Case Studies

Due Diligence, Futile Effort: Copyright and the Digitization of the Thomas E. Watson Papers

Maggie Dickson

Abstract

As archives and libraries digitize and make their collections available online, they are faced with the challenge of meeting growing patron expectations in the online environment while still adhering to copyright statutes. This article reports on a case study investigating the copyright status of materials from a recent effort to make the Thomas E. Watson Papers, a manuscript collection housed at the Southern Historical Collection at the University of North Carolina at Chapel Hill, accessible online. The article also explores fair use as a possible approach for digital publication of archival collections containing materials protected by copyright.

The advent of digital technologies presents archivists with opportunities to provide their patrons unprecedented—and, increasingly, expected—online access. Some archival repositories are now exploring and, in some cases, engaging in, large-scale digitization and Web presentation of their collections. This allows patrons to perform research from their homes or offices at any time of day, which formerly they would have traveled great distances and endured limited hours of operation to accomplish.

In September 2007, the Southern Historical Collection and the Carolina Digital Library and Archives of the University of North Carolina at Chapel Hill

© Maggie Dickson.

The author wishes to thank Amy Johnson, the graduate assistant working on the project at the time of this case study. She proved herself indispensable to the project through her tireless and creative efforts gathering and analyzing the data for this research. The author also wishes to thank Peter Hirtle for his crucial editorial assistance in preparing this article for publication.

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-07-02 via free access

began a two-year, privately funded grant project to digitize and make available online the entirety of the Thomas E. Watson Papers, a manuscript collection housed in the Southern Historical Collection. Watson was a prominent Populist politician in the late nineteenth and early twentieth centuries, and his papers consist of nearly 28 linear feet of correspondence; drafts of books, articles, speeches, and other writings by himself and others; periodicals and pamphlets he edited and published; political materials; legal and financial papers; biographical information; diaries and scrapbooks; family pictures; and other materials.

A major challenge for this project was dealing with the copyright status of the materials. We obtained permission to use any materials created by members of the Watson family prior to beginning work on the digital collection, because the donors for the project were descendants of Thomas E. Watson. But any materials created by third parties found in the manuscript collection were potentially still in copyright. Unpublished manuscript materials are protected by copyright for the life of the author plus seventy years. For us, that meant that any materials created by anyone dying prior to 1939 (2009 was the publication date for our digital collection) were fair game under general copyright rules. However, any materials created by someone dying after 1939 were potentially still in copyright.

Therefore, if we did not claim any exemptions to copyright statutes, and if we wanted to present the entire archival collection on the Web under a strict interpretation of copyright law, we needed to identify all authors of materials in the collection, determine their death dates, locate descendants for those who died after 1939, contact those descendants, and request and then obtain permission to use their deceased family members' materials.

The Case Study

We suspected this effort had little chance of success, but our project was not only about digitizing the Thomas E. Watson Papers. Because it also served as a pilot for a much larger effort aimed at digitizing the entire Southern Historical Collection, we would attempt to research copyrights for the entire correspondence series as a way of thoroughly investigating this aspect of digitizing archival materials. We considered this series, consisting of 7.5 linear feet of letters, postcards, telegrams, and notes written by Watson and his family, friends, and political and business colleagues, a good testbed for this type of research in that it was large enough to yield meaningful results yet small enough to undertake in a reasonable time frame. The dates for the correspondence series range between 1873 and 1986, with the bulk of the letters dating from the 1880s to the 1920s.

Prior to beginning copyright research, the project manager and a graduate research assistant gathered basic metadata (correspondent and recipient names,

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-07-02 via free access

places from which letters were written, and dates, for example) from the more than 8,400 documents in the correspondence series. This information served a dual purpose, being used not only for investigating copyright, but also for creating browse-able and searchable indexes for the digital collection. The project manager and research assistant undertook this work at costs of $28.63 and $16.55 per hour,[1] respectively. It took more than 91 hours to go through the 15 document cases in the correspondence series. The initial cost to compile this data was $1,960. This rough data was iteratively corrected and improved throughout the course of the entire project, but we did not track this effort, and it would be difficult if not impossible to determine the additional costs incurred.

From the information we gathered, we were able to condense and regularize the correspondent list at 3,304 personal names.[2] It is important to note that in many cases, it was difficult to determine whether letters were of a personal or a business nature. If the latter, they might have been works for hire, and unpublished works for hire have a copyright term of 120 years from the date of creation. When we approached the legal counsel for UNC University Libraries on this matter, we were advised to err on the side of personal correspondence, except in the cases where a letter was clearly a work for hire. Additionally, many of the personal letters in the series were written by politicians and other federal or state employees. These works could therefore potentially be considered government documents and subject to another copyright status. We decided that investigating the copyright status of works for hire or government documents would be outside the scope of this study due to our limited project timeline.

The research assistant developed a workflow for identifying and gathering information for determining the death dates of the correspondents in the series. Since we were not sure how long the process would take, we decided to begin work on a 10% sample, which we were advised was a legitimate sample size.[3] The research assistant gathered the sample by selecting every tenth name from the list of correspondents and completed this investigation in 36 hours. Once the sample was completed and we determined that it was feasible, we decided to do the same research on the rest of the names.

We attempted to identify the 3,304 names using a variety of sources. These included ancestry.com, the *Congressional Biographical Directory*, the Historical Marker Database online, the Library of Congress authority database, the *New Georgia Encyclopedia*, print references, the Social Security Death Index, the

[1] These costs include benefits.

[2] 105 letters had signatures that were missing, illegible, or incomplete, and no other contextual information on the documents allowed us to infer the identities of the correspondents. For these materials no further copyright research was possible.

[3] We calculated our sample from our population of 3,300 names using a confidence level of 95% and a confidence interval of 5% (a standard calculation for determining sample size). This yielded a sample size of approximately 10% of the list of names.

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-07-02 via free access

University of Texas WATCH File, *Wikipedia*, and WWI draft registration forms. The *Congressional Biographical Directory* was useful given the political nature of the letters; many of the correspondents had been members of the United States Congress. The *New Georgia Encyclopedia* proved useful as well since many of the correspondents wrote from locations in Georgia. Ancestry.com, for which we purchased a one-year account, yielded the most information by far because it aggregates and makes searchable information from census records.

What resulted was a list of 3,280 confirmed and questionable identifications, and 24 unknowns that were simply impossible to identify, meaning that we didn't find those names in any of the resources we consulted. We felt that we had possibly matched many of the individuals represented in the letters, but there wasn't any way to be certain. We were able to locate birth or death dates for 1,709 (51%) of the correspondents, while for 1,571 (48%) we found no dates in the sources we consulted. For the correspondents for whom we located dates, 1,101 (33%) died after 1939, meaning that their letters are potentially still in copyright, while 608 (18%) died during or before 1939, meaning that their letters are now in the public domain (see Figure 1). This work was almost entirely conducted by the research assistant over the course of 4.5 months working 20 hours a week, at the rate of $16.55 an hour for a total cost of nearly $6,000.

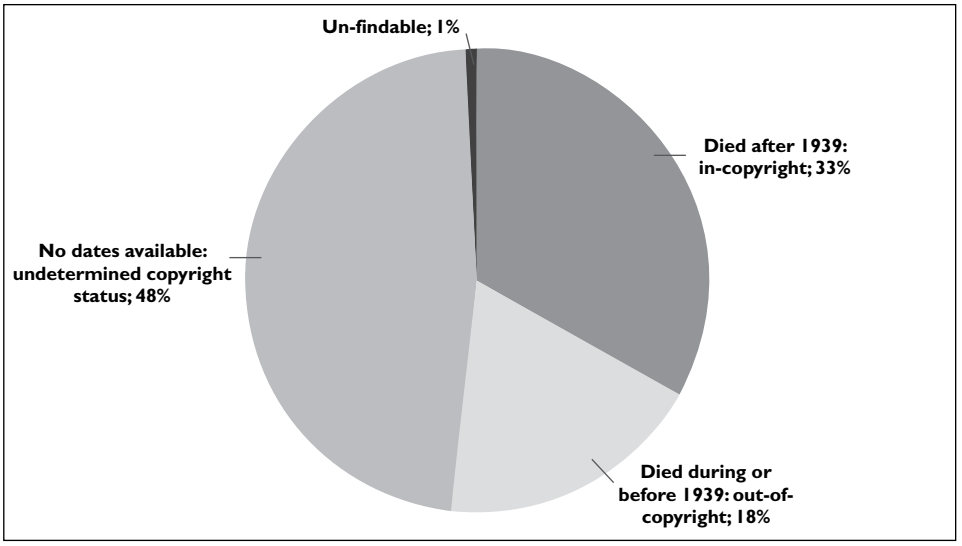

FIGURE 1. Copyright status of correspondents, *n* = 3,304.

Of the identified individuals who died after 1939 or for whom we were unable to determine death dates, we attempted to locate dedicated manuscript collections or materials in the manuscript collections of other individuals

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-07-02 via free access

deposited in repositories. Collections were identified for 50 of the correspondents. We located some of these manuscript collections by Google searching, while others were found by conducting searches using the *National Union Catalog of Manuscript Collections* (*NUCMC*) maintained by the Library of Congress, ArchiveGrid, and the *Congressional Biographical Directory*. In these cases, the project manager contacted the repositories by email, asking for their most recent acquisition information, hoping that it might lead us to descendants of the correspondents from whom we could request permission to digitize their relatives' materials. We received 25 responses to these inquiries. In most cases, no information was available, and when it was, it tended to be outdated, more often than not by more than 20 years.

By contacting repositories, we were able to obtain current, dependable contact information for the copyright holders for only 2 of the correspondents: William Randolph Hearst, a prominent newspaper publisher, and Miles Poindexter, a United States representative and senator from Washington state. We were somewhat generous with our definition of current and dependable contact information, as we were advised to contact William Randolph Hearst IV by sending our permission request care of the *San Francisco Chronicle*. We also found contact information for Upton Sinclair and Hamlin Garland, well-known writers with established literary estates, using the University of Texas WATCH File. Request for permission letters and forms were sent to these addresses by certified mail; 3 of the 4 forms were returned granting us explicit permission to include the letters by those authors in our digital collection. The fourth form, requesting permission to display 2 letters from William Randolph Hearst, was never returned.

To summarize, of the 8,434 documents in the series, 14% were written by members of the Watson family and had been cleared for copyright at the beginning of the project. We excluded from investigation 3% of the letters, which we indisputably determined to be works for hire and therefore outside the scope of the study for practical reasons, but which, of course, should be considered still under copyright. Four percent of the letters were signed with incomplete, illegible, anonymous, or pseudonymous names and contained no other contextual information with which we could identify the authors. Of the remaining 6,706 letters, written by just under 3,300 people, we determined that 21% were in the public domain and 27% were still in copyright, while 31% were of indeterminate copyright status. Going by a strict interpretation of copyright law, these results would allow us to make accessible online 35% of the correspondence series (see Figure 2).

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-07-02 via free access

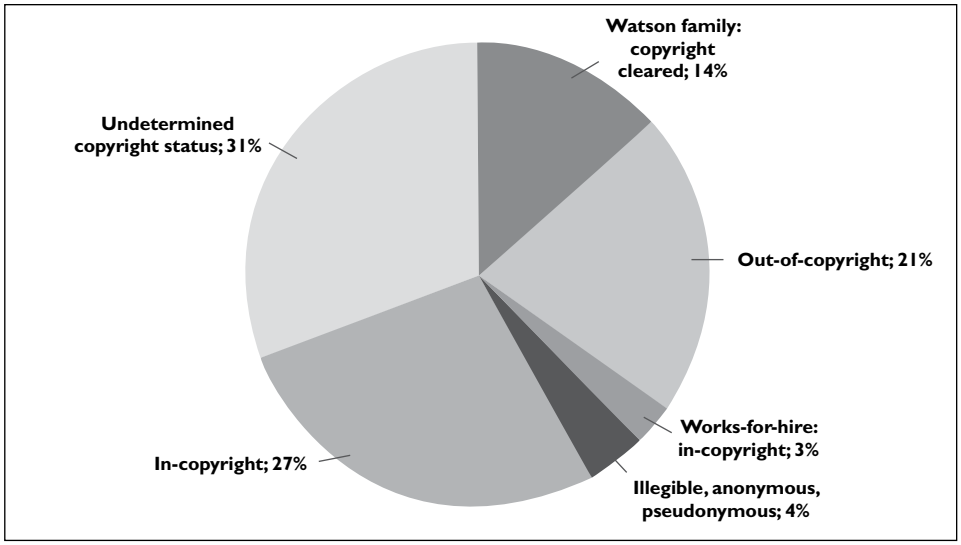

FIGURE 2. Copyright status of letters, *n* = 8,434.

The project manager and research assistant spent more than 450 hours over the course of 9 months to conduct this copyright investigation. The total cost of the research was approximately $8,000, or just over $1,050 per linear foot of correspondence.[4] At the end of the project, we were able to obtain explicit permission to display online only 4 letters, other than those written by members of the Watson family. Looking at the cost of our efforts in terms of the materials for which we were able to obtain permissions, our return on investment was $2,000 per document. And, as thorough as we were, we didn't address the issue of works for hire—had we done so, this research could easily have cost much more, likely without yielding additional results.

Title 17: Copyright Act of 1976

Copyright law was written to protect creative expression and gives the incentive of exclusive rights to creators of works. Under the current law, creators do not need to publish, register, or affix a copyright notice to their works; they need only to fix a creative idea in a tangible medium for that work to be protected by copyright.

However, exclusive rights held by copyright owners are not absolute. In *Feist Publications, Inc. v. Rural Telephone Service Company, Inc.*, Supreme Court justice

[4] This cost does not include the time spent by the project manager communicating with manuscript repositories or other copyright contacts, as this work was interspersed with her regular duties and difficult to quantify.

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-07-02 via free access

Sandra Day O'Connor stated, "The primary goal of copyright is not to reward authors, but to 'promote the Progress of Science and useful Arts'."[5] Limitations to exclusive rights and remedies in sections 107 to 122 of the Copyright Act allow for the use of copyrighted materials under some circumstances. As the current copyright law was written in 1976, however, its authors did not anticipate the ways in which digital technologies would change the potential uses of copyrighted materials, and we must interpret these limitations and remedies to determine which might best apply to the large-scale digitization of archival materials.

Section 108(b)—Limitations on Exclusive Rights: Reproductions by Libraries and Archives

Prior to the Copyright Act of 1976, the exclusivity of copyright included no exception to allow libraries and archives to reproduce works. Since 1976, libraries and archives have referred to section 108 of the copyright statute to support some reproductions of copyrighted materials. Section 108(b) allows for the limited reproduction of copyrighted materials by libraries and archives for the purposes of deposit for research in other institutions as well as for preservation and security.

This section has been applied in cases where manuscript collections are microfilmed, for instance, and it could potentially be considered as a possible exemption for digital reproduction as well. However, as Peter Hirtle notes,[6] this exemption only applies to up to 3 copies of a copyrighted work held by a library or archives, and, should a copy be in a digital format, the section states that the digital copy may not be "otherwise distributed in that format and is not made available to the public in that format outside the premises of the library or archives."[7]

The Section 108 Study Group recently reviewed this section. The group is charged with conducting "a reexamination of the exceptions and limitations applicable to libraries and archives under the Copyright Act, specifically in light of the changes wrought by digital media."[8] In March 2008, the Study Group published a report detailing its findings,[9] and, while the group recommended that the law be amended to accommodate the ways in which libraries and

[5] *Feist Publications, Inc. v. Rural Telephone Service Company, Inc.*, 499 U.S. 340 (1991).

[6] Peter Hirtle, "Digital Access to Archival Works: Could 108(b) Be the Solution?," *Copyright and Fair Use*, Stanford University Libraries (24 Sept. 2006), available at http://fairuse.stanford.edu/commentary_and_analysis/2006_08_hirtle.html, accessed 15 April 2010.

[7] *Copyright Act, U.S. Code* 17 (1976) § 108(b).

[8] Section 108 Study Group, "Mission Statement," available at http://www.section108.gov/mission.html, accessed 15 April 2010.

[9] Section 108 Study Group, *The Section 108 Study Group Report* (March 2008), available at http://www.section108.gov/docs/Sec108StudyGroupReport.pdf, accessed 15 April 2010.

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-07-02 via free access

archives use digital technologies for the preservation and dissemination of their materials, it did not specifically address the digital publication of large bodies of unpublished manuscript collections. Moreover, Congress has not yet taken up these recommendations. Until this section is amended to reflect the advent of large-scale digitization of archival materials, we must look to other provisions in Title 17 for help with copyright and the digital publication of our materials.

Section 107—Limitations on Exclusive Rights: Fair Use

An important limitation on the exclusive rights of the copyright holder is the "fair use" provision. Section 107 of the Copyright Act—Limitations on Exclusive Rights: Fair Use—states that "use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching, scholarship, or research, is not an infringement of copyright."[10] The Supreme Court, in its ruling in favor of the defendant in *Campbell v. Acuff-Rose Music*, stated that "fair use…permits courts to avoid rigid application of the copyright statute when, on occasion, it would stifle the very creativity which that law is designed to foster."[11]

The archival community is currently interested in the fair use limitation as the exemption that could potentially provide the most effective safe harbor against copyright infringement lawsuits resulting from large-scale archival digitization projects. For the Thomas E. Watson Papers digitization project, we explored this exemption as a possible justification for online presentation of our materials. Although there are no hard-and-fast rules determining whether or not a use is fair, the courts must consider 4 factors when determining fair use:

1. the purpose and character of the use, including whether such use is of commercial nature or is for nonprofit educational purposes;
2. the nature of the work itself [whether it is a factual or creative work];
3. the amount and substantiality of the portion used in relation to the copyrighted work as a whole;
4. the effect of the use upon the potential market for or value of the copyrighted work.[12]

Some libraries and archives shy away from using the fair use privilege to make their materials available online because of its vagueness; historically, court cases dealing with fair use and unpublished materials have been limited in scope

[10] *Copyright Act*, § 107.

[11] *Campbell v. Acuff-Rose Music*, 510 U.S. 569 (1994).

[12] *Copyright Act*, § 107.

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-07-02 via free access

and application, and rulings have tended to be *against* fair use.[13] However, in a recent article, Jonathan Band claims, "…fair use pessimism, especially in the educational context, is unfounded."[14] Band cites three recent court cases involving commercial interpretations of the fair use clause. In each of the cases, the courts ruled in favor of the defendant's claim of fair use. None of these cases involved educational institutions, but rather than diminishing the relevance of the cases, that fact bolsters the argument for educational fair use. If federal courts ruled in favor of commercial fair use, Band argues, "similar types of uses in a nonprofit educational context are *a fortiori* fair."[15] Band's analysis applies to published works, but much of his argument could be extended to include unpublished materials as well. Approaching the first factor with Band's fair use analysis in mind, it is likely that our use would be considered fair: the Southern Historical Collection is an educational institution, and the digital collection we intended to create would not be used for direct commercial gain.

It might also be possible to argue that making the letters accessible online as a collection for scholarly research purposes is a *transformative* use of the materials—presenting the letters as a collection adds research value, and new meaning can be gleaned from the body of materials as a whole, rather than as the sum of its parts. A derivative work may be considered transformative if it repurposes or adds new meaning to the original. Transformativeness, especially in an educational context, is a consideration being given increasing weight by the courts.[16]

The second factor—the nature of the work itself—deals with whether the work is creative, making it less reasonable to claim fair use, or of a factual nature, where fair use is more likely to apply. This factor also considers whether or not a work has been published. Traditionally, in cases where the work in question is unpublished, the courts have ruled against fair use. However, while this is certainly still an important consideration, the current trend is moving away "from an apparent 'per-se' bar on fair use for unpublished works."[17] The combination of the other 3 factors can also override this one. For example, in *Bill Graham Archives v. Dorling Kindersley Ltd.*, the court "found that in cases

[13] For a discussion of these court cases, including *Harper & Row Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539 (1985) and *Salinger v. Random House, Inc.*, 811 F.2d 90 (2d. Cir. 1987), see Kenneth Crews, *Copyright Law for Librarians and Educators* (Chicago: American Library Association, 2006), 103–107.

[14] Jonathan Band, "Educational Fair Use Today," *Association of Research Libraries* (December 2007): 2.

[15] Band, "Educational Fair Use Today," 12.

[16] Peter Jaszi, comment on "More on Educational Fair Use—From an Unexpected Source," *Collectanea* blog, posted on 5 August 2009, available at http://chaucer.umuc.edu/blogcip/collectanea/2009/08/more_on_educational_fair_use_-.html, accessed 15 April 2010.

[17] Crews, *Copyright Law for Librarians and Educators*, 106.

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-07-02 via free access

involving transformative uses [as mentioned above], the second factor should be given 'limited weight'."[18]

For the digitization of the Thomas E. Watson correspondence, we intended to deliver reproductions of entire letters rather than selections from them, which, in the context of the third factor, "the amount and substantiality of the portion used in relation to the copyrighted work as a whole," could find against a fair use argument. However, when considering the other factors, our case for fair use is strong, and, as Band argues, the amount and substantiality of the portion used "has less relevance, particularly if the use is transformative."[19] Additionally, a recent decision by the Fourth Circuit Court of Appeals in *A. V. v. IParadigms, LLC* found that "The use of a copyrighted work need not alter or augment the work to be transformative in nature. Rather, it can be transformative in function or purpose without altering or actually adding to original work."[20]

The fact that these materials are letters never intended for publication or to be used in a commercial manner points in favor of fair use in consideration of the fourth factor, the effect of the use on the potential market. Since there was no preexisting and little likelihood of a potential commercial market for these materials—the correspondents almost certainly never intended to publish or sell these letters—the effect on the market value of these works is very low.

One possible reason why some libraries and archives avoid applying the fair use privilege is that they fail to satisfy all 4 fair use factors. Clearly, however, the factors are to be used as a guide: a use need not be considered fair in all 4 factors to be considered fair overall.[21] Unfortunately, since case law determines the application of the fair use law, the only way to determine whether a use is fair is to have it resolved in a federal court.[22] The thought of such a court battle constitutes a worst-case scenario for us, but given the precedents already set by the courts, that is unlikely to happen. Additionally, the plaintiffs would only be eligible for *actual* damages, meaning lost licensing revenue, which, for materials of this type, would be so low that a plaintiff would have very little incentive to bring such a suit.[23]

[18] Band, "Educational Fair Use Today," 10, quoted from *Bill Graham Archives v. Dorling Kindersley Ltd*, 448 F.3d. 605 (2d Cir. 2006).

[19] Band, "Educational Fair Use Today," 13.

[20] *A. V. v. IParadigms, LLC*, 562 F.3D 630 (4th Cir. 2009).

[21] Crews, *Copyright Law for Librarians and Educators*, 42.

[22] Stanford University Libraries, "Fair Use," chapter 9 in *Copyright and Fair Use*, available at http://fairuse.stanford.edu/Copyright_and_Fair_Use_Overview/chapter9/index.html, accessed 15 April 2010.

[23] *Copyright Act*, § 412. Section 504(c)(2) applies as well: if a court found that we had reasonable grounds to believe that our use of a copyrighted work was a fair use, *even if it were to be determined that the use was not fair*, because we are a nonprofit educational institution, statutory damages would be remitted.

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-07-02 via free access

Conclusion

At the completion of our copyright study, we took the results of our research to the legal counsel for UNC University Libraries and explained that we wanted to discontinue any further copyright investigation for the rest of the materials in the Thomas E. Watson Papers. Happily, the level of risk we were undertaking was determined to be an acceptable one, especially given our liberal take-down policy wherein challenged items may be removed from the website quickly and easily. We were given the go-ahead to make the digital collection available online under the auspices of fair use, and we did so in the fall of 2009. To date, the Thomas E. Watson Papers Digital Collection has received no contact, much less challenge, from potential copyright holders.[24]

Extrapolating from our experience with the Watson correspondence, we believe that attempting to explore copyright status in depth and to obtain permission to digitize unpublished archival materials that are under copyright would stymie an effort on the scale anticipated in digitizing the entire Southern Historical Collection. Moreover, such an attempt would be needlessly expensive and futile. If we hope to make large-scale digitization an integral part of processing archival materials, it is untenable for us to consider undertaking this type of research to determine and obtain copyright—we must develop a new definition of due diligence for this type of copyright exploration.

OCLC Research sets forth guidelines "to establish a reasonable community of practice to increase our ability to significantly improve access to *collections* of *unpublished* materials by placing them online for the purpose of furthering research and scholarship."[25] This one-page set of recommendations balances respect for intellectual property concerns with our responsibility to provide research access to cultural heritage materials. Following such guidelines and continuing to investigate fair use as a safe harbor constitutes a much more reasonable course of action for dealing with copyright concerns than the course we investigated for the Thomas E. Watson correspondence. If we are willing to calculate and assume some degree of risk and to document our decisions, archives and libraries can move forward with large-scale digitization, meeting researchers' needs and expectations, and defending our position in the unlikely case that a challenge is brought against us in the form of a lawsuit.

[24] The collection is available at http://www.lib.unc.edu/dc/watson.

[25] OCLC Research, "Well-intentioned practice for putting digitized collections of unpublished materials online," available at http://www.oclc.org/research/activities/rights/practice.pdf, accessed 15 April 2010.

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-07-02 via free access