

VRA Bulletin

Volume 47
Issue 1 *Spring/Summer 2020*

Article 3

August 2020

Copyright Statements in Plains to Peaks Collective Digital Collections

Summer Shetenhelm

Santa Clara University, sshetenhelm@scu.edu

Follow this and additional works at: <http://online.vraweb.org/vrab>

Recommended Citation

Shetenhelm, Summer. "Copyright Statements in Plains to Peaks Collective Digital Collections." *VRA Bulletin*: Vol 47: Iss. 1, Article 3, 2020. Available at: <https://online.vraweb.org/index.php/vrab/article/view/189>.

This Feature Article is brought to you for free and open access by VRA Online. It has been accepted for inclusion in the *VRA Bulletin* by an authorized editor of VRA Online.

Copyright Statements in Plains to Peaks Collective Digital Collections

Abstract

While digital collections provide users with unlimited access to cultural heritage materials, ambiguities about copyright status limit users' understanding of what can or cannot be done with these materials. This study uses a descriptive research method to examine a sample of copyright statements attached to items in digital collections provided by member organizations of the Plains to Peaks Collective. The Plains to Peaks Collective is a partnership of the Colorado and Wyoming State Libraries that, in Fall 2018, introduced over 181,000 items into the Digital Public Library of America (DPLA). This study seeks to answer the following research questions: what copyright statements are included in Plains to Peaks Collective records that have been ingested by DPLA, and what copyright statements are included for public domain Plains to Peaks Collective items at the local system level? The study reveals that a large number of statements included in these collections incorrectly claim copyright over public domain objects and/or provide conflicting rights information. The study concludes with recommendations about how libraries and archives in the Rocky Mountain area and our greater community can better inform users about rights regarding the reuse of digital objects.

This article has undergone a double-blind peer review process.

Keywords

intellectual property rights, copyright, metadata, digital libraries

Author Bio & Acknowledgements

Summer Shetenhelm is the Digital Collections & Scholarship Librarian, Santa Clara University Library.

Technology has revolutionized the ways in which information is accessed and disseminated, allowing for the instant transmission of journal articles, images, videos, and datasets. However, although we have the ability to access and share these resources, copyright restrictions may mean we do not have permission to. Copyright was created in a print environment, and in many cases cultural institutions are faced with the daunting task of interpreting abstract legal concepts into concrete policies (Hirtle et al. 2009, ix). Misunderstandings around digitization practices and how they affect copyright can arise (Monson 2017; Sims 2017), and these misunderstandings may lead to the misrepresentation of rights to users. In some cases, institutions may make available works that are under copyright or include inaccurate copyright statements in their online collections, such as requiring permission for use of a work that is in the public domain (Benson 2019). Complicating this issue is the inconsistency of language used in copyright statements – for example, the International Rights Statements Working Group found over 87,000 different rights statements for objects (either a digitized or born-digital manifestation of a work) included in the Digital Public Library of America (DPLA) (International Rights Statements Working Group 2016). The Illinois DPLA service hub examined 203,895 records from 120 of their participating institutions and found 938 unique rights values in metadata contributed to DPLA (Stitzlein et al. 2018).

Ambiguities about copyright status and how that status is presented limit users' understanding of what can be done with digital objects. If digital libraries exist to support the research needs of users (Borgman 1999), meet the needs of communities (Calhoun 2014), and “ideally...allow new types of use when digitised” (Jordan 2006, 27), then a clear knowledge of how digitized resources can be reused is essential to creating an effective digital collection. In order to empower users to effectively reuse digitized resources, it is imperative that digital libraries clearly and accurately present copyright status.

The goal of this study is to provide insight into the ways in which copyright is presented to users in digital collections created by institutions in the Rocky Mountain area. The institutions examined are members of the Plains to Peaks Collective (PPC) and contributed items to DPLA in the fall 2018 ingest. By examining the copyright statements attached to these items, we can understand if reuse rights are presented to users clearly. Finally, this study makes recommendations about how to improve representation of copyright in these collections.

Literature Review

Large-scale digital libraries contain immense amounts of information resources contributed by multiple institutions. Xie and Matusiak (2016) describe digital libraries as having the following characteristics: large collections of often more than one million items; multiple resource formats; varying collection policies (one main policy for the large-scale digital library as well as policies of participating institutions); and varying degrees of copyright compliance. In *Exploring Digital Libraries*, Calhoun (2014) provides us with a more general definition: digital libraries are systems that “(a) support the advancement of knowledge and culture; (b) contain managed collections of digital content...intended to serve...defined communities; (c) often use an architecture that first emerged in the computer and information science/library domain” (18). Other definitions collocated by Calhoun focus on the following aspects of digital libraries: computerization; collections of services;

being part of a network; and spaces for collaboration. One aspect of large-scale digital libraries most attractive to users is the collocation of many materials from different sources. Users in a study by Matusiak (2017) stated they liked the “one-stop shop” aspect of digital libraries, because they “don’t have to go to twenty different sites to find something” (163). By combining the collections of multiple institutions scattered across geographic areas, large-scale digital libraries create a platform in which users can have quick and easy access to millions of digitized objects.

One example of a large-scale digital library is Europeana, which provides access to more than 50 million digitized objects (Europeana). These objects include images, books, music, and videos. The mission of Europeana is as follows: “We want to build on Europe’s rich heritage and make it easier for people to use, whether for work, for learning or just for fun.” As per Purday (2012), Europeana “has paid particular attention to sourcing content that represents the diversity of European cultural heritage, and encouraging all countries to provide material” (9). Users are able to browse the Europeana collections by category (images, sound records, text, videos, and 3D objects) or by exhibition (for example, *Edible Plants from the Americas*).

Users are also able to search the collection with keywords and narrow their results by media type, language, providing country and/or institution, and a category called “Can I Use This?”. This category allows users to search by copyright status. Europeana also provides users with information about the public domain and various copyright statements included in the collections and has “begun advocating correct public domain attribution and open data licensing” (Purday 2012, 3). Since Europeana’s mission is to make cultural heritage resources easier for the public to use, it makes sense that they provide users with the capability to search collections by copyright and reuse possibility.

Similar to Europeana is the Digital Public Library of America (DPLA). Launched in April of 2013, the DPLA is a “free, open, and accessible national digital resource” that brings together millions of digital objects into one searchable interface (“About us” DPLA). DPLA does not host or own any of the items featured in its library; like Europeana, it is a metadata aggregate, which means that it gathers information about a large number of objects from different institutions into one place. Users can browse collections by partner institution and topic, view exhibitions, and use a keyword search to find objects. DPLA consists of a series of Service Hubs, which aggregate metadata from organizations within their state or region, and Content Hubs, which are the individual institutions that have created their own digital collections (“Prospective Hubs” DPLA). Content Hubs and smaller institutions that participate in Service Hubs are responsible for contributing metadata for objects that are included in DPLA.

The DPLA’s “About Us” page states that all of the materials found through their service “...are free and immediately available in digital format.” However, under their “Frequently Asked Questions,” the DPLA site states that the “copyright status of items...varies,” and instructs users to “check the ‘Rights’ field in the metadata, or follow the link to the digital object on the content provider’s website for more information.” Because they aggregate metadata from other institutions, DPLA does not have first-hand knowledge of the rights information for any objects users may find through their platform. Contributing institutions are responsible for ensuring that objects included

in DPLA have accurate rights information so users are aware of how they may reuse these digital objects.

Intellectual property rights were identified by Calhoun (2014) as a key challenge for digital libraries. Copyright is one way in which a creator's work is protected. The intention behind copyright was to encourage the creation of new works of authorship by giving the creators of those works exclusive rights for a set period of time. Because of these exclusive rights, creators have the opportunity to achieve monetary gain from their works; this monetary incentive encourages creators to continue making new works and generating new scholarship. Copyright owners have the exclusive right to reproduce their work, distribute copies of or sell their work, and display their work publicly. These exclusive rights legally prevent the reproduction or reuse of these works without express permission from the copyright holder.

Copyright is time-limited, and not all works are under copyright. When a work's copyright expires, that work is now in what is called the "public domain." Works in the public domain have no restrictions; users can reuse them in any way they see fit. As of 2019, any work published in the United States before 1924 (excluding sound recordings) is in the public domain. However, despite being considered public domain in the United States, some works may be subject to the copyright laws of other countries. Copyright is often unclear and can be confusing for users; as such, organizations which make works available online should include clear copyright statements with the works they have uploaded. Certain exemptions allow for the use of a creative work without permission from the copyright owner and without paying for a license; an example of this is fair use, which can include the reuse of a work for scholarship and research, teaching, news reporting, and commenting and criticizing (Hirtle et al. 2009). Students who wish to include a work in an academic paper that is only turned in for class, or a professor who wants to include an image in a slideshow to present in class, could reuse in-copyright items under the defense of fair use. However, if an item were reused in a way that did not fall under fair use – for example, published in a textbook or as part of an open educational resource (OER) – the user would require permission for reuse.

Determining copyright status – and, subsequently, determining if one must request permission for reuse – can be a tricky business. Hirtle, Hudson, and Kenyon (2009) and Benson (2019) describe the complicated way that US copyright has changed over time. Copyright for works published after 1924 can be convoluted and difficult to discover. In 1909, US copyright law extended copyright terms for published works to 28 years and allowed for a 28-year renewal. Later amendments increased the available copyright renewal term to cover 67 years, for 95 total years of copyright for published works. Another big change came with the Copyright Act of 1976. This act, which went into effect in 1978, provided federal copyright protection for unpublished works and established copyright as calculated based upon the life of the author plus 50 years, without the necessity of renewing copyright at a certain time. This meant the copyright term was based upon the lifetime of the author and not a set amount of years.

In 1992, all works published between 1964 and 1978 received a 75-year copyright term. Then, in 1998, the Sonny Bono Copyright Extension Act granted an additional 20 years to the term of copyright, establishing copyright as lasting until 70 years after the death of the author. As such, determining the copyright for works published after 1924 can be incredibly complicated. Factors

such as whether or not a work was published, whether copyright was renewed, and the date of death of the creator affect the copyright status of these works.

As we can see, while the call for institutions to include rights information seems straightforward, this can prove to be a daunting task. Even though many cultural heritage institutions provide resources online, they can face a number of risks when digitizing content. Many institutions are worried about negative backlash from digitizing and making items available (Hirtle et al. 2009; Sims 2017) and/or may not want to share copyright information for fear that “users will take it as legal advice, opening the library to unwanted risk” (Schlosser 2009, 383). Additionally, the copyright status of many works may remain unclear (Capell et al. 2018; Dickson 2010), and, as such, institutions are unable to create an accurate statement of copyright for these works. Hirtle, Hudson, and Kenyon (2009) give us a list of some of the complications institutions may face when determining the correct copyright statement to include with digital objects, including unclear copyright status, inability to identify and/or locate copyright holders, and international copyrights on materials that are public domain in the United States. They also suggest that creators “...may believe they have more rights in material than the law allows, and take umbrage when an institution digitizes the material” (Hirtle et al. 2009, 193). Schlosser (2009) notes that “It is possible that working knowledge of copyright law in many libraries is not sufficient for grappling with the complexities involved,” and as such libraries – although unintentionally – may be presenting false information (383). These obstacles – the reluctance to apply an incorrect rights statement and face potential legal backlash, the inability to establish an item’s copyright status, or a general misunderstanding of copyright law – may prevent many institutions from including certain materials in digitized collections.

Despite these risks, some institutions digitize materials anyway. As an act of good faith (and of institutional protection), institutions may include a statement informing copyright holders that, if an item the institution digitized is being used in a way that violates copyright, they will remove the item upon request from the copyright holder. Alternatively, institutions may post that they have been unable to locate a copyright holder and request that, if a user has information about the copyright status of a digitized object, they contact the institution (Hirtle et al. 2009). This method is especially useful for orphan works, which are works for which a copyright holder, and therefore copyright status, cannot be identified. If an institution decided to exclude orphan works for fear of copyright holder retribution, a large number of digitized objects would not be included in digital collections and as such would not be available to researchers.

The process of identifying rights holders of orphan works can be incredibly time-consuming and difficult. For example, in order to make the entirety of the Thomas E. Watson Papers available, the University of North Carolina at Chapel Hill (UNC-Chapel Hill) had 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” (Dickson 2010, 627). Their investigation took over 450 hours of labor, spread over nine months, and cost approximately \$8000, with an end result of only four additional letters able to be digitized and shared online (ibid). As we can see, identification for orphan works requires huge amounts of time and money. Most institutions simply do not have the resources

necessary to undertake rights investigations for orphan works (Calhoun 2014). In the case of the Thomas E. Watson Papers, the entire collection was made available online under a fair use defense, despite not having copyright clearance; additionally, the level of institutional risk was determined to be low (Dickson 2010). As of 2019, the collection website includes the following statement regarding copyright:

Written permission from the copyright owner and any other rights holders must be obtained for any reuse of these images that extends beyond fair use or other statutory exemptions. Furthermore, responsibility for the determination of the copyright status and securing permission rests with those persons wishing to reuse the materials. (UNC-Chapel Hill Libraries 2019)

This collection passes the burden of determining copyright status down to the user, who is now responsible for using their own time and money to determine the copyright status of these objects. Despite being made available to view, the lack of rights information may prevent users from being able to make a “reasonable judgement” about the reuse of these works (Coyle 2005). Additionally, if an institution such as UNC-Chapel Hill was unsuccessful in tracking down rights information after spending hundreds of hours of labor and thousands of dollars, how can one user be expected to successfully find that information? In this instance, legal reuse for any case other than what may be deemed fair use may be impossible.

Institutions participating in DPLA are required to include rights information regardless of how successful they have been at determining actual copyright status of objects. The DPLA Metadata Application Profile (MAP) guides institutions in creating metadata that is compliant with DPLA systems and APIs. The MAP includes several “core classes” that are required for all objects included in DPLA (DPLA MAP Working Group 2017, 5). “Rights Holder” is a required field for cultural resources and references who holds the rights over the “actual” object as opposed to the digital representation. The field “Rights Statement” is also required for all objects and is visible to users. This informs users about the rights they have to use digital objects included in DPLA. However, while this information is required, there is currently no overarching mechanism for enforcing standardization or accuracy of included information. DPLA performs the mapping, the result of which creates “a surrogate that includes just enough data to be interoperable in the aggregated environment, leading the user back to the original source records in its full richness and depth” (Gueguen 2019, 120). Gueguen explains that “Because of the scale and variety of records harvested, DPLA’s in-house data quality activities are focused primarily on automated fixes for structural issues rather than semantic or conceptual issues” (121). There are simply too many records (over 22 million in 2019) for DPLA to be able to recommend changes at the unique record level. However, DPLA has initiated several community-driven approaches to increase metadata quality, one of which is a series of trainings and documents that encourage institutions to use the standardized rights statements available at RightsStatements.org in their metadata records (ibid).

RightsStatements.org provides a set of 12 rights statements that cultural heritage institutions can apply to digital objects in their collections. This project consists of five members – DPLA, Europeana, the National Heritage Digitization Strategy, National Digital Library of India, and the

National Library of Australia – and seeks to help cultural heritage institutions “[make] it easier for users to understand what they can do with a digital object they discover” (RightsStatements.org). Clarifying rights information for users was imperative, as by the time the group was formed, over 87,000 different rights statements were attached to resources included in DPLA (International Rights Statements Working Group 2016). The International Rights Statements Working Group stated that large-scale digital libraries such as DPLA were being “compromised by an inherent complexity of copyright laws combined with a lack of a shared vocabulary to describe legal rights that may affect the use of these items” (11). RightsStatements.org provides a shared vocabulary that institutions can implement in their metadata records to ensure clarity and uniformity of rights statements. In addition to the information provide on the website, other authors, such as Benson (2019), provide recommendations for when libraries could use certain statements.

Some DPLA hubs have made efforts to encourage the use of RightsStatements.org to enrich metadata records before being ingested by DPLA. The Illinois Digital Heritage Hub (IDHH), the Illinois state service hub for DPLA, audited rights statements included in metadata records by their contributing institutions (Stitzlein et al. 2018). Within these records, they found 938 unique values included in the <dc:rights> field. They compared these values to rights statements available at RightsStatements.org and identified seven potential statements equivalent to the information provided in the <dc:rights> field by contributing institutions (the authors note that this identification was based solely on the information provided, without a consideration of the accuracy of the institutionally-provided statements). After their audit, the IDHH created several strategies to assist contributing institutions with determining copyright status, inputting correct information into the <dc:rights> fields, and determining if there is a RightsStatements.org statement to match their copyright needs. These strategies appear to be working; the authors noted that, since undergoing their audit, more than 10% of harvested records now include standardized rights statements (Stitzlein et al. 2018).

Individual institutions are also using these standardized rights statements. Ballinger, Karl, and Chiu (2017) discuss the process librarians at Pennsylvania State Libraries used in early 2016 to prepare collections to be harvested by DPLA. They discovered that 27 out of 65 of their CONTENTdm collections lacked item-level rights information. Additionally, they found that only seven collections were “explicitly identified as being in the public domain,” although they discovered “several more collections” that fit that description (ibid, 146). Prior to assigning rights statements, Penn State Libraries considered “basic questions about the collection’s provenance and rights” to gain an understanding about item copyright and ownership (ibid, 153). Once the copyright status was established, the copyright officer selected an appropriate RightsStatements.org statement to apply to each item and included the URI for that statement in a metadata spreadsheet.

A project at the University of Miami Libraries hoped to “implement the RightsStatements.org recommendations, add item-specific copyright information to each object in the library’s digital collections, and develop workflows for including copyright information in all future digital projects” (Capell et al. 2018, 2). A main motivation for this was to “empower researchers to make better-informed decisions about potential uses of the library’s online resources” (ibid, 7). They identified six RightsStatements.org statements that accurately described materials in

their collections and included the URIs for these statements in their records. Additionally, they provided a Copyright Guidelines page that explains the ways in which “In Copyright” items could be used under fair use. Although they faced many issues when determining copyright for certain objects – for instance, when dealing with orphan works and works that lacked information essential for determining copyright (such as publication dates) – at least some rights information exists for the entirety of the University of Miami Libraries’ digital collections at the item level. Not only are these collections now in “closer alignment” with DPLA recommendations, but they also allow the University of Miami Libraries to “better serve [their] users by enhancing the accessibility of online resources” (Capell et al. 2018, 17).

However, not all institutions are conforming with DPLA recommendations. Gregory and Williams (2014) describe the workflows created by the North Carolina Digital Heritage Center (NCDHC) when serving as a DPLA hub for cultural institutions in North Carolina. When describing the process for checking metadata from contributing institutions, they state that, “Metadata quality beyond mapping accuracy and absence or presence of required fields was not rigorously reviewed” (Gregory et al. 2014). The NCDHC did not check metadata content; they only wanted to ensure that institutions provided metadata for all required fields. Gregory and Williams explain that the intention behind this was to “encourage participation by all types of institutions,” but that going forward, it is unclear how interested institutions utilizing the NCDHC as their hub would be in cleaning up their metadata, such as by providing more accurate rights information (2014). While encouraging participation is important, it is also important that institutions ensure they are presenting copyright in a manner that allows for reuse.

There is a body of evidence that shows how digitized materials are being reused. Users include objects in digital exhibits created by other institutions or feature objects in documentaries (Thompson et al. 2018); share them on social media sites (Kelly 2018; Reilly et al. 2017; Thompson et al. 2018); and publish objects as part of online tutorials or in open educational resources (OER) (Anderson et al. 2019; Terras 2015). Finally, some users may simply want to include them in academic papers and presentations (Matusiak et al. 2019).

As mentioned previously, many academic uses may fall under fair use. Four factors need to be considered when determining if something is fair use: the purpose of use (commercial vs non-commercial/educational); the nature of the work; the amount of the original work used; and the effect upon the market of the work under copyright. While this seems straightforward, most cases of fair use are judged on a case-by-case basis (Hirtle et al. 2009). However, if someone wants to reuse an object in a way that does not fall under fair use stipulations, they need to be aware of the copyright status of that object to ensure that they are not violating copyright and/or so they can obtain permission from the rights holder. Unfortunately, without a clear rights statement included with objects in digital libraries, users may be unsure of how they may reuse these objects. As per Terras (2015), “...how can users trust the content they are finding, and know that they will not be pursued for damages should they choose to reuse it?” (10). Some knowledge of copyright is essential in ensuring users understand how objects included in digital collections can be reused.

Not all educational uses are covered under fair use. The creation of open educational resources (OER) is one of the reasons digital collections should include a clear rights statement with

their digital objects. OER are educational materials that “reside in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions” (“Open educational resources (OER),” 2019). These materials are distributed freely and provide students in developing countries with up-to-date educational materials (ibid) as well as lessening the textbook cost burden in countries like the United States, where students may spend an average of \$1,240 per year on textbooks and supplies (“Average estimated undergraduate budgets, 2018-19,” 2019). However, since OER are meant to be redistributed, objects included in these materials need to be licensed, released with no licensing restrictions, or exist in the public domain. Interviews regarding OER use at Washington State University found that many faculty members had difficulty identifying materials for OER because of misunderstandings around statements such as “No Copyright” in DPLA (Anderson et al. 2019). One faculty member interpreted the phrase to mean that faculty were freely able to use these materials in course slides, one assumed “the item could only be linked to at its source,” and another “assumed unlimited use of the material” (ibid, 9). Faculty in this study noted that they “wanted to see clearer intellectual property designations,” and Anderson and Leachman (2019) encourage technical services and systems librarians to create improvements in metadata to meet this need (10).

Personal and commercial uses are also subject to copyright law. Researchers have found that users reuse images in digital collections for personal reasons, such as sharing on social media sites. Reilly and Thompson (2017) found that fifty percent of users of the Library of Congress’ “Teaching with Primary Resources” digital collection were using these resources for personal reasons. The most common type of reuse was posting objects to social media sites, with the “overwhelming majority of Social Media reuses” being on Pinterest, a social media site that allows users to “pin” images to virtual bulletin boards (Reilly et al. 2017, 62). Users also included objects on social media websites such as Tumblr and Flickr, although to a much smaller degree. Reilly and Thompson (2018) also found that “Popular Culture Research,” comprising primarily entries on personal blogs relating to the blog owner’s specific interests, was responsible for much object reuse (63). Some users also include objects from digital collections in commercial enterprises, such as selling physical prints (ibid). In all these instances, users would need to receive permission before reusing digital objects unless they were in the public domain or released under a license that allowed reuse. Copyright information included with a digital object’s metadata would allow users to determine whether or not they need to seek permission for reuse.

As we can see, murkiness around the copyright status of digital objects and lack of uniformity involved in the inclusion of copyright information for digital objects in large-scale digital libraries like DPLA can provide abundant opportunities for misinformation. This misinformation may have negative effects for users, such as confusion over the ability to reuse works and subsequent lack of use. By examining the copyright statements of objects included in DPLA by the Plains to Peaks Collective (PPC), this study seeks to understand how copyright is presented to users of these objects and makes recommendations regarding how institutions can accurately present copyright to users.

Methods

This study uses a descriptive research method and examines a sample of PPC records that has already been ingested into DPLA. It seeks to answer the following research questions: what copyright statements are included in PPC records that have been ingested by DPLA, and what copyright statements are included for public domain PPC items? As of April 3rd, 2019, the PPC has deposited 181,001 items into DPLA. Three institutions have contributed the bulk of these items – Institutions A, B, and C. These institutions have contributed metadata for 140,509 digital objects, which amounts to approximately 77.5% of the PPC collection, and as such are the institutions on which this study focuses. An additional consortium, while being listed as a major contributor, includes records from collections across the Rocky Mountain area, including Institutions B and C; as such, that consortium was excluded in order to prevent duplicate data.

Table 1

Objects Contributed by Institution

Institution Name	# Objects in DPLA	Relative % for Sample	# Objects Sampled
Institution A	93,043	66.2%	254
Institution B	37,530	26.7%	103
Institution C	9,936	7.1%	27
TOTAL	140,509	100%	384

Since the total collection has 181,001 objects, this study examined a sample size of 384 digitized or born-digital objects (380 digitized images, four born-digital PDF informational “bulletins”), which allows for 95% confidence with a 5% margin of error. The sample size reflects the entire PPC collection, but copyright statements were only taken from the top three contributing institutions. Ideally, follow-up studies would have a larger percentage of confidence with a smaller margin of error, but the required sample size was determined to be too large for the timeframe of this study. The sample size was split to reflect the relative percentage of contributions from each of the three targeted institutions. Institution A contributed 66.2% of objects (93,043) out of the top three institutions’ contribution of 140,509 objects; therefore, the copyright statements of 254 objects were examined for this study. Institution B contributed 26.8% (37,530) of objects for a total of 103 objects examined. Institution C contributed 7.1% (9,936) of objects for a total of 27 objects examined.

The collections of each institution were searched using three keywords – “fort,” “horse,” and “mountain.” These keywords were chosen by the author as words that would potentially be included in the descriptions for objects created in the Rocky Mountain region. Search results were sorted by relevance. Using DPLA’s list feature, lists were made of every other object until the required number of sample objects was met. Using these lists, basic metadata, including title, contributor, and a link to the object, was exported as a spreadsheet. Copyright information was then

added manually after looking at each object as displayed within DPLA and on the originating institution's website.

When reviewing the data, the study focused mainly on the copyright status of digital objects with creation dates that placed them firmly within the public domain. For works published in 1924 and after, the copyright status may not be easily traced, particularly because of the changes in copyright law that were previously mentioned in this paper (requirement to renew, possibility of non-renewal, etc.). However, works (excluding sound recordings) created before 1899 and with a known creator who died prior to 1950 (as of 2019) are within the public domain in the United States. Additionally, all results were examined for conflicting rights information. The 1899 date was thus chosen to include objects with ambiguous information regarding publication.

Results

Out of the 384 objects sampled in this study, 44 objects (~11.5%) included metadata that identified creators with death dates prior to 1950 and creation dates that placed them firmly within public domain years, but none of these objects had copyright statements that informed the user of public domain status.

Table 2

Public Domain Objects by Collection

Institution Name	# Objects Sampled	# Objects with public domain dates	# Objects with public domain rights information
Institution A	254	42	0
Institution B	103	0	0
Institution C	27	2	0
TOTAL	384	44	0

89 objects studied included conflicting rights information (23.2% of sample); when viewed on one platform, the user was presented with a certain statement, and when viewed on another, differing information was given. For these objects, the information in DPLA was that copyright status has not been evaluated. However, upon viewing these objects on the institutional websites, they were listed as being under copyright (usually by the contributing institution). Users are presented with conflicting information and as such cannot be sure how to legally reuse these objects or from whom to seek permission for reuse.

All 254 objects from Institution A included identical rights information. When these objects were viewed in DPLA, they provided "Copyright Not Evaluated" information as per RightsStatements.org. However, when viewed on Institution A's local site, all objects included the following statement: "Copyright restrictions applying to use or reproduction of this image available from [Institution A], at [email address]." Out of the 254 objects viewed, 42 of these objects

(approximately 16.5%) include a date or date range that falls completely within the date range of public domain and were made by a creator with a death date prior to 1950. However, none of these objects were marked as being in the public domain. These objects, which should be freely available for use without copyright restrictions, were marked as both “Copyright not evaluated,” and, essentially, as “rights available upon request.”

Table 3*Copyright Statements from Institution B*

Copyright Statement in Institution B Local Rights Metadata Field	# Objects with Statement
“The works contained in this collection are copyrighted by [Institution B] and other creators. To purchase copies of images and/or obtain permission to publish or exhibit them, contact [Institution B].”	81
(Local rights field not included)	14
“Institution B”	4
(URL to copyright statement saying Institution B retains copyright but also explaining Fair Use and providing contact information for commercial use inquiries.)	4
TOTAL	103

For the 103 objects examined from Institution B, all objects display the “Copyright Not Evaluated” language from RightsStatements.org when viewed in DPLA. When viewed within the Institution B repository, the objects include “Copyright Not Evaluated” language. However, 81 (78.6%) of these objects also include the following rights statement: “The works contained in this collection are copyrighted by [Institution B] and other creators. To purchase copies of images and/or obtain permission to publish or exhibit them, contact [Institution B].” An additional four objects include a “Copyright Not Evaluated” statement but also link to a page for the collection to which the digital objects belong. This page contains yet another copyright statement that says Institution B retains copyright to all images included in that specific collection. This page also includes a link to the institution’s stance on fair use and states that the images can be used for non-commercial, educational purposes so long as the files are not modified and a proper citation is used.

In contrast to the other two institutions, Institution C had no rights information included in DPLA. On their website, the only rights information on the page is a simple disclaimer stating that, if a copyright owner has work that is hosted on Institution C’s site and they believe that work has not been correctly attributed, or used without permission, the copyright holder should contact the institution. This statement is not included with the actual object metadata; it is included as a separate footer at the bottom of each page of the website. Similar to the other institutions, Institution C had items that were within public domain dates and known creators with death dates prior to 1950 but

were not marked as public domain. Institution C had two objects with metadata that indicated they were within the public domain, or about 7.4% of objects sampled, but because the sample size for this institution is so small, the study cannot extrapolate the same percentage to the collection as a whole.

Significance

If copyright statements in digital collections do not clearly indicate the accurate copyright status of a digital object (as seems to be the case in the PPC data set), then users do not have the information necessary to be able to reuse these objects. As per Sims, “providing only online access to collections is of limited value when visitors don’t know how they can make use of these materials” (Sims 2017, 79). Users who do not have essential copyright information for digital objects cannot trust that they will be able to use these objects without facing negative consequences (Terras 2015). Alternatively, if an object is displayed without any copyright statement, users may incorrectly assume objects in digital collections are in the public domain and may unknowingly commit copyright infringement. Coyle argues that libraries should “carry information that helps subsequent users make some assessment of the copyright status of the item” (2005). If an institution is unable to or does not have the resources to establish copyright status of an object themselves, they should at least provide as much information as possible in order to empower users to perform their own rights investigation. Instructing users to contact the institution for copyright information adds an unnecessary step and makes it harder for users to efficiently reuse digital objects.

Additionally, by claiming copyright over digital objects included in their collections without regard to the actual rights state of a work, institutions commit copyfraud. Copyfraud occurs when an entity falsely claims copyright over a public domain work (Mazzone 2006), such as when institutions use blanket statements in their collections or falsely point users to a sales department without regard for public domain status. Misconceptions about ownership may exist, and with the exception of some archival materials for which the copyright has been signed over to the institution by the rights holder, “for the vast majority of archival materials and for commercially and non-commercially published materials, libraries hold no intellectual property rights” (Coyle 2005).

Copyfraud prevents users from reusing items because it misrepresents the freedoms of reuse available to users. It “stifles valid forms of reproduction and creativity and undermines free speech” (Mazzone 2006, 1030). Even the use of Creative Commons licenses on public domain materials that allow users to reuse materials without paying a licensing fee but require attribution are still forms of copyfraud; while institutions may want credit for digitizing an item, under current copyright law, they do not have the right to require users to attribute their institutions (Sims 2017).

Additionally, in some instances, users are under the impression that they must ask permission for or pay to reuse objects that are in the public domain. In this study, users who wanted information about the copyright status of objects in Institution A’s digital collections, including objects with metadata indicating they are within the public domain, were directed to an email account with “sales” in the address. Such misrepresentations can result in “payment from individuals who do not know better or find it preferable not to risk a lawsuit” (Mazzone 2006, 1030). This misrepresentation may also prevent users from reusing objects because they do not want to or

cannot afford to pay for these digital resources. Some institutions claim that “fees for commercial use of public domain images are necessary to sustain its digitization efforts” (Schlosser 2009, 382). However, misrepresentation of copyright to users should not be an answer to funding issues, nor do funding issues give institutions the right to claim copyright over public domain objects, despite the cost of digitization. While institutions are free to charge for physical prints of public domain images, they should not claim ownership over the image itself.

Limitations

For this study, every other object of the returned search results was examined until enough copyright statements were collected to meet the sample size. Future iterations of the project will employ systematic sampling, but time constraints and technology issues caused the study to examine every other object for this specific trial. Technology issues included the breaking of the DPLA list feature after adding more than approximately 35 objects (the list feature should be able to contain 50 objects per list [Gibson 2018]) and the inability to export metadata regarding rights statements. As such, rights statements were collected manually. This process was further complicated because the author received server errors after accessing Institution A’s online collection too many times. The author had to switch between different browsers and computers in order to successfully collect all rights statements from sampled Institution A objects. Additionally, this study featured a sample size of 384 objects, which allowed for 95% confidence with 5% margin of error. Future iterations of this study should include a sample size big enough to allow for more confidence, with a smaller margin of error.

Discussion

One project that was created out of the need to address the problems of copyright in digital collections is RightsStatements.org. Both Institutions A and B have employed one of the statements from this project on their objects: “Copyright Not Evaluated.” However, for objects that are in the public domain, they could replace that statement with “No Copyright – United States.” Additionally, when developing workflows for the creation of metadata for inclusion in large-scale digital libraries like DPLA, cultural institutions should create steps to incorporate including an accurate copyright statement (Benson 2019). By doing so, users will not only be able to view these objects, but also have the information necessary to empower them to reuse these objects in legal ways. Although the copyright status of an object can be complicated or unclear, especially for works published after 1923 (as of 2019), anything published during this year and earlier could be considered low-hanging fruit. By asking themselves a simple question – was this work published in or before 1923? – institutions could add public domain information to all works within their collection that are clearly within public domain and as such free for reuse (although institutions should ensure they update the cut-off every year). They could follow a similar process for unpublished works created in 1899 (as of 2019) and before that have a known creator with a death date prior to 1950 (once again, as of 2019).

Institutions should also make plans to revisit metadata that has already been created and included in their digital collections. Such a task may seem daunting; after all, libraries already face

restrictions of time and funding, and many institutions may not feel that resources should be spent on correcting past mistakes. However, if institutions start by focusing on objects at the collection level – as Capell and Williams did when reviewing large groups of materials by the same creator (2018) – or objects with public domain dates, they could create a simple workflow that could be completed relatively quickly. There are also tools currently being developed that could be used to audit and fix rights information in metadata. The Mountain West Digital Library created a metadata auditing tool which was adapted from a tool created by the North Carolina Digital Heritage Center and allows for “student workers and non-professionals to perform large-scale metadata auditing even if they have no prior knowledge of application profiles or metadata auditing workflows” (Hebron 2018). Using the DC Facet View feature of this tool, institutions could review date metadata from large amounts of objects at one time and easily identify which objects have dates that fall within the public domain range and correct rights information for those objects. Since this tool was designed for beginners, it could easily be adapted into workflows and/or tutorials for student worker or paid intern use.

Another option that institutions could employ is the creation of clear reuse sets. The Library of Congress has a section of their digital collections titled “Free to Use and Reuse Sets.” These are sets for which the “...Library believes that this content is either in the public domain, has no known copyright, or has been cleared by the copyright owner for public use” (Library of Congress). These collections are grouped by subjects like “Maps of Cities,” “Historical Travel Pictures,” and “African-American Women Changemakers.” By having these free-to-use resources clearly collocated and available to users, users can be confident that these resources are eligible for reuse without negative repercussions.

Conclusion

Libraries share things online “...to fulfill their public interest missions of access, preservation, research, and education” (Hirtle et al. 2009, 14). The purpose of digital collections is to fulfill the research needs of users (Borgman 1999), meet the needs of communities (Calhoun 2014, 18), and “ideally...allow new types of use when digitised” (Jordan 2006, 27). If digital collections do not accurately state the rights users have to works included in these collections, then those works cannot be reused to contribute to new scholarship. This study has observed that three organizations in the Rocky Mountain area utilize blanket copyright statements and conflicting rights information when including objects in digital collections. The blanket copyright statements mislabel public domain works as being under copyright and/or contradict other copyright information included in the digital object’s metadata. By not providing accurate copyright statements, the objects in these collections are rendered legally useless. In order to empower users to reuse these materials, PPC institutions should provide clear and accurate copyright statements with all objects uploaded in digital collections. This study hopes to inspire institutions to undertake revisions of copyright statements in order to clarify user rights and increase reuse of their works.

Bibliography

- “About rightsstatements.org.” RightsStatements.org. Accessed April 17, 2019.
<https://rightsstatements.org/en/about.html>
- “About us.” Digital Public Library of America. Accessed April 17, 2019. <https://dp.la/about>.
- “Average estimated undergraduate budgets, 2018-19.” The College Board. 2019.
<https://trends.collegeboard.org/college-pricing/figures-tables/average-estimated-undergraduate-budgets-2018-19>.
- “Free to use and reuse sets.” Library of Congress. Accessed April 17, 2019.
<https://www.loc.gov/free-to-use/>.
- “Frequently asked questions.” Digital Public Library of America. Accessed April 17, 2019.
<https://dp.la/about>.
- “History.” Digital Public Library of America. Accessed April 17, 2019. <https://pro.dp.la/about-dpla-pro/history>.
- “How to use this site.” UNC-Chapel Hill Libraries. May 18, 2019.
<https://docsouth.unc.edu/watson/help.php>.
- “Open educational resources (OER).” UNESCO. 2019. <https://en.unesco.org/themes/building-knowledge-societies/oeer>.
- “Prospective hubs.” Digital Public Library of America. Accessed April 17, 2019.
<https://pro.dp.la/prospective-hubs>.
- “Welcome to Europeana collections.” Europeana. Accessed May 5, 2019.
<https://www.europeana.eu/portal/en/about.html>
- Anderson, Talea, and Chelsea Leachman. “Strategies for Supporting OER Adoption through Faculty and Instructor Use of a Federated Search Tool.” *Journal of Librarianship and Scholarly Communication* 7, no. 1 (February 22, 2019). <https://doi.org/10.7710/2162-3309.2279>.
- Ballinger, Linda, Brandy Karl, and Anastasia Chiu. “Providing Quality Rights Metadata for Digital Collections Through RightsStatements.Org.” *Pennsylvania Libraries: Research & Practice* 5, no. 2 (October 31, 2017): 144-58. <https://doi.org/10.5195/PALRAP.2017.157>.
- Benson, Sara R. “Copyright Conundrums: Rights Issues in the Digitization of Library Collections.” In *Digital Preservation in Libraries: Preparing for a Sustainable Future*, 305-320. Edited by Jeremy Myntti and Jessalyn Zoom. Chicago, IL: American Library Association, 2019.
<http://hdl.handle.net/2142/102161>.
- Borgman, Christine L. “What are digital libraries? Competing visions.” *Information Processing & Management* 38, no. 3 (May 1999): 227-243. [https://doi.org/10.1016/S0306-4573\(98\)00059-4](https://doi.org/10.1016/S0306-4573(98)00059-4).
- Calhoun, Karen. *Exploring digital libraries: Foundations, practice, prospects*. Chicago: Neal-Schuman, 2014.
- Capell, Laura, and Elliot Williams. “Implementing RightsStatements.org at the University of Miami Libraries.” *Journal of Librarianship and Scholarly Communication* 6, no. 1 (2018).
<https://doi.org/10.7710/2162-3309.2254>.
- Coyle, Karen. “Descriptive metadata for copyright status.” *First Monday* 10, no. 10 (2005).
<https://doi.org/10.5210/fm.v10i10.1282>.
- Dickson, Maggie. “Due diligence, futile effort: Copyright and the digitization of the Thomas E. Watson papers.” *The American Archivist* 73, no. 2 (2010): 626-636.
<https://www.jstor.org/stable/23290761>.
- DPLA MAP Working Group. “Metadata Application Profile (version 5.0).” December 7, 2017.
<http://dp.la/info/map>.
- Gibson, Samantha. “Save your favorite DPLA items with new list feature.” Digital Public Library of America (blog). August 1, 2018. <https://dp.la/news/save-your-favorite-dpla-items-with-new-list-feature>.

- Gregory, Lisa, and Stephanie Williams. "On being a hub: Some details behind providing metadata for the Digital Public Library of America." *D-Lib Magazine* 20, no. 7/8 (2014): 25-32. doi:10.1045/july2014-gregory.
- Gueguen, Gretchen. "Metadata quality at scale: Metadata quality control at the Digital Public Library of America." *Journal of Digital Media Management* 7, no. 2 (2019): 115-126.
- Hebron, Teresa K. "Extending and adapting metadata audit tools for Mountain West Digital Library members." *Code4Lib Journal* 41 (August 9, 2018). <https://journal.code4lib.org/articles/13632>.
- Hirtle, Peter B., Emily Hudson, and Andrew T. Kenyon. *Copyright and cultural institutions: Guidelines for digitization for U.S. libraries, archives, and museums*. Ithaca, NY: Cornell University Library, 2009. <https://hdl.handle.net/1813/14142>.
- International Rights Statements Working Group. *Rightsstatements.org white paper: Recommendations for standardized international rights statements*. 2015 (updated January 2016). rightsstatements.org/files/160208recommendations_for_standardized_international_rights_statements_v1.1.pdf.
- Jordan, Mark. *Putting content online: a practical guide for libraries*. Oxford: Elsevier, 2006.
- Kelly, Elizabeth Joan. "Content analysis of Google alerts for cultural heritage institutions." *Journal of Web Librarianship* 12 no. 1 (2018): 28-45. <https://doi.org/10.1080/19322909.2017.1369374>.
- Matusiak, Krystyna K. "User navigation in large-scale distributed digital libraries: The case of the digital public library of America." *Journal of Web Librarianship* 11, no. 3-4 (2017): 157-171. <https://doi.org/10.1080/19322909.2017.1356257>.
- Matusiak, Krystyna K., Anna Harper, and Chelsea Heinbach. "Use and Reuse of Visual Resources in Student Papers and Presentations." *The Electronic Library* 37, no. 3 (2019). <https://doi.org/10.1108/EL-09-2018-0185>.
- Mazzone, Jason. "Copyfraud." *New York University Law Review* 81 (2006): 1026-1100. <https://heinonline.org/HOL/P?h=hein.journals/nylr81&ci=1042>.
- Monson, Jane D. *Getting started with digital collections: Scaling to fit your organization*. Chicago: American Library Association, 2017.
- Purday, Jonathan. "Europeana: Digital access to Europe's cultural heritage." *Alexandria* 23, no. 2 (2012): 1-13. doi:10.7227/ALX.23.2.2.
- Reilly, Michele, and Santi Thompson. "Reverse image lookup: Assessing digital library users and reuses." *Journal of Web Librarianship* 11, no. 1 (2017): 56-68. <https://doi.org/10.1080/19322909.2016.1223573>.
- Schlosser, Melanie. "Unless otherwise indicated: A survey of copyright statements on digital library collections." *College & Research Libraries* 70, no. 4 (2009): 371-385.
- Sims, Nancy. "Rights, ethics, accuracy, and open licenses in online collections: What's 'ours' isn't really ours." *College & Research Libraries News* 78, no. 2 (2017): 79-82. <https://doi.org/10.5860/crln.78.2.9620>.
- Stitzlein, Hannah, Myung-Ja K. Han, and Sara R. Benson. "Unraveling challenges: Rights statements in digital cultural heritage collections." *Journal of Library Metadata* 18, no. 3-4 (2018): 135-150. <https://doi-org.du.idm.oclc.org/10.1080/19386389.2018.1540328>.
- Terras, Melissa. "Opening access to collections: The making and using of open digitised cultural content." *Online Information Review* 39, no. 5 (2015): 733-752. <https://doi.org/10.1108/OIR-06-2015-0193>.
- Thompson, Santi, Genya O'Gara, Elizabeth Joan Kelly, Ayla Stein Kenfield, Caroline Muglia, and Liz Woolcott. "Assessing for digital library reuse: Initial findings from the Measuring Reuse Project." Presentation at the American Library Association (ALA) Annual Conference, New Orleans, June 25, 2018. <https://doi.org/10.17605/osf.io/m46bc>.

Xie, Iris, and Krystyna K. Matusiak. *Discover digital libraries: Theory and practice*. Cambridge, MA: Elsevier, 2016.