put into an editorial context that provides a

strong intellectual and ethical basis for imaginative selection of complex materials.

Many of the collections that may be of the highest research and teaching value will not be digitized for Web access because of the strictures of copyright that might apply. For this reason, library Web sites these days contain a disproportionate amount of public domain material, which distorts the nature of the source base for research restricted to the Web. The notion on the part of many young students that, if it is not on the Web or in an online catalog, then it must not exist, has the effect of orphaning the vast majority of information resources, especially those that are not in the public domain. This is not what the Framers had in mind when they wrote the copyright code into the Constitution, "to promote the Progress of Science and useful Arts." This skewed representation of created works on the Web will continue for quite some time into the future, and the complications that surround moving image and recorded sound rights means, ironically, that these will be the least accessible resources on the most dynamic information source around. And until Optical Character Recognition (OCR), the post-processing technology that makes scanned text searchable, works as well for scripts using non-Latin characters as for those using Latin ones, resources from around the world in vernacular languages will not take their proper place in the scanning queue.

## What is Gained and What is Lost

In contemplating a digital conversion project, an institution must ask itself what can be gained from digitization, and whether the value added is worth the price. Many libraries have begun the difficult task of developing criteria for selecting for digitization and have published their criteria on the Internet. Columbia University, for example, was among the first to post guidelines for selection of materials for digital conversion, which include the criterion of added value (available from http://www.columbia.edu/cu/libraries/digital/criteria.htm). They define the added value of digital capture as

- enhanced intellectual control through creation of new finding aids, links to bibliographic records, and development of indices and other tools;
- increased and enriched use through the ability to search widely, manipulating images and text, and to study disparate images in new contexts;
- encouragement of new scholarly use through the provision of enhanced resources in the form of widespread dissemination of local or unique collections;

- enhanced use through improved quality of image, for example, improved legibility of faded or stained documents; and
- creation of a "virtual collection" through the flexible integration and synthesis of a variety of formats, or of related materials scattered among many locations.

At present, however, the cost of digitization and of creating and maintaining a migration path for preserving the files is very expensive. The benefits of making an underused collection more accessible should be viewed in conjunction with other factors such as compatibility with other digital resources and the collection's intrinsic intellectual value. As the Society of American Archivists has said, "The mere potential for increased access to a digitized collection does not add value to an underutilized collection. It is a rare collection of digital files indeed that can justify the cost of a comprehensive migration strategy without factoring in the larger intellectual context of related digital files stored everywhere and their combined uses for research and scholarship." (Available from http://www.archivists.org/governance/resolutions/digitize.html.)

As Donald Waters of the Digital Library Federation has expressed it, the *promise of digital technology is for libraries to extend the reach of research and education, improve the quality of learning, and reshape scholarly communication*. This is not an extravagant claim for the technology, but rather a declaration of an ambition shared by many who are developing and managing the technology. And the key to fulfilling that promise lies within the communities of higher education, science, and public policy responsible for applying digital technology to those ends. Digital conversion of library holdings has its stake in this ambition, particularly to the extent that it can broaden access to valuable but scarce resources. But the cost of conversion and the institutional commitment to keeping those converted materials refreshed and accessible for the long-term is high—precisely how high, we do not know—and libraries must also ensure the longevity of information that is created in digital form and exists in no other form. We need more information about what imaging projects cost, and about who uses those converted materials and how they use them, in order to judge whether the investment is worth it. In the meantime, libraries must continue to be responsible custodians of their analog holdings, the print, image and sound recording collections that are their core assets and the legacy of many generations. This task requires continuing use of tried-and-true preservation techniques such as microfilming to ensure the longevity of imperiled information.

Analog is a different way of knowing than digital, and each has its intrinsic virtues and limitations. Digital will not and cannot re-

place analog. To convert everything to digital form would be wrong-headed, even if we could do it. The real challenge is how to make those analog materials more accessible using the powerful tool of digital technology, not only through conversion, but also through digital finding aids and linked databases of search tools. Digital technology can, indeed, prove to be a valuable instrument to enhance learning and extend the reach of information resources to those who seek them, wherever they are, but only if we develop it as an addition to an already well-stocked tool kit, rather than a replacement for all of those tools which generations before us have ingeniously crafted and passed on to us in trust.

**U.S. Department of Education**
*Office of Educational Research and Improvement (OERI)*
*National Library of Education (NLE)*
*Educational Resources Information Center (ERIC)*

# NOTICE

## REPRODUCTION BASIS

☒ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").