Copyright Conundrums
Rights Issues in the Digitization of Library Collections

Sara R. Benson

This chapter addresses the copyright conundrum issues that may arise in the digitization of materials held in academic library collections. Developing a clear understanding of the copyright status of digitized materials can be complicated, frustrating, and even (at times with orphan works) impossible. Nonetheless, when curating digitized library collections, a copyright review and management plan should be followed in order to both maximize the availability of open access and public domain items and minimize the confusion of researchers and the public when utilizing such digitized objects. A copyright review protocol will aid librarians in making decisions regarding what rights metadata to include on a digital work and will, in turn, aid patrons in determining how they can utilize the work.

There is a large distinction that readers should note, however, between determining whether to digitize a particular item at all, say for preservation purposes, and whether to make a particular digitized image publicly available. While Section 108 of the U.S. Copyright Act permits librarians to make up to...
three copies of works for preservation purposes, it does not always allow those copies to be available digitally outside of the premises of the library. Furthermore, when works are not in the public domain, a library may conduct a risk assessment and determine that it is willing to make a copy of the work available in a digital format on the library website either with the express permission of the copyright holder or by asserting a fair use right. It is important to note that these kinds of decisions may fall outside the scope of this chapter, which is more focused on workflows relating to opening up the public domain (such as with the HathiTrust digital library copyright-review process) or with accurately labeling previously digitized works (with the University of Miami’s digital library). The decision of whether to begin digitizing a particular collection at all, especially under a fair use analysis, very well may be the subject of another future book chapter, but it is outside the scope of this particular discussion.

The copyright conundrum begins with the initial copyright review to determine whether a digital copy can be made of a particular item. There are many models available for this type of copyright review, including the HathiTrust digital library copyright-review process. Regardless of the method used, the process should be streamlined, and knowledgeable individuals should engage with the materials in order to determine their copyright status and, where possible, obtain permission to place the materials online in an open, publicly accessible digital collection. The copyright discussion continues by presenting the challenge of deciding whether and how to include a copyright notice on the online work. This portion of the chapter will include common mistakes and misperceptions, such as copyfraud and overreaching. The final piece of the copyright puzzle involves responding to public inquiries to utilize the library’s digital copy of the work in further academic or commercial pursuits. Rights metadata, when properly applied to a digital work, will include information allowing the patron to understand how the work may be used.

The Copyright Law Landscape

In the United States today, there are no formalities required for copyright to attach to a given work. Thus, if a literary, musical, dramatic, choreographic, pictorial, graphic, or sculptural work, a motion picture and other audiovisual work, a sound recording, or an architectural work is minimally creative and fixed (generally meaning that it was written or recorded), it is protected by copyright. Copyright for works created today in the United States lasts quite a long time: the length of the author’s life plus 70 years if the author is an
individual. For works authored by a corporation (or works made for hire), the length of copyright is longer: either 95 years from first publication or 120 years from the year of creation, whichever is the shorter length of time. In determining who the actual copyright holder might be in your copyright review, the U.S. Copyright Act provides some guidance. The Copyright Act defines the work of employees and certain independent contractors this way: “(1) a work made by an employee within the scope of his or her employment; or (2) a work specifically ordered or commissioned for use as a contribution to a collective work . . . if the parties expressly agree in a written instrument signed by them that the work shall be considered a work made for hire.” All terms of copyright run through the “end of the calendar year in which they would otherwise expire.”

In contrast to the law that is in effect today after U.S. accession to the Berne Convention, which mandated that the United States omit the use of formalities for copyright protection, the copyright status of works first published in the United States between 1923 and 1963 can get a bit more difficult to decipher. This is due to changing laws in that era, and the requirement of copyright formalities such as including a copyright notice on the work, registering the work with the U.S. Copyright Office, and renewing the registration at the appropriate time with that agency. Failure to comply with these formalities led to many works published in this era falling into the public domain, meaning that they are not subject to copyright protection. (Just what percentage of the total published works from that period are in the public domain is up for dispute.) It is important to understand some basic principles when searching for rights information during this era.

The era between 1923 and 1978, however, is a more challenging period in the copyright realm because of changing rules under U.S. copyright law in that time frame. For instance, a work published in the United States without a copyright notice between 1923 and 1977 is also in the public domain, as is a work published between 1923 and 1963 with a copyright notice but a failure to renew the copyright registration. Furthermore, during that time period, the length of the copyright was 95 years from the date of publication, whereas it is generally the life of the author plus 70 years beginning in 1989. In the period between 1978 and 1989, the failure to include a copyright notice did not push a work into the public domain, so long as the copyright owner registered the work within five years of publication.

However, a significant challenge in digital collections is that many works were never published, which is then further complicated by a lack of knowledge regarding the identity of the work’s creator, let alone the death date of...
the work’s creator. It is worth noting that the definition of an “unpublished work,” while it may seem fairly straightforward, gets complicated quickly. The Copyright Act defines publication as “the distribution of copies . . . of a work to the public by sale or other transfer of ownership, or by rental, lease, or lending.” The act also notes that “the offering to distribute copies . . . to a group of persons for purposes of further distribution, public performance, or public display, constitutes publication.” But the act excludes public performance or the display of a work, alone, from constituting publication. And courts, when interpreting this language, also struggle with the definition of what it means for a work to be “published,” often in cases predating the 1976 Copyright Act, which provided the above definition. In the era before 1976, therefore, it was even more difficult to determine what kind of work was an “unpublished” work, and it was even more crucial in that era to decide that fact due to the attachment of copyright formalities as detailed above. The cases from this era demonstrate that judges tended to “rely heavily on the copyright owner’s intent with respect to authorized copies.” Also “when the facts show that [works] are distributed freely, it weighs in favor of publication and when the work is made accessible in a way that demonstrates that the copyright owner is retaining control over the copies, publication is less likely to be found.” Determining the publication status for works created before 1976 gets complicated quickly and can depend on what a court might interpret as “publication.”

For previously unpublished works, such as diaries, photographs, manuscripts, and the like, the default rule for copyright term in the United States applies when the death date of the author is known: the life of the author plus 70 years. Thus, unpublished works by authors who passed away before 1947 are currently in the public domain. If the death date of the author is unknown, the term extends to 120 years from the date of creation. Finally, for unpublished anonymous or pseudonymous works, or for works that were commissioned by an employer in a work made for hire situation, the length of copyright is also 120 years from the date of creation.

When analyzing copyright ownership, one should also consider any transfer of copyright through contractual agreement. With older copyright agreements, even if the publisher owns the copyright and could (ostensibly) provide permission for the use of the work, the potential exists that the publisher is currently defunct or that the copyright has been further transferred to another party. A further complication in locating the owner of the copyright may arise because in certain circumstances authors or their heirs can “take back” their copyright from a publisher or other owner within a given period of time (35 or 40 years after the execution of the transfer of copyright by the author after

January 1, 1978, depending on the circumstances). Additional rules apply to pre-1978 transfers. Potentially in a copyright review, a reviewer may run into the problem of so-called orphan works, where no copyright owner can be determined at all. Or the reviewer may encounter foreign works, which may have even longer copyright protection terms in the United States, due to restoration, than in their home countries.

If a particular two-dimensional work, such as a painting, is determined to be within the public domain, then a picture of that piece of art is also in the public domain. This is so because, in this instance, a photographer attempting to merely document an exact “slavish” copy in a photograph of a public domain work does not represent the minimal amount of creativity necessary to create an original work capable of copyright protection. Thus, if a library attempts to make an “exact” digital replication of an existing two- or three-dimensional work of art without additional creative choices added to the process of digitization (including the photography), the library likely does not hold a copyright in the digital replication. As such, libraries in the United States generally do not treat duplicate copies of images as having a separate copyright.

Library patrons consulting digital collections will also note that there may be terms of use or terms of service associated with the library’s website. By accessing the web page, patrons consent to the contractual terms of service. Copyright provisions do not preempt contractual terms. Thus, should a library wish to charge for access to a particular item, even if that item is in the public domain, the library may do so under a “value added” (or fee for the service, or for the staff time used in producing the image) proposition, or the argument that the preservation and archiving of the particular item is worth the licensing fee provided for in the terms of service. This is not to say that the author condones such a practice, but rather to note that contract law and copyright law are different legal regimes.

The Potential for “Copyfraud”

The term copyfraud was coined by Jason Mazzone and is intended to describe instances when a non-copyright holder asserts rights in a work that they have no copyright ownership over. For instance, if a work is in the public domain, but a library asserts that it holds copyright over the work with a copyright statement such as “© 2017 Library,” then Mazzone would likely assert that the library is engaging in “copyfraud.” Note that if a person does this intentionally and with a fraudulent intent, that individual may be guilty of a criminal...
violation as well.\textsuperscript{28} The term \textit{copyfraud} is a bit harsh because, technically, fraud is an act involving malicious intent. The library is typically not engaged in an intentionally bad act, but rather is merely a bit unclear or ignorant of the copyright laws. In this way, the library may inadvertently be misleading the public, but not in any malicious or intentional way. Regardless, the consequence is the same: the public may be left with the impression that the copyright to the object in question is owned by the library when, indeed, it is not. Because of these kinds of errors or imprecision in rights metadata and an absence of a standard approach to rights statements in general, the Digital Public Library of America (DPLA) and Europeana created what they call “Standardized Rights Statements” (SRS).\textsuperscript{29}

Local Approaches to Rights Statements Before SRS

In order to accurately describe digital collections, and before the development of standardized rights statements, a few conscientious librarians had developed local approaches to rights management system processes.

One locally produced system was developed by Maureen Whalen in 2009 for the Getty Museum. Whalen developed rights metadata to apply on a consistent basis across the Getty Museum’s online digital collections.\textsuperscript{30} One helpful thing to note about the collection is that all of the work was owned by the Getty family and was explicitly willed to the museum, so in that respect, this particular system may not be as useful to other librarians wishing to apply standardized rights metadata to their own digital collections. Nonetheless, the choices made by Whalen regarding which fields to incorporate into the Getty rights statements is informative. The Getty included five “priority” metadata fields: creator name, copyright status of the work, publication status, copyright notice, and credit line.\textsuperscript{31} Other fields were also included, such as potential claimants, creator role, special notes, and the like.\textsuperscript{32} The copyright status field further broke down into eight separate sub-indicators: copyright owned by institution, copyright limited license to institution, copyright owned by third party, public domain, orphan work, unknown (“research was conducted and, as of that date, no reliable rights information has been found”), additional research required (“research was conducted, but it is [in]conclusive”), and not researched.\textsuperscript{33} Interestingly, many of the copyright fields operate well with the SRS. However, the SRS do not take into account rights other than copyright,
such as potential claimants based on a right of privacy or publicity, while the Getty metadata does (likely due to the fact that many of the individuals featured in the photograph collection are or were famous).

The system for rights metadata developed by Karen Coyle, who worked for over twenty years at the California Digital Library, in 2005 to address copyright in all digital collections was a precursor to the SRS project. Coyle asserts that copyright metadata should be accurate, even when licensing agreements exist, because “a license does not remove the copyright status of an item; it establishes an agreement between the parties that is founded on the ownership rights that copyright law defines.”34 Thus, while she recognizes the importance of licensing, her metadata development does not include a field to input “licensing” information. Rather, she included the following rights fields in her copyright metadata: copyright status “(copyrighted, public domain, unknown),” publication status “(published, unpublished),” dates (year of copyright or creation; year of renewal of copyright); copyright statement “(from the piece),” country of publication or creation, creator (creator name, dates, contact), copyright holder, publisher and year of publication, and administrative data such as the rights research contact.35 This important work predates the SRS and demonstrates the importance of utilizing standardized metadata in digital collections.

Copyright Review Models

First, it is important to note that there may be no “one-size-fits-all” copyright review process, even though there is a standardized metadata input mechanism such as the SRS. Why? Because each institution’s holdings are different and their records may be more or less detailed. Additionally, their staff may be more or less familiar with copyright laws as well, and may feel uncomfortable conducting any type of copyright review if the materials in the digital collection are not owned by the institution. Although it is preferable to conduct a thorough copyright review of each digital item in a given online library collection, it is understandable that a library may not have the staff or the time to do so. However, at the bare minimum, libraries should attempt to input accurate data, and not use a default metadata rights identifier such as “©Library” when it is incorrect.

Second, copyright is often impacted by licensing and/or terms of use limitations. However, when licensing or terms of use are placed in the rights field, the user may be unnecessarily confused. Thus, the information about licensing/
terms of use should be separated from the rights information in a clear fashion; when appropriate, libraries should also use the “No Copyright—Other Known Legal Restrictions” label for the metadata.

**Copyright and Cultural Institutions Book**

Peter Hirtle, Emily Hudson, and Andrew T. Kenyon detail helpful information for librarians wishing to engage in rights status analysis for digital collections in their open-access digital book titled *Copyright and Cultural Institutions: Guidelines for Digitization for U.S. Libraries, Archives, and Museums*. Some of the most useful portions of the book, from the perspective of a librarian conducting a copyright review, are the charts and checklists. Specifically, table 3.2.1 details the copyright term length for unpublished works. Table 3.2.2 similarly details copyright term length, but for works first published in the United States. Flowchart 3.2 explains how to determine the rights status for works published in the United States between 1923 and 1989. Table 3.2.3 demonstrates how to calculate the rights status for foreign works. Finally, flowchart 6.1 details when digitized copies of works may be made available under Section 108 of the Copyright Act, specifically when they are in the last twenty years of the copyright term and the work is not subject to normal commercial exploitation, no copy of the work can be obtained at a reasonable price, or the copyright owner has notified the Register of Copyrights that either of these preceding two conditions have been met. Of course, the entire book is helpful to those wishing to engage in a thorough rights analysis, but these checklists and charts are very handy guides to apply when devising a review process.

More recently, the Society of American Archivists (SAA) further built on this work by specifically writing a guide for the implementation of rights statements for “archivists and other cultural heritage professionals.” The guide is freely available on the website of the SAA.

**HathiTrust Copyright Review Management System Model**

The HathiTrust Digital Library is “a digital preservation repository” that provides “access services for public domain and in copyright content from a variety of sources, including Google, the Internet Archive, Microsoft, and in-house partner institution initiatives.” It has a very detailed copyright review model, the Copyright Review Management System (CRMS), which it documents through an open access e-book. The most interesting part of
this model, perhaps, is that it is collaborative in nature. Many libraries dedicate staff members to participate in HathiTrust’s review process, resulting in a large-scale review of the copyright of published works. In the review process, each book is reviewed by two nonexperts and one expert reviewer, in case of a conflict between the reviews by the nonexperts. The book is made openly available through the HathiTrust Digital Library if it is in the public domain. Alternatively, if a book is still in copyright, it is available to display the page numbers and the number of times the search term appears on relevant pages for the term searched. Verified researchers can access in-copyright materials for nonconsumptive, or text-mining, research, through an application process to the HathiTrust Digital Library.

Note that HathiTrust is currently focused on reviewing works that were published in the United States between 1923 and 1963 (class A books), published in the United Kingdom before 1943, or published in Canada and Australia before 1963. Other areas of work include the evaluation of U.S. state and local governmental material that was published between 1923 and 1977.

The copyright management “decision tree” is one piece of the documentation created by the CRMS project that is valuable to anyone wishing to review the copyright for works published in the United States between 1923 and 1963. Indeed, all of the decision review documents created by the CRMS team, as well as the book published on the subject—Finding the Public Domain by Levine et al.—are invaluable tools for those wishing to engage in a metadata rights determination project without reinventing the wheel.

Once the rights decision tree is followed and an appropriate copyright determination is made, the DPLA/Europeana SRS may be applied. HathiTrust currently uses metadata through Zephir, a bibliographic programming system managed by the University of California and the California Digital Library.

University of Miami Libraries

In 2015, the University of Miami Libraries started a project to assign right statements to their digital collections by using standardized rights statements. The libraries noted that before this project was undertaken, “the majority of the Libraries’ digital collections contained little to no rights-related information on their metadata.” In some ways, this may have made the implementation of the SRS easier, since it may be simpler to add a rights field than to change an imposed rights data. The libraries created a copyright decision matrix, drawing largely on the work of Peter Hirtle (outlined above), which is freely accessible...
on the Internet. In essence, the University of Miami Libraries combined a decision matrix with SRS implementation in order to provide a model approach for libraries wishing to begin assigning rights statements to their digital collections. Interestingly, large amounts of the University of Miami Libraries collection came from Cuba, so the decision matrix includes a “country of creation” designation specifically to notify staff when a given work was created in Cuba. The University of Miami undertook an effort to understand Cuban copyright law for similar reasons. Additional information about the approach taken by the University of Miami is available on the libraries’ website.

Pennsylvania State University’s Workflow for Rights Statements

The Pennsylvania State University Libraries (commonly known as the Penn State Libraries) are the most recent example of the implementation of standardized rights statements. In spring 2016, the Penn State Libraries held a retreat with relevant library stakeholders in order to discuss the workflow surrounding metadata for digital collections. Many currently digitized collections at the Penn State Libraries require metadata revisions in order to satisfy the requirements for deposit to the DPLA collection. Newer collection deposits would, of course, follow the new protocol for rights metadata.

The first step in a digitization process at the Penn State Libraries is for the copyright officer to determine whether the collection items are public domain, whether the library owns the rights, or a fair use determination positively supports digitization. If not, the process may be put on hold pending permission requests to the copyright holder. Next, the copyright officer assigns a relevant rights statement utilizing the language of the standardized rights statements. Finally, the metadata librarian records the rights information in the record for the item. Interestingly, Penn State discovered through trial and error that the URIs for the rights statements are case-sensitive (which in DPLA will resolve “to display the official [rights statement] icon near the item’s thumbnail image”).

Standardized Rights Statements

As noted above, the standardized rights statements were launched in April 2016 by the DPLA and Europeana to create an unambiguous, uniform
vocabulary for inputting rights metadata. There are three main categories of rights statements contained in the standardized rights statements:

- In Copyright
- No Copyright
- Other

The three categories are further subdivided into more specific categories. For “In Copyright” works, for instance, the additional subcategories include:

- In Copyright
- In Copyright—Educational Use Permitted
- In Copyright—Non-Commercial Use Permitted
- In Copyright—Rights Holder(S) Unlocatable or Unidentifiable

The simplest subcategory for “In Copyright” is just that—“In Copyright.” When a work is clearly still protected by copyright and the rights holder is known, this statement is appropriate. The next few designations listed above permit noncommercial or educational uses only. Finally, a designation for “Unlocatable or Unidentifiable Rights Holder” is appropriate when the work is clearly still in the copyright term, but the rights holder is unknown—this would be appropriate for an orphan work. Each category of rights statements is fully described and detailed both on the SRS website and in the SRS White Paper (http://rightsstatements.org/en/; http://rightsstatements.org/files/180531recommendations_for_standardized_international_rights_statements_v1.2.2.pdf).

The “No Copyright” rights statements further subdivide into four categories:

- No Copyright—Contractual Restrictions
- No Copyright—Non-Commercial Use Only
- No Copyright—Other Known Legal Restrictions
- No Copyright—United States

The first category, “No Copyright—Contractual Restrictions,” would be appropriate if a work was no longer under the copyright term restrictions, but by license or contractual obligation the work has additional restrictions. So, for instance, if a vendor has commercialized a public domain work and has imposed terms of use, this designation would be appropriate. The “No Copyright—Non-Commercial Use Only” designation would be appropriate for works that are designated with a license to be open, but only for noncommercial uses. The “No Copyright—Other Known Legal Restrictions” designation would be appropriate if the work is not restricted by copyright, but rather by other legal
rights such as privacy or moral rights. Finally, the “No Copyright—United States” designation would be an appropriate license for U.S. public domain materials, such as those published in the United States prior to 1923. This standardized rights statement type is designated as “United States” because public domain terms can vary by country and by one country’s treatment of foreign works under its own law (such as the restoration of foreign copyrights under U.S. law), and this statement has only been vetted for the U.S. public domain term.

The “Other” category is akin to providing catchall provisions. It is broken down into three subcategories:

- Copyright Not Evaluated
- Copyright Undetermined
- No Known Copyright

“Copyright Not Evaluated” is a provision that lets the user know that the copyright has yet to be evaluated by the hosting institution in any way, for instance, with the mass digitization of works. Although the digitizing institution may believe that the materials it is digitizing belong in the public domain, it has not done an individualized assessment of each piece in the collection. The “Copyright Undetermined” designation is intended for use when the institution has reviewed the item and has “made the item available, but the organization was unable to make a conclusive determination as to the copyright status of the item.” A library may wish to use this designation if it has done a copyright review, but was unable to determine the status of the item because the author’s death date is unknown and it is necessary to calculate copyright length, for instance. Finally, “No Known Copyright” is an appropriate designation for a work made between 1923 and 1968 where a reasonably diligent online search has been conducted and no copyright can be located. It indicates that this is not a simple case, like a work published in the United States prior to 1923, and that a copyright search has been made, but that the organization cannot warrant the accuracy of the information to an infallible degree.

Note that all of the rights statements include a disclaimer that there is no warranty as to the accuracy of the rights statement and that the user of the work is ultimately responsible for his or her own use.
Conclusion

Although it may be time-consuming and difficult to correct errant online digital rights statements, it is important that libraries make the effort to do so. Otherwise, as noted above, there is a risk of copyfraud. Thankfully, the SRS documentation provides librarians with uniform guidance to appropriately include rights metadata for the materials in their online collections. Librarians no longer need to invent homegrown systems for copyright metadata. However, using correct terms for metadata is only half of the copyright battle. The harder part of the challenge is to develop systematic approaches for copyright review. Luckily, HathiTrust, the University of Miami Libraries, and Pennsylvania State University Libraries have taken the lead in providing a road map for such decisions. These examples, taken along with the copyright flowcharts from the Copyright and Cultural Institutions book, provide librarians with guidance on how to appropriately chart copyright decisions in order to determine which rights statements to apply to works housed in a digital collection. Hopefully, many more libraries will follow this lead and SRS metadata will become the norm and not the exception to the rule for digital collections.

NOTES

2. Ibid., § 302.
3. Ibid., § 101.
4. Ibid., § 305.
8. Ibid.
9. Ibid.
11. Ibid.
13. Ibid.

14. Ibid., 204.
15. Hirtle et al., *Copyright and Cultural Institutions*, 42.
16. Ibid.
17. Ibid.
18. Ibid.
20. Ibid., § 304.
22. Hirtle et al., *Copyright and Cultural Institutions*, 49–51.
24. Ibid.
28. Criminal penalties may be assessed under the Copyright Act if a person “with fraudulent intent, places on any article a notice of copyright . . . that such person knows to be false”; 17 U.S.C. § 506(c) (2012).
32. Ibid., 26–28.
33. Ibid., 23–24.
35. Ibid., 7.
36. Ibid., 28.
37. Hirtle et al., *Copyright and Cultural Institutions*.
38. Ibid., 41.
39. Ibid., 45.
40. Ibid., 48.
41. Ibid., 49.
42. Ibid., 110.

44. Ibid.


47. Ibid., 115.


49. HathiTrust Digital Library, “Copyright Review Program.”

50. Ibid.


58. Ibid.

59. Ibid., 153.

60. Ibid.

61. Ibid., 154.

62. Ibid., 155.

63. The Digital Public Library of America is an online collection of cultural heritage materials from American museums and libraries; see https://dp.la/info/.

64. Europeana is a digital collection of cultural heritage materials from European museums and libraries; see www.europeana.eu/portal/en/about.html.


66. There is an additional subcategory for “In Copyright—EU Orphan Work,” but that is unlikely to be utilized in the United States and will not be discussed further in this chapter.


68. Ibid., 24–25.

69. Ibid., 23.

70. Ibid., 26–30.

71. Ibid., 27.

72. Ibid., 26.

73. Ibid., 28.

74. In the United States, the only recognized “moral rights” are those contained in the Visual Artists Rights Act, or VARA: 17 U.S.C. § 106A (2012). In other countries, however, moral rights may include rights such as attribution and integrity (including rights against the mutilation of the work) and may last indefinitely. Cyrill Rigamonti, “Deconstructing Moral Rights,” *Harvard International Law Journal* 47, no. 2 (2006): 357.


76. Ibid.

77. Ibid., 30–31.

78. Ibid., 31.

79. rightsstatements.org, “Copyright Undetermined,” http://rightsstatements.org/vocab/UND/1.0/.