

2010

Preservation in the Age of Google: Digitization, Digital Preservation, and Dilemmas

Conway, Paul

Library Quarterly, vol. 80, no. 1, pp. 61-79 <<http://hdl.handle.net/2027.42/85223>>
<http://hdl.handle.net/2027.42/85223>

PRESERVATION IN THE AGE OF GOOGLE: DIGITIZATION, DIGITAL PRESERVATION, AND DILEMMAS¹

Paul Conway²

The cultural heritage preservation community now functions largely within the environment of digital technologies. This article begins by juxtaposing definitions of the terms “digitization for preservation” and “digital preservation” within a sociotechnical environment for which Google serves as a relevant metaphor. It then reviews two reports published twelve years apart under the auspices of the Council on Library and Information Resources. *Preserving Digital Information* presented an insightful and visionary framework for digital preservation in 1996. *Preservation in the Age of Large-Scale Digitization* explores the implications for preservation practice of the digitization of books and, by implication, our cultural heritage in general. These juxtapositions frame four dilemmas for preservation relating to the impact of environmental storage, new challenges to preservation quality, threats to audiovisual heritage, and an emerging expertise gap. The article concludes with recommendations and observations on making difficult choices.

It’s the end of the world as we know it. (R.E.M.)

Introduction

On November 4, 1966, the Arno River in Florence, Italy, flooded its banks, breaching the basements of museums, libraries, and private residences and burying centuries of books, manuscripts, and works of art in muck and muddy water. The natural disaster of the Florence flood galvanized a fledgling conservation community into action. In the intervening decades, successive generations of preservation professionals have learned to take a

1. This article originated as the keynote address of the “Persistence of Memory” conferences sponsored by the Northeast Document Conservation Center (NEDCC) in 2006–8. The author wishes to thank NEDCC for its support and also recognize two anonymous reviewers for their editorial recommendations.
2. Associate professor, School of Information, University of Michigan, 1085 South University Avenue, Ann Arbor, MI 48109; E-mail pconway@umich.edu.

[*Library Quarterly*, vol. 80, no. 1, pp. 61–79]

© 2010 by The University of Chicago. All rights reserved.

0024-2519/2010/8001-0003\$10.00

global view of the preservation of cultural heritage while immersed in the technical details of materials science—a profession committed to preventing future disasters while focused on sophisticated decision making in support of cost-effective action. The transformative effects of the Florence flood on the conservation and preservation fields are now clear and unambiguous [1–3].

Today, a flood of a different sort is sweeping across the land—a veritable deluge of digital data along with tools and toys designed expressly to tap the flow of digital information for commercial, entertainment, and educational ends. The digital deluge has two streams that converge to give the impression—at least in the more technologically developed places of the world—that we are indeed immersed in an all-digital environment. One source is the collectively massive and accelerating conversion of book and nonbook materials from analog to digital form. The second source derives from the fact that nearly all new information is created digitally, communicated digitally, used in a digital environment, and stored in digital systems, sometimes “for posterity.” As technology devices become increasingly feature rich, usable, and affordable, the proportion of information that makes its way to paper or film is declining, along with the proportion of paper that warrants long-term preservation [4].

Preservation specialists within the broad cultural heritage community of libraries, archives, and museums have explored how to integrate digitization technologies into their suite of reformatting techniques for over a decade while also positioning themselves to embrace the specialized challenges of preserving information in digital form. Commentary lauding new technologies or warning of unintended consequences fills shelves in bookstores; dominates the popular press’s coverage of lifestyle, business, and politics; and sets the agendas of some professional and scholarly associations. The purpose of this article is not to rehearse the open debate between “digerati” and “Luddites” [5, 6] but rather to outline four dilemmas that the preservation community faces in finding a middle path between wild enthusiasm for and deep skepticism about new technologies.

My strategy for tackling this rather large issue is to present a series of juxtapositions. The first of these is definitional, namely, the very important distinction between the terms “digital preservation” and “digitization for preservation,” within a sociotechnical environment for which Google serves as a relevant metaphor. A second juxtaposition reviews briefly two reports published under the auspices of the Council on Library and Information Resources. *Preserving Digital Information* [7] presented, in 1996, an insightful and visionary framework for digital preservation. *Preservation in the Age of Large-Scale Digitization* [8] is a recent exploration of the implications for preservation practice of the digitization of books and, by implication, our cultural heritage in general. These juxtapositions frame the ways that the

cultural heritage sector engages with its technological environment, exposing nearly intractable dilemmas for the preservation community, dilemmas that may only resolve themselves by a fundamental shift in how preservation activities are conceived, operationalized, and funded.

What Is the “Age of Google”?

The preservation enterprise in the cultural heritage sector now functions largely within an environment of digital technologies, organized digital content, and tools to find and use digital information. Because Google and its competitors are such driving forces in this environment, a closer look at what these companies represent to users and preservation professionals is a useful mechanism for exposing the dilemmas that preservation faces today. From one perspective, Google is a powerful technology firm capitalized more generously than most in the United States [9]. The company is leveraging its market share and its wealth to create new browser-based tools, acquire and integrate technology tools built by other companies, and, most recently, build massive collections of digital content through its Google book search initiative (<http://books.google.com/>). For the cultural heritage community, Google may also be viewed as a powerful suite of metaphors reflecting how this community simultaneously is embracing and threatened by a rapid shift from analog to digital information.

Google is a metaphor for the instant gratification expected in information search and retrieval today. For a new generation of users, Google represents anonymous access to information without human mediation. In the environment of social computing, code and interface are the new mediators. Social software provides to builders and contributors alike new incentives, ranging from sheer fun to enhanced reputations [10, 11]. Librarians, archivists, and museum professionals who mediate information services or provide end user training may simply be in the way [12]. Paul Courant argues, too, that “the world of Google’ is emblematic of something more important [than a search engine], namely the world of ubiquitous indexed digitized content” [13].

Perhaps most compellingly, Google is a metaphor for a particular form of digital information, namely, content that is simultaneously fixed and fluid yet decontextualized for use and reuse in ways that may in no way represent the original intent of the creator [14]. The technological environment in which Google thrives is enabling a new digital humanities scholarship—intensely collaborative, interdisciplinary, and enabled by computing tools for finding new meaning through data mining, creative visualizations, and other ways of pushing the boundaries of existing documentation and evidence [15].

The leaders of Google would not necessarily claim to be in the business of new scholarship, but they do claim to be in the cataloging business, with a mission “to organize the immense amount of information available on the web.”³ The business models for Google and its competitors also encompass some elements of preservation, if only through their new responsibilities to manage and deliver bundled content and tools for periods of time not yet defined. What is now clear is that large-scale information providers are competitors with the cultural heritage community in defining what preservation means in the future. In the age of Google, nondigital content does not exist, and digital content with no impact is unlikely to survive.

What Is Preservation?

Preservation is an act of “responsible custody” [16, p. 218], ancient in its origins but with a decidedly modern implementation in the twentieth century, complete with theoretical debates about the relevance of archival thought and social memory [17], as well as a rich and diverse literature [18]. Preservation specialists in libraries, archives, and museums share a commitment to protect the “stuff” of culture in all forms and formats, perhaps even including the built environment that houses the raw material of history and the evidence that feeds our memories.

Twenty years ago, near the peak of the traditional preservation movement, I proposed an influential, research-oriented definition of preservation that highlights its dual nature: one set of actions prevents or slows the inevitable deterioration of mostly organic materials; other actions restore the usefulness of an artifact or information source, usually near the end of its life [19]. I asserted that preservation action should nearly always be taken in reference to use, rather than to the purely intrinsic value of an object, but that in the end, value judgments driven by limited funding are at the heart of the traditional preservation business case. Effective and efficient preservation action also has depended on innovative research and development [20], and the mutually reinforcing relationship between research and practice remains critically important in the digital environment [21].

When approaching the dilemmas of the cultural heritage preservation community in the context of digital technologies, it is important to establish clear distinctions between the terms “digitization for preservation” and “digital preservation.” Digitization for preservation creates valuable new digital products, whereas digital preservation protects the value of those

3. Quoted from corporate information available at <http://www.google.com/corporate/>.

products, regardless of whether the original source is a tangible artifact or data that were born and live digitally. Digitization for preservation and digital preservation are intimately related, but the underlying standards, processes, technologies, costs, and organizational challenges are quite distinct.

The recognition of digitization as a preservation strategy is a relatively new and still-controversial concept within the cultural heritage community, which has generally viewed digitization activities as a form of copying for easier and broader access [22]. The Association for Research Libraries officially adopted digitization as an acceptable preservation strategy, by recognizing the sheer impossibility of redoing digitization work as technology evolves and acquiescing to the notion that technology capabilities and standards are sufficiently mature for the task at hand [23]. If digitization for preservation is defined solely as a set of technical specifications for full information capture [24], then the cultural heritage community neither requires nor can afford to employ preservation-quality digital conversion across the board. It is far more useful, perhaps, to define digitization for preservation as activities that result in the creation of digital products worthy of long-term preservation [25]. These activities include the decisions to select materials for digitization, create digital representations with full and accurate descriptions, and then compile images and text into a coherent product.

If digitization for preservation is an investment in the creation of lasting digital products, then digital preservation is best viewed as the suite of tools, operations, standards, and policies that help ensure that this investment is not squandered.⁴ Digital preservation is informed by digital collection building but encompasses the acquisition, ongoing maintenance, periodic transformation, and persistent delivery of digital assets. Once digital resources are identified as having long-term value, then the technology systems and accompanying policy frameworks that preserve those digital assets must inspire the same level of trust and confidence in users and stakeholders as do traditional preservation and access services. Ross Atkinson predicts that the outcome of digital preservation actions will be “new special collections, which, like its traditional counterpart, will be very expensive and will be warranted for only a small, selected subset of publications” [26, p. 172].

Preserving Digital Information

Since its publication, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* has become a touchstone in digital pres-

4. *Library Trends* 54 (Spring 2005) is a special issue that demonstrates the diversity of digital preservation best practices and program innovation.

ervation [7]. The report's immediate focus was on the challenges of preserving the integrity of materials already in digital form and retaining the ability of users to retrieve, display, and otherwise use these materials in the face of constantly changing technology. The task force that developed the recommendations was composed of an array of individual talent drawn from the cultural heritage community, as well as the information industry, construed broadly enough to encompass publishers and scholarly societies. Most of the twenty-one members of the task force continue to have an impact on the preservation field. For example, six members now have significant leadership roles in academic libraries, two have senior faculty appointments, and three others conduct research on digital preservation issues. The cochairs of the task force, Donald Waters and John Garret, have had a direct impact on the international digital preservation agenda. Waters served as the first executive director of the Digital Library Federation and is now a program officer at the Andrew W. Mellon Foundation. Garrett is affiliated with NASA's Goddard Space Flight Center and has been a key player in the development of international archiving standards.

Preserving Digital Information is densely argued and complex in its line of reasoning. Among its many conceptual breakthroughs, however, the report joins the notion of "digital archives" with the concept of preservation, thereby distinguishing the challenge of long-term preservation from the ongoing development of digital libraries that may or may not consider preservation to be a mandate. In placing first responsibility for preservation on the creators of digital information, paralleling similar thinking in the Australian library community, the report greatly expands the playing field of stakeholders, potentially lessening the traditional monopoly on preservation responsibility held by cultural heritage organizations. The report argues for the development of a widely distributed deep infrastructure to support digital archives. It insists that truly trusted digital archives require a system of third-party certification. Perhaps most radically, the report proposes that digital archives rescue digital information threatened with obsolescence, when the original creators either abandon or fail to preserve the information, posing a direct challenge to mainstream thinking regarding the rights of intellectual property ownership [7, p. 40].

In the decade since *Preserving Digital Information* appeared, significant progress in at least three important areas can be most clearly traced to its recommendations. First, the report's call for the development of "deep infrastructure" helped initiate the development of "cyberinfrastructure" and e-science programs that encompass the technological, socioeconomic, and organizational requirements for using and preserving digital information to further scholarship. Cyberinfrastructure is to the knowledge economy what the infrastructure of railroads, superhighways, and power and telecommunications grids have been to the industrial economy [27,

28]. Second, in placing integrity at the heart of the digital preservation enterprise, the report allies digital preservation with the discipline of archival science, which Luciana Duranti defines as comprising “the ideas about the nature of archival material (i.e., archival theory) and the principles and methods for the control and preservation of such material (i.e., archival methodology)” [29, p. 39]. Third, the Waters-Garrett report identified and defined five distinctive archival facets of integrity: formatted and structured bits (content) “frozen” as discrete objects (fixity) in a predictable location (reference) with a documented chain of custody (provenance) and linkages to related objects (context). This framework of integrity in *Preserving Digital Information* inspired the preservation metadata specification of the Open Archival Information System (OAIS) reference model [30, 31], which in turn spurred development of new preservation metadata schemas [32] and related trusted repository system development [33, 34]. Only the report’s proposed rescue function remains unexplored and untested.

Preserving Digital Information is one of the most successful efforts in the field of preservation. After an initial delay, the report and the people associated with it fueled new preservation initiatives that now have deep technical and organizational roots. This progress, although by no means constituting a “solution” to the digital preservation challenge, has in many ways forcefully captured the limited attention of cultural heritage resource allocators, thereby forcing an assessment of strategic priorities.

Preservation in the Age of Massive Digitization

Just over a decade after the preservation community started a conversation about the nature of digital preservation, another conversation is taking place over the implications and impact of the massively scaled digitization of books from research libraries in the United States and elsewhere. Spurred by Google’s book search partnerships with publishers and libraries and the University of Michigan’s decision to contribute its entire library of books and serials to the digitization effort, it is possible that the majority of the published record held in the United States may exist in digital form ten years from today. *Preservation in the Age of Large-Scale Digitization*, written by Cornell University’s Oya Rieger, is an effort to examine the implications of the digitization of books on the preservation of books themselves and their digital surrogates [8]. The report is a benchmark examination of the transition that cultural heritage institutions are making toward a service model dominated by, perhaps even determined by, digital information technologies. The digitization of books and historical records is the activity that first engaged research universities, and national archives and libraries with next-generation technologies, almost a decade before the advent and dispersion of the Web [35].

Preservation in the Age of Large-Scale Digitization provides a useful summary of the major efforts to affect the wholesale conversion of books and serials from paper to digital form, bringing improved search and retrieval of the digital content and advances in the delivery of content through the Internet. The report compares and contrasts four digitization programs and concludes that they “challenge some of the prevailing [preservation] practices not only because of their scale but also because of the transformation of digital library technologies and user preferences” [8, p. 19]. The report usefully distinguishes between the processes and procedures of digitization and the requirements for preserving the digital books once conversion is complete.

The heart of the report is a four-part framework for analyzing digitization efforts, including selection for digitization, content creation, technical infrastructure, and organizational infrastructure. The report contains thirteen recommendations for the advancement of large-scale digitization efforts, regarding improving or adjusting product quality, service contracts, access to digitized content, preserving digital surrogates, and library priorities. Rieger sees the greatest impact of the digitization efforts on traditional book collections, including the prospect that most library books will reside in remote storage facilities, traditional preservation functions will reorient to support digitization quality control, and the artifactual value of books will pale in comparison to their digital surrogates—readily available for print-on-demand services but stripped of their tangible connection to their origins and history.

The report provides direct and compelling evidence on how large-scale digitization is confronting the preservation community with choices about priorities, as well as its fundamental values. The true significance of this report—and the preservation community’s reaction to digitization in general—is what it says about new relationships with digitization stakeholders and the place of quality in preservation theory. In this regard, Rieger demonstrates quite compellingly that the digitization of books places the preservation community in a nearly impossible bind.

Dilemmas for Preservation of Cultural Heritage

Preserving Digital and Information and *Preservation in the Age of Large-Scale Digitization* bracket a decade of extraordinary technological change that has now clearly precipitated a fundamental change in preservation perspectives. Over a decade ago, I predicted this transformation: “The digital world transforms traditional preservation concepts from protecting the physical integrity of the object to specifying the creation and maintenance

of the object whose intellectual integrity is its primary characteristic” [36, p. 3]. By incorporating commitments to large-scale digitization and digital preservation into a practice that remains deeply committed to the protection and care of cultural and historical artifacts, preservation practice is on the cusp of a fundamental transformation that will continue into the foreseeable future.

On the path to transformation are many choices regarding how preservation practice is conceived, organized, and funded. Some of these choices take the form of genuine dilemmas for the cultural heritage preservation community—dilemmas that have been neither adequately defined nor appropriately addressed. The *Oxford English Dictionary* (2nd ed., 1989) defines a dilemma as the “choice between two (or, loosely, several) alternatives, which are or appear equally unfavorable; a position of doubt or perplexity, a ‘fix.’” Four of the most knotty challenges are described below, followed by some recommendations for resolving them.

The Environmental Dilemma

The first dilemma for preservation derives from a success story, namely, the broad acceptance of proper environmental storage conditions as the first and most cost-effective preservation strategy [37]. Spurred by technically compelling preservation science carried out by the Image Permanence Institute, the Library of Congress, and the Smithsonian Institution, preservation researchers first diagnosed the root causes of the degradation of paper, film, magnetic tape, and other organic media and then demonstrated with a high degree of confidence that creating cool, dry, dark, and stable storage conditions could extend the useful life of the holdings of archives, libraries, and museums across the board [38]. With the leadership of Paul Banks at the Newberry Library in Chicago and then Harvard University Library, research universities pioneered low-cost, high-density storage facilities, essentially codifying the mechanisms for preservation storage coupled with cost effective access [39]. Other universities followed suit, building many dozens of storage modules across the United States and allocating precious renovation dollars toward retrofitting and upgrading existing stack areas [40].

This recent and significant investment in buildings and renovation for preservation purposes places the preservation community in a bind. Preservation storage is so cost effective and carries with it such a dramatically positive impact on the life of organic materials that it is exceedingly difficult to defend the allocation of financial resources for physical conservation treatments, preservation rebinding, special collections rehousing, and almost every other preservation action that involves the handling of individual physical artifacts. Provosts, boards of directors, and chief executive

officers alike are cutting preservation budgets or reorienting expenditures to digital initiatives, based in large part on the demonstrated impact of the preservation environment on university assets.

The clear need to place the funding of environmental control for collections storage at the top of the funding priority list creates a major dilemma for the cultural heritage preservation community about how to allocate any other preservation resources that may be available. The preservation community may have to choose between undertaking physical conservation treatments and collection care or investing in digitization for preservation. This choice presents a dilemma, in part, because resources for preservation are meager in small and medium-sized heritage organizations. The Heritage Health Index reported in 2005 that 26 percent of cultural heritage organizations in the United States have no environmental control of their collection storage areas, and two-thirds have less than \$3,000 per year allocated for all preservation activities [41, p. 73]. On the other side of this “preservation divide,” inflation-adjusted annual funding for preservation in the 111 largest American research universities, totaling over \$112 million in 2008 alone, has remained flat for a decade in research universities and could actually decline in the years ahead [42, pp. 13–14].

The choice embedded in this dilemma has consequences for cultural heritage organizations. To reduce support for item-level treatments of any kind, in the absence of a full commitment to preservation environments, threatens the value of original artifacts. To fail to embrace digitization for preservation of collections, either locally or in collaboration with others, is to risk organizational obsolescence.

The Quality Dilemma

The second dilemma emerges from a tension over how the principle of quality applies to digital preservation and digitization for preservation. For a generation or more, “preservation quality” has been an adjective that advocates of traditional preservation have applied to media, materials, processes, and outcomes. Preservation-quality microfilm and alkaline paper are distinctive for their longevity. Preservation-quality bindings endure more punishing use than trade bindings. Preservation-quality reformatting techniques—film duplication, photocopying, and digitization alike—are designed with a goal of full information capture for a given recording media. The ethics of preservation essentially have made inseparable the notion of “quality” and actions to extend the life expectancy and usefulness of information. Indeed, one of the principal reasons that preservation specialists in the early 1990s were able to accept the risks of newly marketed digital-imaging technologies was the promise of a much higher level of reproduction quality than other well-established, standards-compliant technologies. For conservators, treatment quality is a core goal of the profes-

sion's code of ethics: "While circumstances may limit the resources allocated to a particular situation, the quality of work that the conservation professional performs shall not be compromised" [43].

Does the digital world render obsolete those preservation ethics that embrace maximum quality and full information capture? Recent thinking on this matter from the cultural heritage community contains mixed messages, demonstrating an ongoing struggle with balancing traditional commitments to quality with fiscal and technological realities. The Association of Research Libraries asserts that digitization technologies may replace preservation microfilm because "there are established guidelines for image capture and processing to ensure that images are of high quality and provide faithful representations of the original" [23, p. 2]. Rieger's report on mass digitization efforts reaffirms the importance of image quality but calls for a concerted effort by the preservation community to revisit and revise its outmoded practices best suited for "hand crafted" scanning projects [8, pp. 33–34]. A conference on the implications of large-scale digitization programs on library special collections programs concluded that librarians and archivists should set aside well-understood guidelines on image quality to increase digitization productivity for special collections projects [44, p. 6]. Kalev Leetaru's recent comparison of image quality in two mass digitization projects supported by library preservation programs raises quite serious concerns about the simple usability of the resulting page images and the long-term value of these large-scale projects [45].

Until recently, those who built preservation repositories also exercised significant control over the creation of digital content by specifying digitization best practices. Increasingly, however, digital preservation repositories take what they can get, with, at best, assurances that the submitted content has been validated for quality. Repository best practices call for the inspection of the technical characteristics of deposited objects, in part to detect file formats, in part to check that deposited objects are accurately described, and sometimes to normalize data structures and formats [34]. Even when there is documentation on digitization processes or vendor quality assurance routines, questions do arise about the gap between the capabilities of the repository/content and the expectations of users. For an institution and its community of users to trust that individual digital objects are accurate, complete, and intact and to know that deposited objects have the capacity to meet a variety of uses envisioned for them by different stakeholders, another step will be required. Digital preservation repositories may need to validate the quality and fitness for specific purposes of individual objects they preserve.

The principal dilemma for the cultural heritage preservation community is to sort out a new stance on digitization quality while extending the commitment to preservation quality of the content preserved in digital

repositories, when preservation specialists generally either have little lead responsibility for digitization or sometimes little or no authority over digital preservation specifications. As with the environmental dilemma, the choice may be to find the “least bad” option. Compromises on image and text quality in digital conversion may increase workflow productivity but at the cost of decreasing the overall-use value of the resulting digital products. Upholding or even increasing digitization quality standards may increase cost or diminish the rate and quantity of digitization in an era of declining resources and increased pressure to migrate online. In the realm of digital preservation, increased attention to content quality may lead to contentious decisions not to preserve digital assets that do not meet quality standards. Lack of attention to internal content quality may result in digital preservation repositories that protect bit streams whose intellectual content is of little long-term value.

The Nonbook Dilemma

The third dilemma derives from the success that the preservation field has had in mitigating the problems associated with the acidic paper production. After decades of creative materials-science research backed by multiple preservation strategies, including preservation reformatting, large-scale paper deacidification, and improved storage, the preservation community has essentially rescued 150 years of paper-based cultural heritage and begun to turn the tide against continued loss. The nearly worldwide adoption of alkaline-buffered paper is cause for much optimism about the prospects for greatly extended life expectancies of the published record in the future. In contrast to the clear successes of paper preservation, it is now obvious that magnetic audiotape and videotape and acetate-based films are the great preservation challenges of the twenty-first century [46, 47].

The dilemma for the preservation community is in the choice between accepting the consequences of benign neglect, on the one hand, and accepting the risks of digitizing audiovisual resources in the absence of comprehensive diagnostics and digitization standards, on the other. Magnetic tape and several generations of motion picture film suffer from the same built-in obsolescence that characterizes acidic paper [48]. In spite of a century of film vault fires (some nearly as culturally devastating as the Florence flood) and the existence of vast collections of unreadable and poorly documented audiovisual resources, the preservation community has not made the sort of transformative progress with standards, best practices, appraisal and selection strategies, and other procedural requirements that undergird past successes with book and paper preservation. The profound irony with the preservation of audiovisual materials is that new digital technologies are essentially rendering tape and film media obsolete at precisely the time when societal demand for historical pictures, recorded

sound, and motion pictures is unprecedented. A decision to base the preservation of audiovisual resources on environmental storage alone will likely result in a significant portion of the sound and moving image records from the 1930s to the 1990s being lost.

The larger implication of the nonbook dilemma for the cultural heritage preservation community may rest with professional identity. Can a profession raised and nurtured on the care of books and paper reorient its identity to give equal attention to film and magnetic media? With increased attention to visual and aural cultural resources comes the need to recruit and retain a new generation of preservation specialists whose orientation is toward the development and use of multimedia resources. The challenge of change is one not only of tradition and worldview but also of expertise.

The Expertise Dilemma

Both digital preservation and digitization for preservation carry with them a level of technical complexity that the preservation community has not had to face since attempting to develop very large-scale deacidification plants in the 1980s, along with the subsequent cost overruns, explosions, and setbacks [49]. The workflow challenges of outsourcing large-scale digitization projects are daunting. Large-scale quality control requires a mastery of statistical-sampling techniques and sophisticated signal-processing routines. The underlying infrastructure of digital preservation resides in the science of complex relational databases, distributed storage networks, and application development on the cutting edge of computing.

Formal educational programs are emerging to begin filling the expertise gap. Under the auspices of the DigCCurr program, the University of North Carolina at Chapel Hill is leading an effort to define an openly accessible graduate-level curriculum to prepare students to work in the field of digital curatorship, which is roughly but not exactly equivalent to the concepts of digital preservation defined in this article. The project has developed a matrix of knowledge and competencies and the overall elements of a curriculum [50]. A separate collaborative effort is developing a modular curriculum for training graduate students in the larger field of digital librarianship [51]. Both initiatives are comprehensive in their educational vision and are developing specifications and content that may, in the long run, create a new generation of experts capable of functioning in a mostly digital environment while mastering and upholding the core principles that have motivated preservation for several generations.

The cultural heritage preservation community has only begun to articulate or understand the implications of aggressive training, recruitment, and technical collaboration on the management of preservation in the cultural heritage sector [52]. The dilemma for the cultural heritage preservation community derives from the lag between immediate need and

the long-term transformation of digital preservation expertise. The choice now is between retraining or recruiting new staff with new technology skills and outsourcing critical digital preservation functions to experts at arm's length either inside or, collaboratively, well beyond the organization. Recruiting assumes that digital preservation responsibilities rest with preservation administrators who in essence take on the mantle of systems administration. Outsourcing requires the preservation community to trust technology development to experts who may see what they do on a day-to-day basis as risk management rather than long-term access.

Recommendations and Conclusion

What can those of us who are so committed to the idea of preserving cultural memory do to embrace the innovative aspects of the digital world while preserving the values that have motivated stewardship efforts for generations? Cultural heritage organizations must create clear distinctions between efforts to create new digital content by digitizing substantial parts of their collections and efforts to preserve digital data for use into the future. For most cultural heritage organizations, the dilemmas of preservation will require choices regarding resource allocation between creating digital products worth preserving (digitization for preservation) and investing in the long-term preservation of digital assets (digital preservation). Only the very largest organizations can address both areas at a scale worthy of the need. Everyone else has to choose.

Digital ubiquity presents the entire cultural heritage sector with an opportunity. When expectations for digital access reach the point where only digital information will satisfy the vast majority of user needs—as may already be the case in universities, corporations, and government agencies—then libraries, archives, and museums must digitize extensively and to a level of quality that supports a wide variety of actual and potential uses. In this regard, a decision to digitize is a commitment to create coherent products. From a preservation viewpoint, digitization activities must create lasting value—value that is embedded in the digital collections themselves rather than primarily derived from their association with original source materials. The creation of new digital collections that have long-term value is a new and truly appropriate business function that creates immediate value for users and long-term value for the organization. When the construction of digital collections through digitization activities is embraced as a core service function, then preservation initially can be reconceived as those functions that help guarantee that more value is added through the digitization process than is lost through the transformation of original resources to a new medium.

If digitization is an investment in the creation of lasting digital products, then digital preservation is a value proposition, informed by immediate value assessments and the need to guarantee persistent access but motivated by the desire to facilitate a form of generational transfer that the preservation mandate has expected for centuries. Where does this recommended trade-off between digitization and digital preservation leave traditional preservation practices? At a practical level, cultural heritage organizations have some tough decisions to make about their preservation priorities and how to allocate limited resources in new directions. Here is a set of recommendations derived from the choices implied by the four preservation dilemmas outlined above:

- Continue to give pride of place to preservation quality environments and the buildings required to sustain them. Perhaps for decades to come, material culture artifacts will serve as the ultimate backups for their digital surrogates. Preservation environments buy time for careful decision making and represent highly tangible commitments to long-term preservation.
- Shift precious preservation resources away from the item-level treatment of books to the digital rescue of audiovisual resources. Even though serious questions remain to be resolved about the quality and usefulness of millions of digitized book and serial volumes, the crisis of audiovisual media deterioration cannot wait for a full answers to emerge.
- Embrace digital technologies for collection building and collection management. The era of experimentation with imaging technologies has ended, and the time has come to build digital conversion factories that are managed collaboratively by the cultural heritage community. Collaboration is also essential in the development of preservation repositories, which should not depend for their existence on resources allocated by any single organization.
- Orient digitization decisions toward increasing impact and distinctiveness at home while leveraging digitization work already completed elsewhere. When digitization projects are small and experimental, it is perhaps justifiable to invest significant intellectual energy in devising sophisticated strategies for selecting a small subset of a collection for digital conversion. In the face of large-scale digitization, the cumulative effect of many simple and relatively arbitrary decisions to digitize high-use or high-value resources in a richly collaborative environment should relegate “selection angst” to the past.
- Reassert leadership regarding standards and best practices for digital collection building on par with the kind of leadership that now exists in the digital preservation arena. Digital preservation has benefitted

more directly than digitization for preservation from the development and promulgation of innovative research agendas. A similar effort to articulate the research and development requirements underpinning digitization for preservation would assist cultural heritage organizations to collaborate as a community.

We are approaching the end of digital reformatting for all but the rarest and most specialized books. What we are experiencing is a shift in preservation away from its nearly exclusive focus on books and paper to new priorities on other formats. We are at the end of the era of “boutique” digital scanning projects for which the principal goal is experimentation with new technologies and extraordinary attention to the unique properties of each artifact. Finally, we may be witnessing the end of preservation as we know it. In the digital environment in which the cultural heritage community will increasingly find validation and new relevancy, the previously specialized, localized, and almost marginalized field of preservation is on the cusp of becoming the universally endorsed activity that Pelham Barr called for over sixty years ago. There is no small irony in the fact that digital technologies forced this hand and may well prove to be preservation’s salvation.

Seemingly intractable dilemmas often invite a choice between compromise and inertia in an environment of uncertainty. The dilemmas of preservation within the cultural heritage community of librarians, archivists, curators, and associated professions are no exception. Michael Stipe of the band R.E.M. has something to say about this. He sings about the mixed-up, chaotic world that we live in, capturing in a blur of word images the emotions of the nightly news and the energy of new media, blurting a refrain of change and, ultimately, optimism [53]:

It’s the end of the world as we know it.
It’s the end of the world as we know it.
It’s the end of the world as we know it and I feel fine.

It is the end of preservation as we know it, too. But the future of preservation in the age of Google is precisely where it has always been: transforming artifacts into useful new forms and extending their useful life.

REFERENCES

1. Ogden, Shereilyn. “The Impact of the Florence Flood on Library Conservation in the United States of America: A Study of the Literature Published, 1956–1976.” *Restaurator* 3 (1979): 1–36.
2. Waters, Peter. “The Florence Flood of 1966 Revisited.” In *Preserving the Word, 103rd Library*

- Association Conference Proceedings*, edited by R. E. Palmer. London: Library Association, 1987.
3. Ray, Emily. "The Prague Library Floods of 2002: Crisis and Experimentation." *Libraries and the Cultural Record* 41 (Summer 2006): 381–91.
 4. Lyman, Peter, and Varian, Hal R. "How Much Information? 2003." School of Information Management and Systems, University of California, Berkeley, 2003. <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>.
 5. Negroponte, Nicholas. *Being Digital*. New York: Vintage, 1996.
 6. Tenner, Edward. *Why Things Bite Back: Technology and the Revenge of Unintended Consequences*. New York: Vintage, 1996.
 7. Waters, Donald, and Garrett, John. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*. Washington, DC: Commission on Preservation and Access and the Research Libraries Group, 1996.
 8. Rieger, Oya Y. *Preservation in the Age of Large-Scale Digitization: A White Paper*. Washington, DC: Council on Library and Information Resources, 2008.
 9. Stross, Randall E. *Planet Google: One Company's Audacious Plan to Organize Everything We Know*. New York: Free Press, 2008.
 10. Nov, Oded. "What Motivates Wikipedians?" *Communications of the ACM* 50 (November 2007): 60–64.
 11. Lessig, Lawrence. *Code: And Other Laws of Cyberspace*. New York: Basic, 1999.
 12. Lippincott, Joan. "Net Generation Students and Libraries." In *Educating the Net Generation*, edited by Diana G. Oblinger and James L. Oblinger. Boulder, CO: EDUCAUSE, 2005.
 13. Courant, Paul N. "Scholarship and Academic Libraries (and Their Kin) in the World of Google." *First Monday* 11, no. 8 (August 2006).
 14. Levy, David M. "Fixed or Fluid? Document Stability and New Media." In *Proceedings of ECHT '94, the ACM European Conference on Hypermedia Technology, Edinburgh, September 18–23*, pp. 24–31. New York: Association for Computing Machinery, 1994.
 15. McCarty, Willard. *Humanities Computing*. New York: Palgrave Macmillan, 2005.
 16. Barr, Pelham. "Book Conservation and University Library Administration." *College & Research Libraries* 7 (July 1946): 214–19.
 17. Blouin, Francis X., Jr., and Rosenberg, William G. *Archives, Documentation and Institutions of Social Memory: Essays from the Sawyer Seminar*. Ann Arbor: University of Michigan Press, 2006.
 18. Schnare, Robert E., Jr.; Swartzburg, Susan G.; and Cunha, George M. *Bibliography of Preservation Literature, 1983–1996*. Lanham, MD: Scarecrow, 2001.
 19. Conway, Paul. "Archival Preservation: Definitions for Improving Education and Training." *Restaurator* 10 (1989): 47–60.
 20. Cummings, Martin M. *Influencing Change in Research Librarianship: A Festschrift for Warren J. Haas*. Washington, DC: Council on Library Resources, 1988.
 21. Hedstrom, Margaret. *It's about Time: Research Challenges in Digital Archiving and Long-Term Preservation*. Washington, DC: Library of Congress and National Science Foundation, 2003.
 22. Puglia, Steven; Reed, Jeffrey; and Rhodes, Erin. *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files—Raster Images*. Washington, DC: Digital Library Federation, 2005.
 23. Arthur, Kathleen; Byrne, Sherry; Long, Elisabeth; Montori, Carla Q.; and Nadler, Judith. *Recognizing Digitization as a Preservation Reformatting Strategy*. Washington, DC: Association of Research Libraries, 2004. http://www.arl.org/bm~doc/digi_preserv.pdf.
 24. Frey, Franziska S., and Chapman, Stephen. "Developing Specifications for Archival Digital Images." In *IS&T 2001 PICS Conference Proceedings*, pp. 166–71. Springfield, VA: Society for Imaging Science and Technology, 2001.
 25. Conway, Paul. "Overview: Rationale for Digitization and Preservation." In *Handbook for*

- Digital Conversion Projects: A Management Tool for Preservation and Access*, edited by Maxine K. Sitts. Andover, MA: Northeast Document Conservation Center, 2000.
26. Atkinson, Ross. "Transversality and the Role of the Library as Fair Witness." *Library Quarterly* 75 (April 2005): 169–89.
 27. Atkins, Daniel E. *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Washington, DC: National Science Foundation, 2003.
 28. ACLS (American Council of Learned Societies). *Our Cultural Commonwealth: The Final Report of the American Council of Learned Society's Commission on Cyberinfrastructure for the Humanities and Social Sciences*. New York: ACLS, 2006.
 29. Duranti, Luciana. "The Impact of Digital Technology on Archival Science." *Archival Science* 1 (2001): 39–55.
 30. Lavoie, Brian. "The Open Archival Information System Reference Model: Introductory Guide." DPC Technology Watch Report 04-01, Digital Preservation Coalition, London, 2004.
 31. Higgins, Sarah, and Semple, Najla. *OAIS Five-Year Review: Recommendations for Update*. London: Digital Curation Center and Digital Preservation Coalition, 2006.
 32. Caplan, Priscilla, and Guenther, Rebecca. "Practical Preservation: The PREMIS Experience." *Library Trends* 54 (Summer 2005): 111–24.
 33. *Trusted Digital Repositories: Attributes and Responsibilities*. Mountain View, CA: Research Libraries Group, 2002.
 34. Center for Research Libraries. *Trustworthy Repositories Audit and Certification (TRAC): Criteria and Checklist*. Chicago: Center for Research Libraries, 2007.
 35. Borgman, Christine. *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*. Cambridge, MA: MIT Press, 2000.
 36. Conway, Paul. *Preservation in the Digital World*. Washington, DC: Commission on Preservation and Access, 1996.
 37. Reilly, James M.; Nishimura, Douglas W.; and Zinn, Edward. *New Tools for Preservation: Assessing Long-Term Environmental Effects on Library and Archives Collections*. Washington, DC: Commission on Preservation and Access, 1995.
 38. Porck, Henk J., and Teygeler, Rene. *Preservation Science Survey: An Overview of Recent Developments in Research on the Conservation of Selected Analog Library and Archival Materials*. Washington, DC: Council on Library and Information Resources, 2000.
 39. Nitecki, Danuta A., and Kendrick, Curtis L. *Library Off-Site Shelving: Guide for High-Density Facilities*. Englewood, CO: Libraries Unlimited, 2001.
 40. Payne, Lizanne. *Library Storage Facilities and the Future of Print Collections in North America*. Dublin, OH: Online Computer Library Center, 2007.
 41. Heritage Preservation. *A Public Trust at Risk: The Heritage Health Index Report on the State of America's Collections*. Washington, DC: Heritage Preservation, 2005.
 42. Young, Mark, and Kyrillidou, Martha. *ARL Preservation Statistics, 2005–2006: A Compilation of Statistics from the Members of the Association of Research Libraries*. Washington, DC: Association of Research Libraries, 2007.
 43. American Institute for Conservation of Artistic and Historic Works. "Code of Ethics and Guidelines for Practice, 1994." <http://www.conservation-us.org/>.
 44. Erway, Ricky, and Schaffner, Jennifer. *Shifting Gears: Gearing Up to Get into the Flow*. Dublin, OH: Online Computer Library Center Programs and Research, 2007. <http://www.oclc.org/programs/publications/reports/2007-02.pdf>.
 45. Leetaru, Kalev. "Mass Book Digitization: The Deeper Story of Google Books and the Open Content Alliance." *First Monday* 13 (October 2008).
 46. Edmondson, Ray. *Audiovisual Archiving: Philosophy and Principles*. Paris: UNESCO, 2004.

47. Gracy, Karen F. *Film Preservation: Competing Definitions of Value, Use, and Practice*. Chicago: Society of American Archivists, 2007.
48. Smith, Abby; Allen, David Randal; and Allen, Karen. *Survey of the State of Audio Collections in Academic Libraries*. Washington, DC: Council on Library and Information Resources, 2003.
49. Porck, Henk J. *Mass Deacidification: An Update on Possibilities and Limitations*. Washington, DC: Council on Library and Information Resources, 1996.
50. Lee, Christopher. "Curation Junction, What's Your Function? Defining What It Means to Do Digital Curation in Order to Teach It: Proceedings of the 2009 Annual Conference of the Association for Library and Information Science Education (ALISE), 2009." <http://ils.unc.edu/digccurr/alise2009-lee.html>.
51. Pomerantz, Jeffrey; Wildemuth, Barbara; Fox, Edward A.; and Yang, S. "Curriculum Development for Digital Libraries." In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 175–84. New York: Association for Computing Machinery, 2006.
52. Gracy, Karen F., and Croft, Jean Ann. "Quo Vadis, Preservation Education? A Study of Current Trends and Future Needs in Graduate Programs." *Library Resources and Technical Services* 50 (October 2006): 274–95.
53. R.E.M. "It's the End of the World as We Know It (and I Feel Fine)." In *Document*. I.R.S. Records (EMI-Capitol Entertainment Prop), 1987.