Symposium: Digital Archives: Navigating the Legal Shoals

If Only We Could Reach the Shoals: Barriers to Archives Digitization

William J. Maher[*]

When I first joined the archival profession over thirty years ago, I noticed how my new colleagues seemed to be hallucinating. Witnessing the power of mainframe computers and imagining the possibilities of bibliographic networks, they had visions about how this technology might liberate us so we could actively connect our archives to users and potential users—students, scholars and the general public—without being bound by the physicality of the archives. While most thought in terms of automating finding aids and indices, a few even dared to imagine that new technologies might also allow us to store, retrieve and deliver the full text of archival documents. I have to admit that the notion that somehow computer terminals could be linked to a network to enable even small archives to "broadcast" archival documents worldwide was a bit beyond my imagination. Where was the technology, the infrastructure and the mass market to implement such a system?

In light of all we have experienced since the advent of ever cheaper and ever more powerful "personal" computers and since the World Wide Web burst onto the scene, it seems clear that mine was a failure not of vision but of imagination. However, in the years right after the new U.S. copyright law added unpublished material to its coverage, set aside most complex notice and formality requirements, added fair use provisions and provided for library and archival copying, no one could have imagined that copyright issues would emerge as infinitely greater barriers than technology on any voyage to the dreamed-of lands where we would be able to make our records broadly and freely available. Some recent developments, such as the semi-promising but stalled efforts to enact an orphan works exception or to amend § 108, are encouraging, but it is clear that the voyage to archival

[*]   William J. Maher is University Archivist and Professor of Library Administration at the University of Illinois at Urbana-Champaign (UIUC). He was Assistant University Archivist at UIUC (1977–85 and 1986–95) and Program Officer at the U.S. National Endowment for the Humanities (1985–86).   He was President (1997–98) and Treasurer (1991–94) of the Society of American Archivists. In addition, Mr. Maher was President (1987–89) and Secretary-Treasurer (1981–85) of the Midwest Archives Conference. He holds degrees from Case Western Reserve University (1972), Washington University (1975) and UIUC (1991). As the author of one book and over twenty articles, he is a regular speaker on archival administration and copyright law.

digitization involves many significant navigational challenges, some of which will be the focus of today's symposium.[1]

As difficult as the legal issues are, however, for many archivists digitization may be only one aspect of the present challenges and opportunities they face. To set the context so we can best understand the legal challenges involved with digitization, first we need to understand the range of issues that contemporary archivists face. That means we first must clarify what it is we mean by, or could mean by, the term "archives." The Society of American Archivists' *Glossary of Archival and Records Terminology* provides six different meanings, the three most relevant of which are: 1) "Materials created or received by a person, family, or organization . . . in the conduct of their affairs."[2] These are preserved as evidence of the functions and responsibilities of their creator, or because of the enduring value contained in the information they contain.[3] 2) "The division within an organization responsible for maintaining the organization's records of enduring value," or "an organization that collects the records of individuals, families, or other organizations."[4] 3) The building or facility housing archival materials.[5]

Classically, though, "archives" refers to the organized noncurrent records of an institution or organization, which are retained for their continuing value in providing evidence or information on institutional functions or on those affected by the organization. For managerial purposes, there is value in understanding the traditionalist's distinction between the concepts of "archives" as the repositories of official records of an organization and "manuscript collections," which are collected or generated by private individuals and subsequently donated to or acquired by another institution. For the copyright issues before us today, however, this distinction fades; so we can use the term "archives" broadly to refer to published and unpublished—currently a very problematic term of limited utility—documentary material created on behalf of both organizations and by individuals for personal reasons.

Archives may be in any physical format, including handwritten or typescript text, formal publications and "ephemera," photographs, sound recordings, motion pictures and audiovisual recordings and electronic records ranging from punched cards, backup tapes, disks, thumb drives and SAN servers. From a copyright perspective, the works found in archives range from factual to fiction, from mechanical to creative.

As such, archives are found in all kinds of institutions: national, state and local governments; for-profit businesses; not-for-profit organizations; labor unions and professional societies; universities and colleges; research libraries; and large and small private historical societies. Although archives, especially the archetypical institutional archives, perform an important role in supporting institutional and

---

1. 17 U.S.C. § 108 (2006).
2. RICHARD PEARCE-MOSES, A GLOSSARY OF ARCHIVAL AND RECORDS TERMINOLOGY 30 (2005), *available at* http://www.archivists.org/glossary/term_details.asp?DefinitionKey=156.
3. *Id.*
4. *Id.*
5. *Id.*

governmental accountability, they are essentially cultural institutions existing as "loss centers" and rarely ever as profit centers. Generally, they are valued by their publics for their preservation of authentic and historical works and occasionally by their parent institutions for protecting or enhancing the parent's core business and even supporting brand identity or providing sources of exploitable intellectual property.

To understand the kinds of issues that archivists face on their voyage to digitization, we need to consider what should be happening in professionally run archives. That is, what are the core functions or activities that the archivist in any type of repository should be doing to justify the resources at his or her disposal? To distinguish archives from just a cabinet of curiosities, there should be a systematic application of theory and principles to practice in seven areas or "domains":

1.   Establishment of administrative authority for the program to define its powers, ground its authority and justify the resources its work will require.

2.   Authentication of material brought into the archives to ensure that its value as evidence is not lost by either breaks in the chain of custody or lack of provenance information when custodianship has changed.

3.   Appraisal, evaluation or selection of those materials that have enough enduring value to justify retention consistent with the archives' mission and resources.

4.   Arrangement—that is, management of archival material—in a way that retains and reflects its original order, and in less than ideal situations deciphers and establishes the order which should have existed prior to accessioning.

5.   Description via inventories, catalog records, databases, indices and other tools commonly known as "finding aids."

6.   Preservation by providing appropriate physical protection and other measures so that the material survives from generation to generation.

7.   Enabling, promoting and facilitating the use of the archives for research, administration and study.

While careful work in each of these domains is essential, it is the last domain, the use of archival materials, which is *the* reason to do archives. There is really no justification to preserve material or take any particular care in selecting, arranging or authenticating material, unless the archivist is going to assign a priority to its use, whether in the original form or some copy. Professional principles and standards are especially important in the domain of "use," where archivists must both promote access and ensure protection of the documents. The archivist must also protect the rights within those documents, including privacy, public access and copyright. Of course, we cannot ignore the other domains; all seven have to be held in equilibrium.

I. A ROUGH RIDE BEFORE WE CAN EMBARK

Archival success requires a balancing act in applying professional practices in relation to available resources and present demands. This is especially difficult

now in the midst of three changes confronting the profession: changes in the fundamental nature of records, changes in the tools and resources at our disposal and changes in the societies we serve. These changes set the context for considering the major issues and challenges facing archives as we consider voyaging on to archival digitization.

No doubt archivists of all generations would have little difficulty in naming a set of seemingly overwhelming new challenges they faced in their own times as well as the legacy of the burdens they inherited. Yet, the real focus needs to be not on the burdens we carry from the past but on the challenges of the present and future. Any archive that conceives of its present only in terms of the past will likely find itself slipping into obscurity and irrelevance. While nothing is certain about the future, archivists know for sure that digital technologies are a key way to carry out our mission, the means by which we can sail to future ports and have a life in as yet undiscovered lands. Before we can embark on that voyage, we need to consider all the barriers, hazards and threats to fulfilling our mission under present day conditions.

Note that the phrasing of these conditions as "hazards and threats" should not be dismissed as just curmudgeonly negativity. What follows are ten genuine challenges which we must take into account before we can assign significant resources to digitization. There are solutions for some of these, and maybe some of them could be put on hold for a time, but it is the archivist's responsibility to engage with all of them. Thus, the conscientious archivist, like the prudent prospective sailor, needs to scan the horizon to chart a course that takes each into consideration at some level. While I am sure there are others, I will focus on these:

1. A fundamental challenge that underlies and is interwoven with all the others is the breadth and depth of changes in the nature of the organizations and people who author and use records and creative works of interest to archives. Thanks to information technology, social trends, the combination of a workaholic culture and personal passions for connectivity, the borders between work and personal life, between public and private, between formal and informal are no longer fixed but permeable. Thus, rather than having bearing walls on which we can build our archival house, we have folding partitions that are neither fixed nor sturdy enough for systematic, planned records transfers, collection development plans or user services programs. As an example, one need only look at the effect of the Freedom of Information Act or discovery requests that indiscriminately sweep vast quantities of undifferentiated recorded information from personal and official technological systems into a single bin and treat them as ad hoc records.[6] In the past, the only things that would have been considered records would have been items more consciously and formally created, classified, filed and maintained; thus, the need for archival intervention would not arise until a much later time in the so-called "life cycle of records." Furthermore, today such transactions are often being conducted without dependence on any archival intervention or opportunity to exercise control.

---

6.   5 U.S.C. § 552 (1996).

2. Implicit in the first challenge are several other impediments to capturing the archival records posed by what the old world has called "electronic records." Without going into the details evoked by such terms as metadata, migration, emulation, data warehouses, trusted repositories or "bit decay," it is sufficient to note that the content of electronic records will not survive for future generations unless archivists are able to exercise significantly more archival intervention in records-keeping practices than they have in the past. One need only think about how one would go about capturing "records" from the following five developments: large scale enterprise systems such as PeopleSoft, SAP or SunGard's Banner, scientific research data sets, a proliferation of personal information and personal communication systems, social media and "cloud computing."

3. Not so much a new trend, but an enduring challenge nevertheless, is the problem of completing the arrangement and description of the materials selected for retention and use by our publics. Over the past forty years, new technologies have enabled greater dissemination and sharing of descriptive data about archives. While I would never deny the value brought by this sharing, it comes only at the significant cost of both standardization and endless data conversions as finding-aid systems progress on cycles of progressive obsolescence. Without such developments as the MARC-AMC format in the 1980s, EAD in the 1990s and DACS (Describing Archives: A Content Standard) in the past ten years, we would be unable either to share our collections' information or support digitization. Yet, each of these standards has created methodological and resource demands, not to mention the costs of renewing equipment and software that previously did not need to be supported.

4. Even as archives move into a world of virtual rather than physical holdings, we are still faced with the analog legacy of the recent and distant past in the "person" of processing backlogs. Whether a matter of unopened, unappraised and undescribed boxes or minimally processed records, papers and so-called "hidden collections," every archivist's responsibility is to avoid passing on static processing backlogs to his or her successors. Yet, often large, conventional processing backlogs are not terribly complicated problems, just ones that compete for time and resources with more interesting new things, such as digitization.

5. There is a further arrangement and description challenge that we will face more immediately in our voyage to the lands of archival digitization. In the current era, the profession has finally awakened to the common sense voiced by archivists Mark Greene and Dennis Meissner that we need to apply a "more product, less process" (MPLP) approach to processing, and thus reduce archival backlogs.[7] Yet, it has become equally clear that either conventionally or MPLP-processed collections require significantly more detailed arrangement and description if they are to be digitized. Indeed, digitization requires archivists to cross the river Styx into the underworld of item-level description. Moreover, because of the metadata

7.   Mark A. Greene & Dennis Meissner, *More Product, Less Process: Revamping Traditional Archival Processing*, 68 AM. ARCHIVIST 208 (2005).

demands of such systems as ContentDM, archivists now have to describe not just individual items but also pages of items and facets of items, such as original copyright owner and current right holders.

6. On top of all this, preservation remains an ever present issue for archivists, especially for the more complex, and often the more interesting, materials in archives, such as sound recordings, photographs, video and electronic data. All these pose considerable challenges if we truly want them to be usable in the future. Generally, not only are the needed solutions expensive, but each special format comes with its own special twist because of the dependence on specific machinery and machine protocols. Preservation requires immediate action to migrate or reformat before obsolescence sets in. If, for instance, we wait until U-Matic video players are obsolete and gone from the marketplace, it will be too expensive and probably too late to carry that content into the future. Meanwhile, the clock keeps ticking on the wood-pulp paper, nitrate and acetate negatives and audiotapes that are already deteriorating in our archives.

7. There should be nothing more important to the archivist of yore or of today than enabling and promoting use of archival material. Technology has facilitated this primary archival objective through Internet databases and search engines, as well as quick communications through email. Some archives have even reached out via Facebook and Twitter. But even as these approaches can expand our audiences, this new environment brings its own particular challenges, namely an increase in the sheer number of research users and a rise in the casual requests that, no matter how trivial or how insoluble, still have to be investigated by reference staff. This, of course, takes time away from digitization efforts and vice versa.

8. There is nothing wrong with a wider audience for archives, and the broader societal and educational use of archives makes this a satisfying time to be an archivist. Yet, it comes with its own issues. Students not only are learning and studying in quite different ways—for example, with more collaboration and personal informality in creating their final products—but they are also coming in far less prepared for use of either archives or libraries. As subjects such as spelling, penmanship and languages are de-emphasized in primary and secondary schools, archivists increasingly have to be personal coaches and translators for their younger users, another drain on our finite amount of time.

9. A constant set of issues facing archivists in any generation is the range of matters relating to the authorization and foundations for our programs. On the one hand, this is a matter of having sufficient personnel, in terms of both quantity and quality, and having enough space of sufficient quality for our collections. Now, however, it is also a matter of having access to the range of staff expertise needed for managing electronic records and making them available via new technologies. The matter of space no longer can be measured solely in terms of NASF and Isoperms but also in terms of terabytes and petabytes, as well as backup and data refresh cycles. In short, technology has not reduced the resource needs of archives; rather, it has increased them exponentially.

10. However, perhaps more significant is how the revolution in our user community's expectations interacts with our resource allocators' patience and

demands.   That is, our users have come to expect twenty-four-seven access, unlimited by spatial boundaries. Who among us can sustain such expansion, no matter how much we and our institutional superiors want us to be on the front lines of a new archival world order?

II. READY TO GET ON BOARD?

While every archivist might have his or her own top ten challenges, the ones on my list are not insoluble; nor are all these issues faced equally by all archivists. Nevertheless, they are simply a good map of the twists and turns we have to navigate; they are not unbreachable barriers that argue against travel to the port from which we will embark and set sail for digitization. As significant as these challenges are, digitization is fundamental to the future of archival operations and thus must be part of our current programs. After all, it represents an opportunity to reach new audiences; an opportunity to respond to increased audience demands; an opportunity to fulfill the ancient archival mission and fundamental goal of promoting use; a means to fend off intrusion and competition from other information professionals by emphasizing the distinctive professional archival domains; a solution to some current pressing problems such as document deterioration from the wear and tear of reference use while also substituting for staffed reference points in an era in declining budgets; and a means to preclude obsolescence and demonstrate public engagement. Unfortunately, it also creates a drain on our resources.

Nevertheless, walking the gangway to get on board for the voyage to digitization entails at least four significant challenges of its own: 1) We need well selected archival series/collections that respond to user needs and are logical fits for our particular repository's core mission. 2) We need staff time for reprocessing and preparation of materials for scanning. 3) We need resources to support digitization beyond just pilot and demonstration projects.   4) We need a technological infrastructure to provide efficient input mechanisms that can reduce item-intensive work, effective public interfaces that represent individual archival items and considerable server space that will be absolutely reliable into the indefinite future.

More directly relevant to this symposium's focus, however, is that we need a pathway through intellectual property challenges—the rocky shoals. Indeed, we are likely to encounter diverse villains on this journey, whether they be tax and toll collectors, highwaymen or pirates (i.e., ISP vendors, librarians or digitize-it-yourselfers), but the greatest problem we may face comes from outside of our professional toolkit, i.e., copyright, which may be understood better in terms of ghosts and zombies.

This is a particular problem for archival digitization because of the very nature of archives. To be authentic, an archives or manuscript collection must be broad in scope and deep in coverage, if not complete and exhaustive. While the terminology may come from an analog world, these goals are certainly portable to the digital lands. No archival series or collection is effectively digitized until every box, every folder, every document and every page is included. Anything less is not an

authentic archive, and any compromise undermines the full value of both the original archival custody and the subsequent digital project.

A few years ago, I conducted a study of thirteen digital archives sites, many of which were multi-institution portals, thus representing a rather large number of collections and projects.[8] While the most common content seemed to be individual photographs, often of unclear copyright status, a major component was works clearly in the public domain—ones published before 1923 or U.S. government works, such as the Making of America digital library.[9] In far too few instances were there actual primary sources or unpublished textual documents, and these were predominantly taken out of their context, selected items from larger collections. On many sites today, the content presented is less than complete, and thus far from archival. In some cases, such as those sites that include pre-1923 published material from a collection but no correspondence from relevant individuals, one suspects that the reason is probably copyright. A prime example of incompleteness is the Paul Eliot Green Papers at the University of North Carolina, where only about fifteen letters, comprising 111 pages, have been digitized out of a 192 linear foot, 110,000 item collection.[10] The Library of Congress's Leonard Bernstein collection makes available only 1,100 letters written by nine correspondents out of 15,500 letters written by 3,300 individuals, and the finding aid indicates that permission was secured for these nine correspondents.[11] In a case I know more intimately, from the James B. Reston Papers at the University of Illinois, we have been able to digitize only about 2,000 pages of an estimated 146,000 pages in the collection.[12] Again, the reason for the limit is

---

8. William J. Maher, *Heritage Under Lock, but No Key: The Troubled Status of Unpublished Works in Digital Archives Projects*, in FREE CULTURE AND THE DIGITAL LIBRARY SYMPOSIUM PROCEEDINGS 2005 141 (Martin Halbert ed., 2005).

9. For the University of Michigan's *Making of America* pages, see *Making of America*, UNIV. MICH. LIBR., http://quod.lib.umich.edu/m/moagrp/ (last updated June 13, 2007). For Cornell University's *Making of America* pages, see *Making of America*, CORNELL LIBR., http://digital.library.cornell.edu/m/moa/ (last visited Oct. 15, 2010).

10. *Paul Eliot Green Papers*, U.N.C. CHAPEL HILL LIBR., http://docsouth.unc.edu/ wwi/greenletters (last visited Sept. 20, 2010). The Green letters were digitized as part of what appears to be a special project: "North Carolinians and the Great War." In that project at least one nineteen-page diary and several outgoing letters were found for a person who died in 1959, but this is still a relatively small part of a 2.0 linear foot (340 item) collection of the papers of Robert March Hanes. Without comparing the Web version of materials in this project with what is in hard copy in Chapel Hill, it is not entirely clear whether the limitations on content are driven more by copyright than by the subject parameters of the project; but regardless, for either Green or Hanes what is presented on the Web is something significantly less than the archival collections themselves. *See Robert March Hanes Papers*, U.N.C. CHAPEL HILL LIBR., http://docsouth.unc.edu/wwi/hanesdiary (last visited Sept. 20, 2010).

11. For the finding aid of the collection, see *The Leonard Bernstein Collection*, LIBR. CONGRESS, http://lcweb2.loc.gov/service/music/eadxmlmusic/eadpdfmusic/mu998001_x.pdf (last updated May 2010). For correspondence, see *The Leonard Bernstein Collection, ca. 1920–1989*, LIBR. CONGRESS, http://memory.loc.gov/ammem/collections/bernstein (search "correspondence" in the "search this collection" field) (last visited Sept. 20, 2010). For the rights and permissions for the digitized correspondence to and from Bernstein, see *Rights and Reproductions*, LIBR. CONGRESS, http://memory.loc.gov/ammem/collections/bernstein/lbres.html (last visited Sept. 20, 2010).

12. For the James B. Reston Papers, see *James B. Reston Papers*, U. ILL. LIBR.,

copyright. Also at Illinois, where we hold the American Library Association Archives, we digitized all of its Board of Education for Librarianship Subject File (1914–56), 6.4 cubic feet, or around 19,200 pages, but we cannot make it available via a live link because of copyright.[13] When individual documents are "cherry-picked," the project is a digital scrapbook or a digital exhibit but nothing much more. The result may even be of some research usefulness, but this is hardly a digital archives.[14]

Thoughtful archivists know that no archives can be fully trusted as repositories of an objective historical truth. Instead, they are merely tellers of a particular narrative. Thus, we also need to recognize that the more we apply selectivity to what we digitize within a particular record series or collection, the less authentic of a voice we give to those documents. This general condition is relevant to the copyright difficulties we encounter when planning digitization. In either a modern or ancient archives, so much of what we have are documents attributable to authors whose names we sometimes, but do not always, know, but who are basically anonymous. In Congress's rush during the 1990s to extend the term of copyright, there was no serious consideration of this reality, or indeed the broader cultural impact of decades added to term. Instead, the push was driven by a bald-faced commercial imperative. At the time, we lacked a name for the problem. Today, we know it as "orphan works." Over the past six years, having this moniker has certainly helped us identify the problem and attempt to devise solutions. However, it may be that we could be more successful if we called this not so much a problem of "orphans" but of "ghost works" and "zombie authors." They are the papers left by what seem like the living dead who arise and walk the earth for seventy to 120 years after having lifted pen from paper, hands from keyboard or thumb from the shutter release. These zombies surround the documents that represent the evidence of their existence as our users seek to examine them in their search to understand the context of the creative moment. Thus, while it seems that technology and professional development have brought our ship within sight of the dream land of archival digitization, we have not just rocky shoals to consider but also some very persistent and pesky undead to deal with as well. I look forward to this

http://www.library.illinois.edu/archives/archon/index.php?p=collections/controlcard&id=5060 (last visited Oct. 15, 2010).

    13.   For the *Board of Education for Librarianship Subject File*, see *Board of Education for Librarianship - Subject File*, U. ILL. LIBR.,http://www.library.illinois.edu/archives/ala/holdings/ ?p=collections/controlcard&id=7095 (last visited Oct. 15, 2010).

    14.   Indeed, all of these practices are quite at odds with the rhetoric by which such projects are promoted. For example, we are told that a driving concept for the Valley of the Shadow project was that it be "a research library in a box, enabling students at places without a large archive [sic] to do the same kind of research as a professional historian." *The Story Behind the Valley Project*, U. VA. LIBR., http://valley.lib.virginia.edu/VoS/usingvalley/valleystory.html (last visited Sept. 20, 2010). The Online Archive of California site is advertised as providing "all" with "access to information previously available only to scholars who traveled to collection sites." *See, e.g.*, *Online Archive of California*, BANCROFT LIBR. (Dec. 3, 2009), http://bancroft.berkeley.edu/collections/oac.html. However, this claim is clearly undermined by the fact that only 212 of the over 9,000 collections listed in the database are marked with the rubric of "Online items available." *Browse the Collections*, ONLINE ARCHIVE CAL., http://www.oac.cdlib.org/institutions/UC+Berkeley::Bancroft+Library? (last visited Oct. 7, 2010).

14   COLUMBIA JOURNAL OF LAW & THE ARTS   [34:1

symposium's efforts to help us on our course.