

Symposium: Digital Archives: Navigating the Legal Shoals

Undue Diligence?

Peter B. Hirtle*

My purpose today is to suggest to you why I think that archivists have by and large failed to use new digital technologies to make the riches in their holdings more widely accessible. Money, of course, has a lot to do with it. The funding for massive archival digitization projects has just not been available. But there has also been a reluctance on the part of archivists to pursue the small amount of funding that is available because of our professional practices.

The problem is that archivists (and our librarian colleagues) may be the last people in the United States who blindly and passionately attempt to respect copyright. Our professional codes tell us we must. For example, the *Code of Ethics for Archivists* asserts that archivists must obey all federal, state and local laws—which would include copyright law.¹ Furthermore, the *Code of Ethics* makes it plain that we must also respect personal privacy, especially when it involves third parties whose works may appear in archival collections, but who have not actually donated the material to the repository themselves.² Similarly, the *ALA/SAA Joint Statement on Access to Research Materials in Archives and Special Collections Libraries*, while encouraging archivists to make their holdings as widely available as possible and on an equitable basis, also states that we must respect copyright laws when making reproductions for users.³

So, archivists are imbued with the desire to respect the law. Yet, in spite of the

* Peter B. Hirtle is a senior policy advisor for Digital Scholarship Services at Cornell University Library with a special mandate to address intellectual property issues. He also serves as the selector for U.S. and general history, genealogy and the history of the book. Previously at Cornell, Mr. Hirtle served as Director of the Cornell Institute for Digital Collections and as the Associate Editor of *D-LIB MAGAZINE*. He is an archivist by training with a Masters in History from Johns Hopkins and a Masters in Library Science with a concentration in archival science from the University of Maryland. Mr. Hirtle is a Fellow and Past President of the Society of American Archivists and is a member of its Working Group on Intellectual Property. He was a member of the Commission on Preservation and Access/Research Library Group's Task Force on Digital Archiving and the Copyright Office's Section 108 Study Group and is a contributing author to the LibraryLaw.com blog.

1. *SAA Council Handbook: Code of Ethics for Archivists*, SOC'Y OF AM. ARCHIVISTS, § IX, http://www.archivists.org/governance/handbook/app_ethics.asp. (last visited Oct. 7, 2010).

2. *See id.* § VII.

3. Am. Library Ass'n & Soc'y of Am. Archivists, *ALA/SAA Joint Statement on Access to Research Materials in Archives and Special Collections Libraries*, SOC'Y OF AM. ARCHIVISTS (June 1, 2009), <http://www.archivists.org/statements/ALA-SAA-Access09.asp> (last visited Oct. 7, 2010).

efforts of Bill Maher to ensure that every archivist has taken a workshop on copyright, there is great uncertainty in the profession as to what the law says, what is legal and what is illegal. It is a riddle that most archivists do not know how to solve, and I dare say many here would say cannot be solved. Coupled with this uncertainty is a great desire on the part of archivists to stay out of court—even though as far as I can tell, only one collecting organization has ever been sued over its reproduction practices, and the circumstances of that case were unusual in that the person who actually used the reproduction was closely connected to the repository, and so it was as much a suit against a scholar as against a repository.⁴ Because archivists are risk averse, it is easiest to assume that the law stops us from doing many socially desirable things.

It doesn't help much that the law itself offers little in the way of assistance to archivists. Section 108 should be a clarion call to archivists to make their materials more available.⁵ The section, which currently covers both libraries and archives, was first introduced at the request of archivists in order to increase access to unpublished materials.⁶ It was Congress's intent to allow archivists to use the prevalent reproductive technology of the day—microfilm—to replicate archival collections in repositories across the country, thus increasing their availability for research, while at the same time helping ensure their preservation.⁷ There was general agreement that archival reproduction, distribution and access to research collections did not constitute publication and would not interfere with the ability of copyright owners to later commercialize the work. Current § 108, however, does not allow archivists to use the modern day microfilm equivalent—digital technologies—to provide distributed access to archival repositories.⁸ Access to digital copies of archival material is limited to the premises of the archives.⁹

Fair use is of little additional help. I won't rehash the arguments about why fair use is underemployed by librarians and archivists. Two points are worth making, however. The first is that in spite of the explicit addition to the statute in 1992 that the use of unpublished material can be a fair use, many of my colleagues were badly scarred by the *Salinger* decision and still work from the assumption that if it is unpublished, any specific use is less likely to be fair.¹⁰ Secondly, at first glance the current emphasis on transformativeness in fair use decisions would seem to work against a finding of archival fair use when making digitized collections available.¹¹ Archival practice is not to comment or criticize when presenting

4. *Sundeman v. Seajay Soc'y, Inc.*, 142 F.3d 194 (4th Cir. 1998).

5. 17 U.S.C. § 108 (2006).

6. Peter B. Hirtle, *Digital Access to Archival Works: Could 108(b) Be the Solution?*, COPYRIGHT & FAIR USE: STANFORD U. LIBR. (Sept. 24, 2006), http://fairuse.stanford.edu/commentary_and_analysis/2006_08_hirtle.html.

7. *Id.*

8. Distribution is limited because access can only be granted on the premises. So, by implication, § 108(b) does not allow general distribution. 17 U.S.C. § 108(b).

9. 17 U.S.C. § 108(b).

10. *Salinger v. Random House, Inc.*, 811 F.2d 90 (2d Cir. 1987).

11. See, e.g., *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994) and subsequent cases that have built on its discussion of transformativeness.

information; that is for the user of the material to do. Instead, archivists want to present records in the context in which they were created, but without editorial or other commentary. In the minds of many (but not all), such practice would argue against a finding of fair use.

And of course, securing the permission of the copyright owner is not an option. Archival collections are perhaps the purest representation of the “orphan works” problem—those works still protected by copyright whose current rights owners cannot be found because they either can’t be identified or can’t be located. As an early speaker at this conference noted, they might better be called “zombie” works. They are the living dead that nevertheless threaten us all with ruin. Archivists live every day with billions of “zombie” copyrighted works created by the “life plus seventy” term.¹² Consider this: I was recently told that the oldest work still protected by copyright in the U.K., which has a “life plus seventy” term for published works, was published in 1859.¹³ (The author died in 1940.) That means, conceivably, any work created since 1859 could be protected by copyright. But we can assume that one is at least in young adulthood when the first publication appears (seventeen in the case of this poem), whereas an unpublished work could easily be from someone even younger. Furthermore, published authors have some degree of prominence and it may be possible to trace them; authors of unpublished works may be incredibly anonymous. Lastly, with unpublished work it is often impossible to tell if the work was a work made for hire and copyright in the work belonged to an employer.¹⁴ One can spend days searching for an heir of a correspondent when in reality copyright belongs to someone else.

A good indication of the staggering problem that zombie works present for archival digitization projects is the example of the Thomas Watson Papers at the University of North Carolina.¹⁵ As a research project, they attempted to identify right holders in the papers of Thomas Watson, a somewhat prominent Senator from Georgia who died in 1922 with a small (7.5 linear feet) collection of correspondence.¹⁶ They decided to ignore the work-made-for-hire problem, instead assuming that all the works were created in an individual capacity, and set to work to locate current right holders.¹⁷ They found 8,434 items in the 7.5 linear feet.¹⁸ It took archivists more than ninety hours of work to extract the names, dates and

12. See 17 U.S.C. § 302(a) (providing for “life plus seventy” term).

13. Peter B. Hirtle, *Factoids: What Is the Oldest Work Protected by Copyright in the U.S.? What Work Will Have the Longest Protection?*, LIBRARYLAWBLOG.COM (Mar. 28, 2010), <http://blog.librarylaw.com/librarylaw/2010/03/factoids-what-is-the-oldest-work-protected-by-copyright-in-the-us-what-work-will-have-the-longest-protection.html>.

14. 17 U.S.C. § 302(c) (providing duration for anonymous works, pseudonymous works and works made for hire).

15. See THE THOMAS E. WATSON PAPERS DIGITAL COLLECTION, <http://www.lib.unc.edu/dc/watson/> (last visited Sept. 27, 2010).

16. Maggie Dickson, *Due Diligence, Futile Effort: Copyright and the Digitization of the Thomas E. Watson Papers*, 73 AM. ARCHIVIST 618, 619 (forthcoming 2010); THOMAS E. WATSON PAPERS DIGITAL COLLECTION, <http://www.lib.unc.edu/dc/watson/> (last visited Sept. 27, 2010).

17. Dickson, *supra* note 16, at 619, 623.

18. *Id.* at 622.

geographical locations of authors of all of the incoming correspondence, obtaining 3,304 names.¹⁹ Using the genealogical database Ancestry.com, and other online resources, they found that 608 of the correspondents had died more than seventy years ago, putting their work in the public domain.²⁰ Archivists could determine death dates for only 1,101 of the remaining correspondents.²¹ This work required the full-time labor of an archivist for *fourteen weeks*.²² After all of this effort, only a handful of the representatives of correspondents were successfully contacted, and they freely gave their consent for the old letters to be published.²³

What about orphan works legislation you might ask? It is important and it should be passed. The Society of American Archivists has already prepared guidelines on what might constitute a reasonable search for the owner of an orphan work.²⁴ But at the same time, the proposed legislation is really aimed at those institutions and individuals who wish to exploit a limited number of orphan works. Even the author of the Copyright Office's *Report on Orphan Works* has admitted that the diligent search aspect of the proposed legislation was not specifically intended to address the problem of mass digitization and would not be an appropriate solution for that problem.²⁵

So, given the professional requirement to follow the law, and the absence of clear and useful exceptions for archives in the law, it is perhaps not so surprising that few archivists are willing to risk engaging in large scale digitization of collections still technically protected by copyright.

I would argue, however, that even in the absence of explicit and clear legal authorizations, repositories face little real risk in digitizing many collections that are still technically protected by copyright. I know this because many institutions have unknowingly and mistakenly assumed that they had the right to make available archival material in digital form, often to great acclaim. Two examples:

The Judaica Sound Archives at Florida Atlantic University makes available pre-1923 sound recordings on its website; later recordings are accessible through research stations that are distributed to other universities.²⁶ Its efforts to make the voices of early cantors and rabbis more well known has led to praise rather than suits. This is in spite of the fact that its understanding of sound recording copyrights is seriously flawed.

Second example: many archives have elected to add images to the Flickr

19. *Id.* at 620.

20. *Id.* at 621.

21. *Id.*

22. *Id.*

23. *Id.* at 623.

24. SOC'Y OF AM. ARCHIVISTS, ORPHAN WORKS: STATEMENT OF BEST PRACTICES (2009), available at <http://www.archivists.org/standards/OWBP-V4.pdf>.

25. Julie Sigall stated this in his remarks given on the *O is for Orphans* panel at the *D is for Digitize* conference at New York Law School. *O is for Orphans*, NEW YORK LAW SCHOOL, 1:14:45–1:16:55 (Oct. 10, 2009), http://www.nyls.edu/centers/harlan_scholar_centers/institute_for_information_law_and_policy/events/d_is_for_digitize/program/ (follow “O is for Orphans” hyperlink).

26. FAU JUDAICA SOUND ARCHIVES, <http://faujsa.fau.edu/jsa/home.php> (last visited Sept. 27, 2010).

Commons.²⁷ In order to post images to the Commons, repositories are supposed to have reasonably concluded that a photograph is free of copyright restrictions. They are then permitted to share that photograph under a new usage guideline called “no known copyright restrictions.”²⁸ There are four bases for concluding that a photograph has “no known copyright restrictions”:

1. The copyright is in the public domain because it has expired;
 2. The copyright was injected into the public domain for other reasons, such as failure to adhere to required formalities or conditions;
 3. The institution owns the copyright but is not interested in exercising control;
- or
4. The institution has legal rights sufficient to authorize others to use the work without restrictions.²⁹

It would seem, though, that many repositories assume that if they do not know of copyright restrictions, then there are none. Here is an example of a photograph from the George Eastman House in Rochester, New York, entitled “Woman and Boy Sitting in Chair.”³⁰ No author is given, and there is no evidence or indication that the photograph was ever published. Copyright law says that copyright in an unpublished anonymous work expires 120 years after creation.³¹ The Eastman House, however, has posted this photograph with the “no known copyright restrictions” rights statement. It suggests that they feel that since no copyright owner can be identified, one can treat the photograph as if it were in the public domain. I am unaware of any objections to the Eastman House’s use of the photograph. You find this sort of behavior everywhere: treating orphan/zombie copyrighted works as if they were in the public domain.

And there are, of course, elements in the law that work to the advantage of repositories that make unpublished copyrighted works available. Chief among these is the absence of statutory damages or attorney’s fees for unregistered works.³² In addition, repositories are absolved of statutory damages if they can make a good-faith assertion that their use is a fair use.³³ Because of these exemptions, it is much more likely that the repository will receive a take-down request than a lawsuit.

There is a growing recognition that repositories, in their desire to obey the law and avoid litigation, may have been overly cautious. In recognition of this state of affairs, the Online Computer Library Center recently held a workshop on what they suggested could be described as “undue diligence” among archivists.³⁴ The

27. *The Commons*, FLICKR, <http://www.flickr.com/commons> (last visited Sept. 19, 2010).

28. *The Commons: About the Rights Statement*, FLICKR, <http://www.flickr.com/commons/usage> (last visited Sept. 19, 2010).

29. *Id.*

30. *Woman and Boy Sitting in Chair*, FLICKR, http://www.flickr.com/photos/george_eastman_house/2677418229/ (last visited Oct. 19, 2010).

31. 17 U.S.C. § 303(a) (2006).

32. *Id.* § 412.

33. *Id.* § 504(c)(2).

34. See *Introduce Balance in Rights Management*, ONLINE COMPUTER LIBRARY CENTER, <http://www.oclc.org/research/activities/rights> (last visited Sept. 19, 2010).

workshop concluded that archivists should not reject the use of digital distribution out of hand, but rather use the practices that they have followed in the analog world when making things available digitally.³⁵ Draft recommendations on well intentioned practice have been prepared and are being distributed for comment throughout the community. Common elements include sensitivity to the materials, disclaimers and takedown provisions.

There are other approaches that might be of use. The University of California, Irvine Libraries, for example, have created for a born digital collection a “virtual reading room” that mimics analog practice: registration, delivery of the copy of the rules, etc.³⁶ Of course, tricky issues still remain. Should, for example, the contents of the documents in a virtual collection be accessible to commercial indexing services? To do so would greatly increase the utility of the collection (and also make it easier for people to find documents to which they might object and request that it be taken down). At the same time, though, it may seem more like “publication” and an inappropriate intrusion on the rights of the copyright owner.

While there is talk of fundamental copyright reform in the wind, I see little chance any time soon that the law will change in meaningful ways to aid archival repositories. We need a different approach. Just as the great documentary editing projects of the last century ignored the copyright laws in order to make access to the papers of the Founding Fathers more widely available, so too must archivists give up their hope of black letter rules on copyright and instead embrace responsible risk management as the appropriate way of managing our social goals.

35. *Id.*

36. *Richard Rorty Born Digital Files, 1988–2003*, UCISPACE @ THE LIBRARIES, <http://ucispace.lib.uci.edu/handle/10575/7> (last visited Sept. 19, 2010).