Copyright Risk Management: Principles and Strategies for Large-Scale Digitization Projects in Special Collections

Kevin L. Smith, Director of Scholarly Communications, Duke University

The foundational premise of this article is that librarians and archivists frequently practice a form of self-censorship when making decisions about digitization of special collections and unique local holdings. This is hardly a controversial assumption, and it was nicely documented in ARL's 2010 report on "Fair Use Challenges in Academic and Research Libraries."[1] In that document, interviewees report reluctance to undertake digitization projects because of uncertainty, and a tendency to select only the safest and most homogenous collections. As one interviewee expressed this view, "We have a lot of things in the public domain, that's the 'easy pickins' for digitization…. We haven't gotten into controversial ground."[2] The authors of the report elaborate on this tendency when they write:

> The challenge is particularly steep when librarians confront mixed collections that include "orphan" works (works whose copyright-holder is unknown or unreachable) and works, such as musical recordings or video, that implicate multiple rights and rights holders.[3]

In the early stages of library digitization projects, this preference for collections that were "safe" and easy to understand in terms of copyright analysis was not particularly problematic. But as the pace of projects increases, it is more and more troubling to realize that decisions are being made not based on scholarly needs or the importance of the material itself, but merely to avoid controversy and risk. In some cases the attitude is that it is better to have some digital material and to avoid risk than to have digital collections that are truly useful and beneficial to the scholarly community. These decisions are made in spite of the discomfort many librarians feel with the "distortion of mission and the incompleteness of [the] resulting digital collections."[4]

Risk Management as a General Practice in Libraries

What often is not recognized in these discussions about copyright is that this is one of the only legal issues in higher education in which the attitude of "no risk at all" prevails. For a wide variety of other areas we undertake to manage risk for the very sound reason that we know we cannot eliminate it entirely. As administrators of large and heavily used buildings, for example, librarians know that there is always a risk of tort claims based on negligence. They put procedures in place to deal quickly with spills or broken furniture to reduce the likelihood of injuries and negligence claims, partly because this is good risk-management practice. In a similar way, libraries, like other employers, post information about channels and protections for employees reporting discrimination or harassment. This practice also is managing a risk that is omnipresent yet capable of reduction but not elimination. In this context, it is curious that copyright is often treated differently—not as a subject of risk management but as an obstacle that must either be avoided completely or allowed to completely block a desired digitization project.

This article contends that copyright should be treated in the same way as other risks of legal liability, as a subject of risk management. One reason that this may not often be the case is that copyright law

seems more complex than negligence or employment discrimination law. In those areas there are well-established practices that library administrators can follow in order to avoid some of the potential risks, whereas copyright law seems like a morass out of which it appears too difficult to select the right questions to ask and principles on which to rely for complex digitization projects. The purpose of this paper is to outline two fundamental principles of copyright risk management for mass digitization and four strategies to implement those principles.

Let me emphasize that nothing presented here is legally innovative or startling in any way. The principles and strategies proposed are entirely straightforward and commonsensical. The goal of this article is to prompt library practitioners to reconsider how they regularly think about copyright law and large-scale digitization, not to make any creative legal arguments.

Principles in Copyright Decision Making

Librarians tend to focus on a single copyright principle when considering a specific potential digitization project. The reasoning often seems to be that digitization can only proceed if **all** of the subject materials are in the public domain, or only if a convincing fair-use argument can be made that applies to **all** of the material. There is no basis in the law for this assumption, and a risk-management approach can help clarify the way in which the different exceptions and limitations in copyright law can work together to reduce the risk of conflict or liability.

> **First, try to reduce the number of risky items that a collection contains. Second, try to reduce the number of people who are likely to want to sue you over the collection.**

For large-scale digitization projects that involve heterogeneous materials from the period when copyright protection may persist, there are two simple principles that a library administrator seeking to manage risk should apply. First, try to reduce the number of risky items that a collection contains. Second, try to reduce the number of people who are likely to want to sue you over the collection. This may seem almost laughably obvious, but thinking about a project in terms of potential points of contention and potential litigants happens fairly infrequently and can be very productive in terms of risk management.

One obstacle to this kind of reasoning is sometimes that librarians simply do not want to think in terms of potential lawsuits. Even abstract reasoning about the potential of getting sued can make one nervous, and if librarians are talking with their university counsel it also may provoke an adverse reaction. But in reality we are thinking about potential lawsuits when we post discrimination and harassment procedures as well, although we are also considering the health of our working environment. All legal considerations in libraries involve some attention to avoiding legal conflicts, and there is no sound reason that copyright should be treated differently. Indeed, considering large-scale digitization in this way has a significant benefit. Once the questions about potential points of contention and potential plaintiffs are considered, the relatively low risk involved in many projects will become apparent, as will the strategies that can be pursued to further reduce that risk. What initially may seem a frightening subject to consider—who might sue us over what material—actually proves to be quite empowering when applied to many collections that might be considered for large-scale digitization.

Four Strategies for Evaluating Risk

The **first strategy** for evaluating the risk associated with a digitization project is to recognize that, in many collections, at least some of the material will be in the public domain. This recognition helps us reduce, in our perception of a project, the number of risky items involved. As has been noted, the public domain is usually considered, in digitization planning, only when an entire collection is likely to be in the public domain. That means that all of the material must be published before 1923 or unpublished and created by persons who died prior to 1940.[5] But mixed collections of 20th-century material will also contain public-domain materials, although determining exactly which items are which may be difficult. Publication date and creator's death date are usually discoverable; what is more difficult is the determination of the copyright status of materials published between 1923 and 1989. John Wilkin has written a comprehensive analysis of this problem, which he calls "bibliographic indeterminacy."[6]

Indeterminacy is only a problem, however, if one is seeking certainty. From the perspective of risk management, it is enough to recognize that most collections of 20th-century materials will contain some public-domain materials. Some materials, for example, will have been published between 1923 and 1963 without copyright notice, or will not have had their copyrights renewed. This category of works is estimated to include about 55% of the books published during this period,[7] and it will often be true of newspaper or magazine clippings as well. Another group of materials will be works of the US federal government, which are not eligible for copyright protection. Amongst the unpublished works in a given collection, such as letters, some will have been created by people who have been dead more than 70 years. Once a risk-management approach is adopted, these categories can be seen as broad groupings that reduce the number of risky items in a potential digital collection. Even though exact determinations cannot be made, recognition of the categories and their potential application is an important step in deciding with some accuracy how risky or safe a particular proposal may be.

The **second strategy** for risk management in digitization projects is to ask permission from the people or organization that would be most likely to object to the digital display. Again, we should recognize that it will usually be impossible or impractical to identify every rights holder and ask permission, and no project need depend on meeting such an impossible standard. The principle of reducing the number of people likely to sue suggests that asking permission even from only a few rights holders, especially those who seem likely to hold rights in a substantial portion of the included material, is an important step in risk management. If there is a large number of clippings from a particular newspaper whose publisher still exists, or a large number of letters by a single author whose heirs can be identified, these are good candidates for permission. Literary estates, which often police the use of works by a particular author, are also good candidates for permission. But it cannot be emphasized too often that asking permission from some large or prominent rights holders does not mean that permission must be obtained for every item in a digital collection. The goal is to reduce the number of likely plaintiffs and to head off those who seem most likely to object as part of an overall risk-management strategy. This approach can significantly increase confidence without creating an insurmountable obstacle.

Also stemming from the principle of reducing potential plaintiffs is a **third strategy** of having, in advance, a take-down policy for any materials made subject of a complaint. Digital collections generally

garner few complaints, but in the rare circumstances where a family member objects that an ancestor's letters, for example, are being displayed to the world, they will often be mollified if the material is removed from public view and the objector is invited to discuss the matter. Sometimes these discussions may result in eventually reposting the work(s) in question, in paying a small licensing fee, or in deciding to leave the material out of the collection. But a responsive take-down policy will inevitably have the effect of preventing most complaints from ever becoming lawsuits.

It is important to note that a take-down policy in the context of library-created digital collections does not have a legal status; it does not create the "safe harbor" that the take-down process outlined in the Copyright Act offers to Internet service providers.[8] There is therefore no guarantee that a rights holder could not or would not sue for an alleged infringement even after the offending materials were removed. But from a risk-management perspective, this is an effective way to defuse conflicts if they arise and will further reduce the anxiety around a digitization project.

The **final strategy** in this arsenal of risk-management techniques is, of course, recognizing that many collections will be supported by a strong fair-use argument. While it is not necessary to rehearse all of the details about what such an argument would look like, two points are important. First, most of the mixed digital collections of 20th-century material to which these strategies would apply will clearly be transformative; such collections will repurpose the individual materials around a research theme, in most cases, that will be far different than the original purpose of the works and will not in any way compete with that original purpose. Second, there is unlikely to be any market for the original in many of these cases, even if licensing markets are taken into account. So the two arguably most important fair-use factors, the first and fourth, will often favor the creation of these digital collections.

When we discuss fair use it is import to consider its application and impact on digital collections decision making. In a mixed and heterogeneous collection, any fair-use argument fabricated in advance of a specific complaint will not apply equally well to all materials. But as we have said, this approach does not require absolute certainty or universal application. When the goal is to evaluate the level of risk in order to undertake sensible digitization projects, it is enough to recognize that fair use would be a plausible defense, and that a good-faith fair-use defense reduces the availability of damages when the user is a nonprofit educational institution.[9] This is especially the case where fair use is understood to be a part of a wider strategy; a "last line of defense" that would further deter potential plaintiffs if all of the other strategies proved, in some rare instance, to have failed to prevent a complaint. For that limited pool of material that is neither in the public domain nor subject to permission, and for that rare plaintiff who is not satisfied by a take-down process, fair use still provides a boundary to the copyright, and raising it would increase the probability that the plaintiff would decide that a lawsuit was likely to prove too expensive and too unprofitable to undertake.

Applications

This four-prong strategy has been successfully applied to two projects in which I have been involved. In the case of a collection of historic TV commercials, recognition of the public domain and efforts to obtain permission from major rights holders were instrumental in a decision to proceed with the digitization.[10] The fair-use argument provided a kind of "backstop," especially in a couple of instances where a putative rights holder told the library that they were not comfortable giving the asked-for permission. In those

cases, where the rights holders were careful not to deny permission but only to refuse to grant it, the library recognized that its fair-use argument still provided enough security to proceed. In almost two years, this project, which seemed very risky on first evaluation, has generated no complaints (and thus no need to use the take-down policy associated with it) while proving very popular with researchers and the general public. In the other case, a joint venture between four libraries, the importance of the strategies was more administrative; it allowed the libraries involved to convince all of the four provosts of their universities to endorse the project, which was an important step toward obtaining grant funding.

Orphan Works

It is an interesting exercise to consider how orphan works—works that are putatively still protected by copyright but for which no rights holder can be located or successfully contacted—fit into the strategies that have been outlined above.

Orphan works are, by definition, not part of the public domain. But the first strategy, that of recognizing the scope of the public domain relative to a proposed digitization project, may have the effect of helping the librarians planning the project understand that the orphan works problem is smaller than they feared. As Wilkin points out in his article on bibliographic indeterminacy, we often lack enough information to decide, for example, whether a work did have its copyright renewed, so that it is potentially an orphan, or whether it did not and is therefore in the public domain.[11] So analysis of a collection in regard to the likelihood of public-domain materials would also help reduce anxiety over the size of the orphan work problem.

Permission is impossible for orphan works, again by definition. But we should recall that the reason for seeking permission from major or potentially litigious rights holders is to reduce the number of likely plaintiffs a library might provoke. Because rights holders in orphan works cannot be located, and in many cases probably do not even know that they hold rights (as is probably the case, for example, with the heirs of a letter writer), they should not be counted in the pool of potential plaintiffs. Also, if a rights holder for a work previously considered an orphan does surface, a take-down policy and the willingness of the library to discuss the matter will be especially likely to defuse the problem, since the rights holder would have little expectation of profit from the work.

Fair use, of course, applies equally to orphan works as it does to in-copyright works for which the rights holders are known or discoverable. Indeed, fair use is probably itself the best "solution" to the orphan works problem, at least in the context of large-scale digitization of library collections. In addition to the support for a fair-use argument that has already been discussed, the fact that a rights holder has not been discoverable or willing to respond to a permission request further strengthens the fourth-factor argument that the use in question does not harm the market or potential value of the work. By definition, again, orphan works are not subject to normal commercial exploitation or to regular licensing, so the fair-use defense becomes quite strong.

> **...fair use is probably itself the best "solution" to the orphan works problem, at least in the context of large-scale digitization of library collections.**

The fair-use defense for digitizing large special collections is not, of course, entirely uncontroversial, as is shown by the recent litigation brought against HathiTrust and five of its university partners by the

Authors Guild.[12] In considering that lawsuit, however, it is important to distinguish between the legal arguments being made in the case and the long-term goals of the plaintiffs. The actual legal arguments do not involve orphan works for the simple reason that no one who holds rights to an orphan work, properly defined, is a plaintiff in the suit. Although the alleged errors on HathiTrust's initial list of potential orphan works received a good deal of attention, none of those works were ever distributed to the public, and the attention received by the case actually showed that the system of making a list of possible orphan works available in advance of their actual distribution was very effective. In an ideal situation, the Authors Guild would work with HathiTrust to be sure that similar errors do not occur in the future.

In a filing made in this lawsuit in February 2012, the Authors Guild has made a unique and troubling argument about fair use and libraries.[13] In essence they suggest that the explicit library exceptions contained in section 108 of the Copyright Act are the sole provisions for libraries, such that fair use is unavailable as a defense for library activities. If accepted, of course, this argument would severely curtail the options for digitizing special collections. But it seems very unlikely that a judge would accept such a suggestion. For one thing, this position would place libraries at a distinct disadvantage against all other potential users of copyrighted content, an outcome clearly at odds with the privileged position usually afforded to libraries by Congress and the courts. Even more decisive, however, is the inclusion in section 108 itself of a provision that reads, "Nothing in this section…in any way affects the right of fair use as provided by section 107."[14] So while this case bears watching and should be a matter of concern to all librarians, the clear intention of Congress ought to prevail, so that fair use will remain a significant option for libraries contemplating digitization projects.

Conclusion

None of the strategies outlined in this article are unique or innovative. The important thing is for librarians to understand how they can work together to provide a more complete picture of the copyright situation involved in a proposed digitization project and a more accurate assessment of the potential risk. Copyright law often seems unmanageably complex, leading librarians to focus too much on a single aspect of a project and, when that aspect proves inapplicable, to give up the proposed digitization. But the multifaceted nature of the law, especially its variety of limitations and exceptions, should really be seen as an invitation to a holistic evaluation that focuses on risk and considers how each facet can contribute to a risk-reduction strategy. If this is done consistently as digitization projects are undertaken, the risk of infringement litigation will usually be seen to be much more manageable, and a great deal of unnecessary self-censorship will be avoided.

1   Prudence Adler, Brandon Butler, Patricia Aufderheide, and Peter Jaszi, "Fair Use Challenges in Academic and Research Libraries," December 20, 2010, http://www.arl.org/bm~doc/arl_csm_fairusereport.pdf.
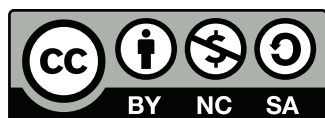
2   Ibid., 12.

3   Ibid., 11–12.

4   Ibid., 12.

5   The best resource for comprehensive consideration of what items may be in the public domain is still Peter Hirtle's chart on "Copyright Term and the Public Domain in the United States," http://copyright.cornell.edu/resources/publicdomain.cfm.

6   John P. Wilkin, "Bibliographic Indeterminacy and the Scale of the Problems and Opportunities Of 'Rights' in Digital Collection Building," Ruminations series, Council on Library and Information Resources, February 2011, http://www.clir.org/pubs/ruminations/01wilkin/wilkin.html.

7   Ibid., citing data from the Copyright Review Management System at the University of Michigan.

8   17 U.S. Code § 512.

9   See 17 U.S. Code § 504 (c)(2).

10 "AdViews," Duke University Libraries, http://library.duke.edu/digitalcollections/adviews/.

11 Wilkin, "Bibliographic Indeterminacy."

12 The documents in this case can be found on the Justia.com website at http://dockets.justia.com/docket/new-york/nysdce/1:2011cv06351/384619/.

13 See entry number 55 in the above docket report.

14 17 U.S. Code 108(f)(4).