Collection Management

ISSN: 0146-2679 (Print) 1545-2549 (Online) Journal homepage: https://www.tandfonline.com/loi/wcol20

From Analog to Digital

Anne R. Kenney & Paul Conway

To cite this article: Anne R. Kenney & Paul Conway (1998) From Analog to Digital, Collection Management, 22:3-4, 65-77, DOI: 10.1300/J105v22n03_07

To link to this article: https://doi.org/10.1300/J105v22n03_07

From Analog to Digital: Extending the Preservation Tool Kit

Anne R. Kenney
Paul Conway

DIGITAL IMAGING AND PRESERVATION– UNCHARTED TERRITORY

[KENNEY] The 1990s have been called by some the decade of the image. For those of us concerned with the preservation and access of research library materials, this claim has a specific meaning tied to the emergence of digital-imaging technology, which represents a powerful new way to manage, store, and retrieve information. Its use stems from a convergence of technological capability and opportunity, including the ubiquitous nature of personal computing, the development of high-speed networks that are accessible to an increasing number of individuals and organizations, the declining cost and increasing capacity of mass storage, and the availability of reasonably priced, high-quality, production scanning systems. By providing for immediate, simultaneous, multiple, and random access to resources located at geographically distant places, digital technology has the potential to expand dramatically the rapid availability of information to users world-wide. Increasingly, business, government,

---

Anne R. Kenney is Associate Director of the Department of Preservation and Conservation, Olin Library, Cornell University.

Paul Conway is Head of the Preservation Department, Sterling Library, Yale University. Conway was unable to attend the symposium, and his remarks were delivered by Barclay Ogden, Head of the Conservation Department, Doe Library, University of California at Berkeley.

[Haworth co-indexing entry note]: "From Analog to Digital: Extending the Preservation Tool Kit." Kenney, Anne R., and Paul Conway. Co-published simultaneously in *Collection Management* (The Haworth Press, Inc.) Vol. 22, No. 3/4, 1998, pp. 65-79; and: *Going Digital: Strategies for Access, Preservation, and Conversion of Collections to a Digital Format* (ed: Donald L. DeWitt) The Haworth Press, Inc., 1998, pp. 65-79.

industry, and the professional and scientific communities are turning to the use of digital technology to manage information and to make the full text of important sources routinely available.

Scholars in the arts and humanities, however, continue to rely heavily on the books, serials, archives, and special collections materials stored on library shelves. With notable exceptions, few historical sources have been converted to electronic form. It is for this group of users–who incidentally represent the primary beneficiaries of preservation reformatting–that the application of digital-imaging technology could prove most significant.

Indeed, there is a rising level of interest throughout the library community in the use of digital imaging for preservation reformatting as evidenced by the many conferences devoted to this topic over the past several years. Additionally, major funding agencies report receiving an increasing number of proposals for projects that involve the use of information technology to capture and make available research materials.

While interest is high in the use of digital technology, the knowledge base–including the development of commonly accepted protocols and standards for the use of digital technology in a preservation context–is low. As a consequence, funding agencies and research institutions alike are slow to implement programs for scanning and digitization.

An important first step towards the recognition of the value of digital-imaging technology in this context has come in the form of *Considerations for Converting Materials to Electronic Form*, which was recently produced by the Joint Federal Funders Group, representing agencies that make grants relating to archival, library and other primary research materials, including the Department of Education, the National Archives, the National Endowment for the Humanities, and the National Science Foundation. This publication covers some of the major issues associated with conversion that will be addressed during this symposium.

The symposium is designed to provide participants with a baseline level of knowledge about the use of digital-imaging technology for preservation reformatting. Throughout the next several days, we will come to understand that digital reformatting does not end with the conversion process itself, but encompasses a host of related processes associated with imbuing the digital images with "intelligence" and the requirements for associated indexing of structure and content.

The symposium is intended to raise as many questions as it answers. One of the basic truths to remember about this emerging technology is that nothing is as sure as change itself. My rule of thumb is that when I really think I have mastered the situation, I'm in trouble because the ground rules change all the time and my complacency means that I have failed to keep

up. I suggest you consider the information presented here as providing a framework for the theoretical as well as the practical–and to recognize that the technical information is but a snapshot of what is available at this particular point. Nonetheless we will begin there. To give you the technical background against which to assess the role and requirements of using digital technology, I turn this presentation over to Paul Conway.

THE RIGHT TOOL FOR THE RIGHT JOB

[CONWAY] A colleague of mine once proposed that when someone goes to the hardware store to buy a 1/4-inch drill bit, they really need a 1/4-inch hole. The more we know about the problem we need to solve and why we need to solve it, the easier it is to select the right tool for the right job. The job that brings us to this symposium is the need to make sure that any investment we make in converting library materials with long-term value to digital image is well spent on behalf of our patrons–present and future.

Digital-imaging technology is a tool with many possible applications. It is only a tool not necessarily a solution, as some vendors would have us believe. In this regard it is extremely important to distinguish between acquiring imaging technology to solve a particular problem on the one hand and adopting it as a preservation option. Acquiring an imaging system primarily to improve access to information now is almost as simple as choosing the right combination of available features to meet immediate management goals. Adopting the technology for preservation, on the other hand, requires a deep and long-standing institutional commitment, the full integration of the technology into our information management procedures and processes, and significant leadership in developing appropriate definitions and standards of quality.

It is our job, then, to reconcile what seems like a fundamental contradiction between our traditional preservation responsibility and the promise of new and emerging, but rapidly changing, information technologies. Library and archival administrators who wish to add digital-imaging technology to their preservation tool kits must take to heart the following statement: the fundamental goal of digital preservation is to preserve continuing access to digital data for as long as that data has value.

Subsumed in the goal statement are assumptions that should be made explicit. Administrators who have responsibility for selecting imaging systems for materials with long-term value also bear responsibility for providing long-term access. This commitment is a continuing one that requires that decisions about preservation and access not be deferred in the

hope that technological solutions will emerge. Decisions on the long-term value to researchers and scholars of library and archival collections that may be converted to digital images are made independently of the decision to adopt the technology. Although increased access by image conversion may indeed add value to library holdings, the point of departure must be assessments of the research value of materials in their original format.

The process of converting library materials to an electronic form is distinct from the medium upon which the images are stored. This distinction allows for a continuing commitment to the digitized information while entertaining the possibility that other, more advanced storage media may render today's optical media storage obsolete. The best digital-imaging system is an integrated collection of hardware and software components that may serve short-term functions in addition to the conversion of library materials for long-term retention.

FOUR PRIMARY ISSUES

Four major issues must be considered by library administrators who wish to adopt digital-imaging technology as part of a comprehensive preservation strategy.

1. System functionality over time.
2. Storage media deterioration and migration of data.
3. Digital image data quality.
4. Integrity of information sources.

I should note in passing that the imaging industry has not reached consensus on some of these points, and there is certainly room for other viewpoints.

1. System Functionality Over Time

It is ironic that today's optical media will most likely far outlast the current hardware and software systems that retrieve and interpret the data stored on them. Since libraries and archives can ill-afford to become museums of obsolete computer technology, we must work simultaneously with manufacturers and within our own institutions to maintain the functionality of the systems we acquire and upgrade their capabilities as the technology evolves. There are at least five aspects to this issue.

- *Open Systems Architecture.* Open systems architecture is a systems design approach that permits users to interchange system hardware

components with minimal impact on the primary operating software and to upgrade the system over time without significant data loss. Open systems architecture should be required for new digital-imaging applications.

- *Nonproprietary Systems.* One of the keys to open systems is the development of nonproprietary standards. Because the barriers imposed by proprietary system configurations can create serious problems for long-term access to documents stored on optical media, vendors with proprietary products should be required to build linkages to systems with nonproprietary configurations.
- *Backward Compatibility.* A useful way to mitigate the impact of information technology obsolescence is to require that new system generations be *backward compatible*, that is, able to read information written by an older generation of technology and convert it to a newer one.
- *Technical Documentation.* Full technical documentation of system components, application software, and operating systems is essential to facilitating long-term access. Administrators should require the delivery of a complete set of documentation, including source code, object code, and maintenance documentation.
- *Responsible Custody.* Digital-imaging systems cannot solve access problems stemming from inefficient manual or computerized information systems and practices. It is necessary to document all aspects of the design and use, including administrative procedures for imaging, retrieval, and storage; problems encountered over time; and measures taken to address them, including hardware and software modifications. It is our responsibility, rather than that of vendors and manufacturers, to ensure that policies and procedures for long-term access are developed and consistently applied.

2. Storage Media Deterioration and Migration of Data

At the heart of any imaging system's functionality is the ability to retrieve data from optical storage media as reliably as possible for as long as possible. Before data error rates become unacceptable (or even fatal), it is necessary to migrate digital data (and the accompanying index information) to newer generations of imaging systems. Simply refreshing data by copying to new disks, as is typical in a magnetic media arena, is not an acceptable long-term solution when imaging systems themselves are becoming obsolete in three to five years. In the area of media longevity, there are at least five issues to consider.

FOUR PRIMARY COMMITMENTS

I suggest, therefore, that we need to make four primary commitments if we plan to integrate imaging technology into our preservation programs. We need to:

- Transfer valuable information across technology systems as these systems emerge.
- Deemphasize storage media formats as the central focus of preservation concern.
- Shift that concern to the fundamental challenge of specifying and then obtaining digital image quality.
- Recognize the importance of maintaining structural, that is, contextual, as well as content indexes.

Given these four commitments, imaging technology is not simply another reformatting option. It is far more than that, and we ought to find a new term to describe what imaging is all about for libraries and archives. I suggest that *transformatting* instead of *reformatting* might be a more accurate term. Digital imaging involves transforming the format of information sources, not simply providing a faithful reproduction of these sources on a different medium. The power to enhance, the possibilities for structural indexing, and the mathematics of compression and communication together fundamentally alter the concept of preservation in the electronic era. These capabilities alone, along with the new responsibilities imaging technology places on us as information professionals, will force us to transform our library services and programs in turn.

[KENNEY] Having considered the digital framework, I would like now to turn to preservation's niche in this brave new world. First and foremost, I believe that digital-imaging technology will take its proper place as an alternative or complement to microfilm and photocopy for the reformatting of endangered library materials.

SOME ADVANTAGES OF DIGITAL IMAGING TECHNOLOGY IN PRESERVATION

Duplication Without Degradation

Indeed, digital-imaging technology offers several important preservation advantages over light lens processes. Chief among them is the ability

to duplicate without degradation. A digital image can be reproduced over and over again with absolute fidelity; those strings of 0s and 1s replicate remarkably well–in marked contrast to light lens copies, which suffer a 10 to 15 percent informational loss with each succeeding generation of copy.

Ease of Manipulation and Enhancement

Digital technology also offers the ability to manipulate and enhance images in ways not possible with light lens technology. This includes removing stains, underlining, and bleed-through; increasing legibility by heightening contrast between text and background; and, by segmenting an illustrated page, capturing both text and image in a manner that optimizes both–the illustration as gray scale, reproducing much of the subtlety of tone, and the text as high-contrast black and white, ensuring readable clarity. By comparison, high-contrast light lens technology requires one to choose which to optimize: text or illustration. Newer continuous-tone microfilms are coming onto the market, which may ease this situation, but the fact remains that digital imaging provides infinitely more flexibility than do light lens processes.

Preview Capability

Digital technology also offers the advantage of previewing the image before keeping it so that adjustments can be made to the brightness, contrast, and color balance beforehand. With light lens processes, there are no preview capabilities. If the lighting is incorrect or the image out of focus, the image must be retaken at a later point. It is also easy with digital technology to reorder the sequence of images after the fact. By comparison, microfilm requires laborious splicing in of missing or corrected pages.

Relative Permanence

A further preservation advantage is that digital images do not decay with use, unlike microfilm, which can become scratched with improper handling, or books whose bindings fail or paper deteriorates over time. In fact, as Paul indicated, use in the digital world can become an important preservation consideration as digital media can utilize error detection and correction codes when used. The more they are used the higher the rate of accuracy maintained.

Flexibility in Output

Digital-imaging technology offers great flexibility in output and can thus meet a variety of user preferences and equipment constraints. With digital technology, it is possible to separate the medium for preservation (film) from the medium for production (digital files) and use (paper, film, on-screen display, etc). These various outputs need not be created at the same time. For example, it may be the case that microfilm is created for archival purposes from the digital images at the time the book is scanned. At another time, a print-on-demand copy (which is superior to a printout from microfilm) can be made in response to local user needs, or the digital images themselves can be transmitted over national networks to researchers at distant institutions. This flexibility obviates the need to make choices about the final format at the point of preservation. Such choices can be use or convenience driven. Thus, digital technology combines the desirable quality of hardcopy reproduction with microfilm's ease of duplication, long-term stability, and space saving compression.

SOME DISADVANTAGES

However, there are a number of disadvantages associated with the use of this rapidly changing technology, including obsolescence and incompatibility of hardware and software; an absence of standards for image capture, file format, compression, and transmission; and a lack of experience with library/archival applications. Because digital images are not eye readable but are coded representations, we are solely dependent on supporting the system configuration to "read" the digital files and are thus vulnerable to total loss on a number of fronts.

Intensive Maintenance

It is fortunate that digital images can indeed be copied without loss of information because preservation using digital technology will require maintaining access to the information through periodic copying, refreshing, and migrating the data to keep pace with changes in hardware and software that are characteristic of an emerging technology. The system requirements needed to access the images are likely to change several times over during the life span of the digital media, and the data will need to be copied to newer formats long before the medium itself degrades. Given such a period of flux, one must also consider the wisdom of provid-

ing a human-readable backup to the digital images themselves, either through maintaining the original or creating a paper or film-based copy either before or after digitizing.

Ever-Changing Technology

As will be seen, digital images are extremely large, and their effective use will not only require large storage capacities but will also affect the requirements for user workstations. Low-end PCs and Macs, for example, will not support their storage, decompression, display and manipulation. In the long run, however, the technology itself is the least of our concerns; storage capacity doubles each year, personal workstations are becoming more powerful, dwarfing the big computers of a decade ago, and the costs of conversion and maintenance will decline over time.

Long-Term Commitment of Funds and Resources

The real problem may lie in adjusting institutional mind sets to accept a long-term commitment of funds and resources to accomplish this task. Libraries and archives, in a period of transition, will have to support two systems–the traditional library and the digital library–and the expense, conflicting requirements, and changes required will be difficult. In this period, the preservation administrator's role will not be so much one of technological guru but of watchdog to keep before the eye of the library administration the need to maintain access to information in digital form over the very long haul.

USE AND EASE OF USE

In this time of uncertainty and flux, I believe the role of preservation will be considerably broadened. As resources for supporting our collections diminish, hard decisions will need to be made about what materials to maintain and at what level to provide access to them. Endangered materials will no longer be defined as just those items printed on acidic paper, but will extend to those that languish from lack of use. As researchers become comfortable with online access to sources, they develop rising expectations for improving access to all information and may become increasingly uninterested in materials that are not easily accessible–to the extent that some will restrict their searches to those materials that are most quickly identified and available.

A telling comment was made by one Cornell faculty member whose reaction to the online catalog was, "Don't tell me about information; give me the information itself. That's when I'll be really interested in what you are doing." It does appear that use and ease of use are highly correlated. Despite claims to the contrary, we have seen this occur when deteriorated materials are converted to microfilm or when paper copies are stored off campus. Their use goes down. It may be an apocryphal story, but our math library has been collecting microforms of material it does not own for the past 20 years and has yet to record a single use of that film. Indeed, there have been instances when the faculty have requested interlibrary loan paper copies for items that are available locally on film!

In the future, materials available in our main research libraries may be considered too hard to access and use in their current formats. This change has already begun in technical, professional, and law libraries where researchers have come to equate nearly instantaneous access with relevancy. The conversion and availability online of retrospective sources could prove critical in revitalizing their use and in providing the needed justification for their share of institutional resources. In the not too distant future, the role of the preservation administrator will take on new meaning as retrospective conversion is extended beyond bibliographic information to the sources themselves.

PRESERVATION ADMINISTRATORS–ARBITERS OF QUALITY

In the present, however, there are immediate demands for preservation administrators to become involved in testing, monitoring, and promoting the use of digital technology for preservation purposes. Chief among the many tasks requiring our expertise is the development of benchmarks for image quality. It will fall to us to define them for a wide variety of source materials.

As will be seen in the tutorials to come, image capture will be affected by the electronic format being used, which in large measure will be dictated by the physical format and condition of the material being scanned. We must come to understand the role of resolution, tonal reproduction, enhancement, and compression in the process of defining quality standards.

The preservation community will also have to consider other variables, such as cost of image capture, institutional capabilities and commitments, the quantity and type of material to be captured, the relevance of current standards for reformatting that have been established for photocopy and microfilm, issues associated with fidelity versus legibility, user require-

ments and perceptions, and the uses to which this material will be put. The application and intended end products of the reformatting effort should be the driving force in making decisions on image capture. As John Stokes once remarked, the tendency is to define solutions for image management in terms of equipment or process with insufficient thought concerning objectives–current, medium, and long range.

The DPI Question

Some first attempts have been made in defining quality benchmarks for preservation reformatting. At Cornell we have spent the past four years evaluating the quality of high-resolution binary scanning to capture printed text. From the outset we were interested in determining whether the quality of the digital image was comparable to that obtained through conventional reformatting techniques–and at an affordable price.

After considerable experimentation with the Xerox prototype scanner and an analysis of the printing processes used during the 19th and first half of the 20th centuries, we concluded that 600 dpi scanning represents a sufficient resolution to capture the vast majority of printed material published during the period of paper's greatest brittleness. Thus, while 600 dpi scanning does not provide the resolution that can be obtained through microfilming, it will do, given the nature of the material we want to capture. Books published during the 19th and early 20th centuries were produced using metal type, which has a tendency to spread, so printers were limited to how small or closely spaced letters could be. All common typefaces used during this period were produced at 5- or 6-point type and above. Six hundred dpi binary scanning can adequately capture 4-point type and below, including the rendering of fine detail and uneven thickness that characterize typical fonts used during this period, which were prone to elaborate serifed script, italics, and small body heights.

Most scanning projects to date have utilized a lower resolution–in the range of 200 to 400 dpi–with 300 dpi being the most common. While lower-resolution scanning can produce satisfactory copies from modern texts of 6-point type and above, many of the deteriorating volumes in our project contained irregular features typical of the production typography and printing techniques of the past century and a half, or they were heavily illustrated with fine line drawings and halftones, or came in languages such as Japanese where characters comprised of varying strokes are difficult to reproduce at lower resolutions. Over half of the initial 1,000 volumes scanned came from our math library and contained annotations and formulae that were really challenging to capture. The mathematicians insisted, for instance, that we not turn pluses into minuses in the scanning

process, which actually occurred in a number of cases when using 300 dpi resolution. The 600 dpi copies successfully captured most of these printing challenges to represent faithful and legible reproductions of the originals.

For printed material, then, which consists largely of text and line art, the key to digital image quality is determined primarily by the capture resolution. The higher the number of dots per inch, the better the reproduction, and we feel confident in calling for 600 dpi as the minimum acceptable resolution for preservation replacement purposes. As higher-resolution production scanning systems come on the market, we may well convert to them, but we will not have to go back and rescan the material already captured. It is a chimera to think that we ever could, and so preservation administrators must take care in defining at the outset quality benchmarks for replacement purposes.

The Success of High-Quality Paper Facsimiles

In addition to a technical assessment, we were interested in a subjective analysis as well. Our faculty advisory committee reviewed paper facsimiles and pronounced them of sufficient quality to replace the deteriorating originals, which in most cases were subsequently discarded. Interestingly, in many cases the faculty preferred the scanned facsimile to the original, finding that the slightly heightened contrast between text and background of the new version made them more readable in the age of bifocals. Faculty acceptance of the printed copy is an important point in that we can move to a single preservation cost for replacement purposes. In microfilming projects at Cornell, we have had to return the original volumes to the shelves to satisfy local users. This has resulted in a substantial additional cost to conserve or protect the originals after filming. It is my belief that the use of digital-imaging technology for preservation and the ability to produce a high-quality printed facsimile will lead to increased scholarly support for the national preservation reformatting effort.

These findings on digital quality requirements come from only one institution. The broader preservation community must determine whether there is general consensus regarding Cornell's position and must also move to determine quality benchmarks for other materials–photographs, both black and white and color; archival sources; and works of art on paper–for which resolution alone will not be the only determinant of image quality. This process must be a collaborative one among preservation administrators, the keepers and users of this material, and the providers of digital-imaging equipment and services.

WORKING WITH VENDORS–CRITICAL TO SUCCESS

Which brings me to my last point: vendor relationships. As mentioned earlier, this technology is an emerging one, and vendor capabilities–for providing scanning services of high quality and low cost as well as storage and maintenance services for digital masters–are rapidly developing. The editor of *Imaging Service Bureau News* estimates there are over 2,000 imaging bureaus in the country, one half of which are using digital technologies. To date, very few have established a relationship with the research library community. But this is a chicken-and-egg phenomenon. While there is a great deal of interest in this technology, libraries and archives have been tentative in its use, citing the lack of standards, quality definitions, system support, funding, and technical expertise. This reluctance quite naturally tends to depress the vendor market. It is time now for us to come to understand the value and uses of digital technology and to work with vendors and service bureaus to ensure that they can understand and meet our needs. This relationship will be critical because the vendor role for services will be large. Few institutions will be able to make an in-house system economically viable, considering the flux associated with changing hardware and software requirements.

Margaret Byrne, head of the preservation section at the National Library of Medicine, addressed the National Preservation Planning for Agriculture Conference in 1991 and identified a number of issues that must be resolved before the use of digital technology can become a viable preservation option. Among them was expertise: "It seems to me that the situation today is very similar to that of ten years ago when cooperative preservation microfilming projects were starting up. Specifications for filming brittle bound materials had to be developed and vendors had to learn new ways of doing things. . . . Today we are facing a similar lack of specifications for scanning brittle library materials or for producing preservation quality film from digital files. And library staff must become much more familiar with hardware, software, image capture, and quality control procedures if they are to communicate successfully with the vendors who will do the work."

We hope this symposium will lay the cornerstone in our quest for building that requisite level of expertise to meet the challenges and opportunities posed by the use of digital technology for preservation of and access to our intellectual heritage.