



# Journal of Copyright in Education and Librarianship

ISSN 2473-8336 | jcel-pub.org

Volume 3, Issue 3

## **Digital Cultural Heritage and Wikimedia Commons Licenses: Copyright or Copywrong?**

Elizabeth Joan Kelly

Kelly, E. J. (2019). Digital Cultural Heritage and Wikimedia Commons Licenses: Copyright or Copywrong? *Journal of Copyright in Education and Librarianship*, 3(3), 1-25. <https://doi.org/10.17161/jcel.v3i3.9771>



© 2019 Elizabeth Joan Kelly. This open access article is distributed under a Creative Commons Attribution-NonCommercial 4.0 International license: <https://creativecommons.org/licenses/by-nc/4.0/>.

## Digital Cultural Heritage and Wikimedia Commons Licenses: Copyright or Copywrong?

Elizabeth Joan Kelly  
Loyola University New Orleans

Correspondence regarding this article should be addressed to Elizabeth Joan Kelly, digital programs coordinator, Loyola University New Orleans, [ejkelly@loyno.edu](mailto:ejkelly@loyno.edu).

### Abstract

Cultural heritage institutions can contribute to public knowledge and increase awareness of their collections by uploading digital objects to Wikimedia Commons for use on Wikipedia and other Wikimedia Foundation projects. However, prior research has established the difficulty of and/or hesitation by many cultural heritage institutions in clearly and accurately labeling the copyright status of their born-digital and digitized collections. With this knowledge, how likely is it that digital cultural heritage will be findable and usable on Wikimedia Commons? This study seeks to determine how accurate rights statements for cultural heritage objects on Wikimedia Commons are, and whether inaccuracies can be linked to problematic rights statements in cultural heritage digital libraries or whether the inaccuracies stem from Wikimedia Commons. By evaluating the rights statements, licenses, and sources for 308 Wikimedia Commons objects from 57 cultural heritage organizations and comparing that information to corresponding licenses from digital libraries, we can begin to develop best practices and educational needs for digital librarians, archives, museum curators, and Wikipedians alike to improve the user experience for those using digital cultural heritage on Wikimedia projects.

*Keywords:* Digital library, copyright, Wikipedia, rights statements, licenses

## **Digital Cultural Heritage and Wikimedia Commons Licenses: Copyright or Copywrong?**

### **Introduction**

Many cultural heritage institutions are working to increase awareness of their collections and contribute to public knowledge by participating in nonprofit projects like those created by the Wikimedia Foundation. Uploading digital surrogates of cultural heritage objects to Wikimedia Commons, the media repository for the Wikimedia Foundation's projects, allows users to access previously difficult to find or even hidden collections and reuse them in ways that boost their visibility and contribute to the credibility of Wikipedia. However, cultural heritage institutions struggle greatly with an integral part of making collections available online—copyright. The complexities of copyright law for organizations that house collections they may not own the rights to, or may—whether ill-intentioned or not—wish to restrict access to, have led to a number of reports of incorrect rights statements in digital libraries. If even trained library, museum, and archives professionals grapple with licensing and determining rights statements, what are the chances that licenses for digital cultural heritage on Wikipedia and other Wikimedia Foundation projects are correct?

This study seeks to determine whether rights statements for cultural heritage objects on Wikimedia Commons are applied correctly (that is, copyrighted, public domain, and open-licensed objects are accurately labeled) and whether inaccuracies can be linked to problematic rights statements in cultural heritage digital libraries or whether the inaccuracies stem from Wikimedia Commons. This research will aid cultural heritage organizations that are hoping to increase their visibility and improve public access to digital cultural heritage in determining where to focus their training efforts when it comes to working with Wikimedia Commons. This research may also be of use to Wikipedians and developers who are actively attempting to improve the infrastructure of Wikimedia Commons, particularly the way rights statements and licenses are inputted and coded, to improve accuracy and increase reuse across Wikimedia projects and beyond. By evaluating the rights statements, licenses, and sources for 308 Wikimedia Commons objects from 57 cultural heritage organizations and comparing that information to corresponding licenses on digital libraries, we can begin to develop best practices and educational needs for digital librarians, archives, museum curators, and Wikipedians alike to improve the user experience for those using digital cultural heritage on Wikimedia projects.

## Overview of Wikimedia Commons

Wikimedia Commons ([commons.wikimedia.org](https://commons.wikimedia.org), hereafter referred to as WMC) serves as the media repository for photographs, artwork, video, sound, diagrams, and other files that can then be reused in Wikimedia projects such as Wikipedia, Wikiquote, and Wikinews as well as outside of the Wikimedia landscape. Like other Wikimedia Foundation projects, anyone can sign up for an account and contribute media to WMC. Files can be uploaded individually as well as into themed galleries, and the following controlled vocabulary categories are assigned to enable easy browsing of media by the following facets (“Commons:Categories,” n.d.):

- Topic (including but not limited to timespan, location, creator, subject)
- Copyright status
- Original source (book, collection, website)
- Media type
- User categories

Tools such as “Wiki Shoot Me!” encourage Wikipedia editors, often referred to as Wikipedians, to enrich articles about nearby locales by taking photographs and adding them to the repository for use in under-illustrated articles, and the Picture Requests page lists articles that need images by category.

Structured data on Wikipedia projects is licensed as public domain; all unstructured text is licensed under the Creative Commons Attribution-ShareAlike License (CC BY-SA 3.0), meaning the entirety of the repository can be shared and adapted so long as appropriate credit is given; any subsequent material must also be given a CC BY-SA 3.0 license if it is remixed or builds upon the original material (Creative Commons, n.d.). Individual media uploaded to WMC must also conform to specific rights criteria in order to meet WMC’s mission to “empower and engage people around the world to collect and develop educational content under a free content license” (“Resolution:Licensing policy,” n.d.). The media must fall into two categories: licensed explicitly using a free license such as Creative Commons, GNU, Free Art, and others, or be in the public domain, at least in the United States and the country of origin (“Commons:Licensing,” n.d.).<sup>1</sup> Licenses that restrict the reuse of material to noncommercial purposes are not allowed, nor are copyrighted materials uploaded with claims of fair use, a doctrine of US copyright law that allows copyrighted materials to be used without permission from the copyright holder under certain conditions. In some cases, however, copyrighted media may be uploaded with permission from the copyright holder (“Commons:Email templates,” n.d.).

Beyond copyright law, which defines ownership of intellectual property and provides the legal parameters under which copyrighted material may be used, some

---

<sup>1</sup> See <https://freedomdefined.org/Licenses> for a full list of available licenses.

materials on WMC may be subject to noncopyright restrictions such as trademark law, personality rights, database rights, and authors' moral rights, to name a few ("Commons:Non-copyright restrictions," n.d.). WMC provides tutorials and reference tools to Wikipedians to help them choose an appropriate license when uploading content—these tools include the adoption of famed archivist and copyright expert Peter Hirtle's copyright and public domain determination chart ("Commons:First steps/License selection," n.d.; "Commons: Hirtle chart," n.d.). Finally, in addition to declaring the license of the uploaded media, Wikipedians insert copyright or public domain tags using wiki markup, enabling easy categorization of WMC media by license ("Commons:Copyright tags," n.d.).

### **Literature Review**

The article "Use of Louisiana's Digital Cultural Heritage by Wikipedians" detailed the ways in which Wikipedia is increasingly used as a research tool by students and professional researchers (Kelly, 2018b). More recent scholarship by librarians posited Wikipedia as an important resource for academic information-literacy instruction by illustrating ways in which using Wikipedia as a pedagogical tool aligns with the Association of College and Research Libraries' Framework of Information Literacy for Higher Education (Dowell & Bridges, 2019). And while much previous research focused on attitudes and perceptions of Wikipedia's reliability (Snyder, 2013), an increasing amount of research is being conducted as to whether Wikipedia is actually a reliable research tool—and evidence is emerging that it is (Bauder & Emanuel, 2012; Chesney, 2006).

Archives and digital libraries can enrich Wikipedia by contributing their digital collections as open research. To meet the needs of a large user population relying on Google as the world's most popular search engine, cultural heritage institutions can situate themselves well by embracing this trend and embedding their resources where users are already looking. As detailed by Kelly (2018b), many cultural heritage institutions are already implementing initiatives to enhance Wikipedia articles and drive traffic back to institution websites and digital libraries. This is done by linking to archival finding aids on Wikipedia articles and uploading digital cultural heritage to WMC. Librarians, archivists, museum curators, Wikipedians-in-residence, and student workers who have engaged in this work have seen large increases in web traffic to their websites and digital libraries and, in some cases, increases in email and in-person consultations regarding their collections (Brinkerink, 2015; "Collections Are for Use," 2009; Cooban, 2017; Elder, Westbrook, & Reilly, 2012; Ferriero, 2011; Galloway & DellaCorte, 2014; Lally & Dunford, 2007; Perrin, Winkler, Daniel, Barba, & Yang, 2017; Szajewski, 2013; Vetter & Harrington, 2013).

A study of Wikipedians-in-residence and archivists on their efforts to promote collections through Wikipedia found that “although adding links back to their catalogues can be worthwhile for both archivists and Wikipedia, editing articles, uploading content to Wikimedia Commons or investigating Wikidata may be more valuable still” (Cooban, 2017, p. 267). Uploading images to WMC is a particularly fruitful strategy because of the relative success of Wikipedia articles with images over those without. Wikipedia pages with public domain images on them saw between 17% and 18% more traffic than articles without images, and attempts to apply a monetary value to the tens of millions of freely licensed media on WMC put the value of these materials at over \$200 million (Erickson, Heald, Homberg, Kretschmer, & Mendis, 2015).<sup>2</sup>

However, cultural heritage institutions should not just contribute to Wikimedia Foundation projects for marketing purposes. As David Ferriero, archivist of the United States, put it, “Our work with Wikipedia is not only good enough, it’s great for us because it takes our goals of transparency, public participation, and collaboration to a new level” (2011, p. 368). Initiatives like OpenGlam provide resources and workshops to aid cultural heritage institutions in publishing their collections and data to Wikimedia Foundation projects and other venues. The advantages of doing so include greater public awareness of collections, increased discoverability, opportunities for public participation in curation and enrichment, increased use by teachers in the classroom, better understanding of cultural heritage by communities, and use of cultural heritage materials in new creations (Terras, 2015a).

Unfortunately, some of the biggest barriers to cultural heritage institutions making their collections openly available are rooted in the intricacies, vagaries, and inconsistent (or incorrect) application of copyright to cultural heritage materials. Numerous research projects have found evidence that “despite their mission to make their holdings available, archives and libraries are invoking copyright in ways that may impede or discourage access to, and use of, online documentary heritage” (Dryden, 2012). This can range from copyfraud, the practice of falsely asserting copyright (whether intentional or unintentional), to intentionally using vague or unclear statements out of liability concerns about accidentally using the wrong statement, confusing assertions of rights over digital surrogates despite statements that physical objects are out of copyright, and conflating licenses with terms of use or other matters outside of copyright law (Ballinger, Karl, & Chiu, 2017; Dryden, 2011; Schlosser, 2009; Sims, 2017; Stitzlein, Han, & Benson, 2018).

While institutions may be able to assert copyright for materials they own, doing so can affect reuse of these materials, as found in a study comparing

---

<sup>2</sup> See <https://commons.wikimedia.org/wiki/Special:Statistics> for up-to-date statistics on the number of items in WMC.

readership and reuse of both public domain and copyrighted issues from *Baseball Digest* (Nagaraj, 2018). This is compounded when copyright statements are erroneously applied to materials that are in fact in the public domain, or when rights statements are phrased ambiguously. Melissa Terras evaluated the adoption of cultural heritage objects in maker culture—such as fabric patterns on the custom design website Spoonflower or the use of images on mugs, T-shirts, or other objects on the e-commerce website Etsy—and found several leading causes as to why such adoption has been limited. Poor browsing interfaces in cultural heritage digital collections, the lack of available open-licensed images from the 20th century, confusing and restrictive licensing statements, and low-quality images (often used by cultural heritage institutions as another means of restricting reuse) were all determinants in limiting reuse (Terras, 2015b). A 2008 survey of library professionals on their motivations for trying to control the noncommercial use of cultural heritage objects identified three general themes: controlling descriptions and representations; legal risks and complexities as well as getting credit; and fiscal and social costs and revenue (Eschenfelder & Caswell, 2010).

Focusing on the museum sector in particular, *Bridgeman Art Library v. Corel Corp* resulted in a landmark copyright-law decision that created the precedent that digital reproductions of two-dimensional works—such as photographs of paintings in a museum—do not merit their own copyright. This means that digital reproductions of public domain two-dimensional works are also in the public domain. While no other district court has since ruled otherwise, variances occur in how this same issue is dealt with in non-US jurisdictions (Ballinger, Karl, & Chiu, 2017). The International Rights Statement Group “strongly encourag[es] those contributing organizations that create scans to adopt the position that the scan should not create a new layer of copyright as a matter of policy and choice by participating institutions, and to the extent that rights are created, those rights are waived” (Ballinger, Karl, & Chiu, 2017, p. 147). Still, cultural heritage institutions seem to struggle with, or intentionally ignore, this ruling. Some museums protested the *Bridgeman* decision because the advent of consumer-grade, high-quality cameras means that anyone can take photos of museum objects, which affects gift shop sales of postcards and other reproductions. These same museums may instead try to enforce physical control over collections by not allowing photography of them or by attempting to enforce copyright when they should not (Needham, 2017).

Cultural heritage institutions may also sometimes misuse Creative Commons licenses. Lawyer and librarian Nancy Sims provided a clear overview of some of the most commonly misused copyright and Creative Commons licensing applied by cultural heritage institutions when she wrote, “it is legally inappropriate for people or organizations other than a rightsholder to try to apply an open license to a work. They are intended to be used by rightsholders to let others know that a work is

preapproved for use under certain conditions. While we (libraries) may have the right to digitize works in our collections and share them online by statute or permission, we far less frequently have the right to authorize uses by others” (2017, pp. 79–80).

Another common misuse of open licenses occurs when Creative Commons licenses are applied to public domain works, whether out of a misguided intent to increase sharing or a more restrictive goal of trying to enforce attribution. Unlike the *Bridgeman* ruling on two-dimensional works, two-dimensional reproductions of three-dimensional works (such as a photograph of a sculpture) can be copyrighted; therefore, Creative Commons licenses can be used for photographic reproductions of public domain three-dimensional objects. However, Sims cautioned that a waiver, such as the CC0 “No Rights Reserved” public domain waiver, might be more appropriate “to convey that we do not believe we have rights in the item, or that we are aware that laws vary across jurisdictions about whether there may be rights in the item, and we don’t want to claim them even in the jurisdictions where they exist” (p. 80). Cultural heritage institutions may use Creative Commons licenses for collections where the institution has been granted copyright, but they may still run into issues if donation and licensing language is unclear in donation agreements or if only one rightsholder is represented in the agreement despite materials in the collection being created by others (for example, if the collection contains photographs of the collection creator that are copyrighted by the photographer).

Attempts to simplify licensing statements for cultural heritage materials on a large scale have been recently undertaken by [rightsstatements.org](http://rightsstatements.org), an international collaborative project to develop standardized, interoperable rights statements and reuse information. The [rightsstatements.org](http://rightsstatements.org) consortium is currently comprised of the Digital Public Library of America, Europeana, the National Digital Library of India, the National Heritage Digitization Strategy, and Trove. The 12 rights statements use linked data to make them both human and machine readable, and they articulate rights under three main categories: “In Copyright,” “No Copyright,” or “Other.” The statements are meant to be used by institutions and aggregators when Creative Commons licensing is not applicable. Creative Commons licenses are meant to be used by creators, and cultural heritage institutions are often stewards, not creators, of open collections. [Rightsstatements.org](http://Rightsstatements.org) provides a means for clearly articulated rights and reuse information for noncreators, thus satisfying Sims’s concern about the application of Creative Commons licensing by someone other than the rightsholder.

As cultural heritage institutions begin to adopt these standardized rights statements, many have identified problems in their legacy collections’ licensing that need reconciling. Example best practices and roadmaps for converting to [rightsstatements.org](http://rightsstatements.org) have been published or presented by librarians at the

University of Miami (Capell & Williams, 2016), the Illinois Digital Heritage Hub (Stitzlein, Han, & Benson, 2018), and Pennsylvania State University (Ballinger, Karl, & Chiu, 2017), to name a few.

Cultural heritage institutions have long wrestled with issues of copyright, but the advent of the internet has exacerbated these issues. Jean Dryden wrote in connection to Canadian archivists, “responding to requests for copies in the analog world was a mediated process that included an opportunity for communication between the archivist and the user that would allow the archivist to differentiate between copyright matters and other interests. But the user copying from the website sees only what is online, and if the guidance on reproduction and use is linked to copyright, it is difficult to escape the conclusion that any restrictions arise out of copyright” (2011, p. 541). Peter Hirtle put it similarly: “Technology has not changed the law, but it has affected how people interpret the law—with potentially disastrous results as far as unpublished material is concerned” (2001, p. 2).

### **Research Questions**

Cultural heritage institutions can contribute to public knowledge and increase awareness of their collections by uploading digital objects to WMC for use on Wikipedia and other Wikimedia Foundation projects. Doing so may also increase perceptions of Wikipedia’s credibility as a research source. But considering the difficulty and/or hesitation of many cultural heritage institutions to clearly and accurately label the copyright status of their born-digital and digitized collections, how likely is it that digital cultural heritage will be findable and usable on WMC?

In light of the issues uncovered in the literature review, this study seeks to answer the following questions:

- Can WMC be used as a reliable source to find accurately labeled public domain and open-licensed media?
- What are the most common inaccuracies found in WMC cultural heritage object rights statements?
- When digital cultural heritage institution media is uploaded to WMC, can the rights statements found in digital libraries be reconciled with their rights statements on WMC?
- When rights statements for digital libraries and for corresponding uploads in WMC cannot be reconciled, which rights statement is correct?

### **Methodology**

Cultural heritage institutions were selected for inclusion in this study using a list of repositories from a previous research project; this list was randomly generated from the archival discovery tool ArchiveGrid (Kelly, 2018a). The purpose of using this list of repositories was to include institutions of various types, sizes,

and geographical locations. Sixty-six institutions, ranging from museums to historical associations to university archives, were searched in WMC (Appendix A). Images of an institution's building or grounds were not analyzed; only media that appeared to be from the institution's collections, determined primarily via the "source" field in the WMC metadata, were included. These images did include user-generated photos of collections, such as images of a painting or sculpture in a museum taken by a Wikipedian.

The purpose of this study was not to determine how many cultural heritage objects are in WMC or how many objects an individual cultural heritage institution has in WMC—not all entries from the repositories investigated were analyzed. In some cases a cultural heritage institution had hundreds of objects from their collections in WMC, but only a small sample of these (no more than 20) were analyzed for each institution.

A total of 308 WMC entries were analyzed, and data was collected and analyzed via spreadsheet. For each entry the following information was entered:

- Repository name
- Search term used to locate result
- WMC object URL
- Source details
- WMC object creation year
- WMC license (abbreviated)
- WMC use permissions
- WMC license (full text)
- Accuracy of WMC license
- Original source URL (when applicable)
- Original source license (full) (when applicable)
- Notes
- License compatibility (when applicable)

Open coding was used to code free-text data points such as source details. Open coding is a procedure from the grounded theory research method in which data is examined and then grouped into thematic categories without any predetermined codes (Strauss & Corbin, 1998). The "source" field in WMC is optional and no-standard, so each had to be examined closely to determine from where the uploaded WMC object might have come. The codes that developed from this exercise were *blog*, *book scan*, *digital exhibit*, *digital library object*, *Find A Grave*, *history website*, *institution website*, *photographer website*, *promotional or public relations article*, *research guide*, *uploaded directly to WMC by institution*, *user photo*, and *unknown*.

When possible, WMC object creation year was determined from the date listed on the WMC record page. However, some of these dates referred to the date the digital surrogate was taken while others referred to the creation date of the original object. In the case of photographs of three-dimensional objects, the date of the digital surrogate would indeed be eligible for its own copyright date, so these dates were inputted on the spreadsheet. Since digital reproductions of two-dimensional objects most likely would not be eligible to be copyrighted, the creation date of the original object was instead analyzed, when available. Accuracy of the WMC license was primarily decided based on the date (to determine whether the object was in the public domain) and related information (such as the creator, to determine if an object was created by a federal government employee as part of their duties and was therefore in the public domain) found in the WMC record. In some cases enlightening information was found once the original source was investigated, but in large part, creation dates and creator information on the WMC page were assumed to be correct.

If a URL was provided for the original source—such as in the case of images downloaded from cultural heritage digital libraries, personal or institutional blogs, institution websites, and digital exhibits—this source was also evaluated to see what copyright information was made available there and whether that information corresponded to the information on the WMC page. If the copyright statements did not reconcile, a determination was made as to which (if any) of the statements were correct in the “license compatibility” data field.

After the data was entered, coded, and cleaned, pivot tables and charts were created to analyze and visualize the results.

## **Results**

Of the 69 cultural heritage institutions that were searched in WMC, 57 had at least one collection object surrogate uploaded to WMC. Ultimately, 308 WMC objects from 57 cultural heritage institutions were examined for an average of five WMC objects per institution. Appendix B lists the number of results analyzed for each institution.

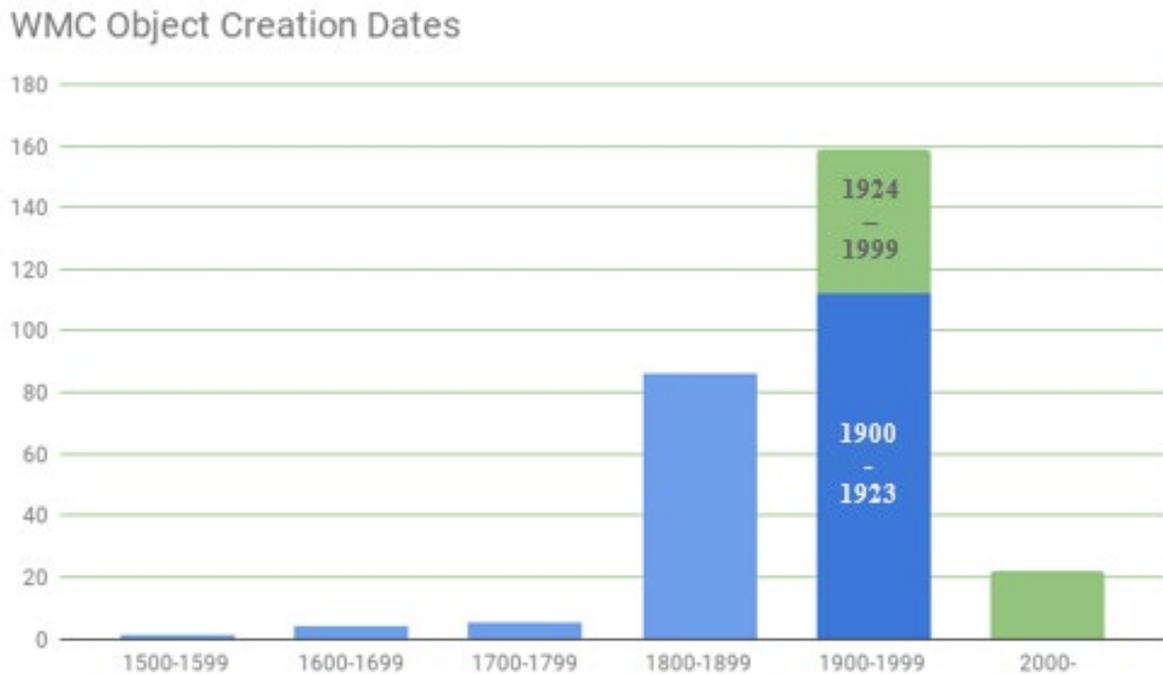
It was not possible to determine the source for 103 of the analyzed results, either because no source was given or because the link to the source no longer worked. Of the remaining 205 WMC results, 60% were the outcome of cultural heritage digitization and were from digital libraries. These included objects from CONTENTdm or other common digital library programs as well as mass digitization projects that were uploaded to the Internet Archive and then to Flickr Commons (Miller, 2014). Media uploaded directly to Flickr were checked as to whether the Flickr account was that of a cultural heritage institution or that of an individual. Cultural heritage Flickr account media were coded as digital library objects, while

individual Flickr account media were coded as user photos. These user photos, either taken from Flickr or uploaded directly to WMC, accounted for the second-most-common source of digital cultural heritage examined on WMC (25%). Digital exhibits (5%) and blogs (3%) contributed a small number of objects to WMC, and the remainder categories accounted for only a few results each (see Table 1).

Source Details (n = 205)	Number	% of Total
Digital library object	123	60
User photo	52	25
Digital exhibit	11	5
Blog	6	3
Uploaded by institution	3	1
Book scan	2	1
History website	2	1
Research guide	2	1
Find A Grave	1	0.5
Institution website	1	0.5
Photographer website	1	0.5
Promotional or public relations article	1	0.5

*Table 1.* WMC object source details. Broad categories for the original source of analyzed digital cultural heritage objects on Wikimedia Commons.

Thirty-one WMC objects did not have an available creation date, either on WMC or the original source (when available). Perhaps unsurprisingly, the majority of results with obtainable dates were from years prior to 1924, as copyright in the United States has expired for works published through 1923. These accounted for 208, or 68%, of all WMC objects analyzed, or 75% of WMC objects with creation dates. A chart of all dates inputted by century for the WMC objects analyzed appears in Figure 1.



*Figure 1.* WMC object creation dates. Dates given for the creation or capture of analyzed cultural heritage objects uploaded to Wikimedia Commons.

More than three quarters (77%) of the WMC objects analyzed were purportedly in the public domain, with another 8% of results classified as author public domain (usually a museum or archive visitor who took a photo of a collection or item). Fourteen percent of objects were licensed using Creative Commons licenses, and the remaining 1% of results were either copyrighted and published with permission from the copyright holder or released under a Free Art License, a copyleft license for creative works that has been deemed compatible with BY-SA by Creative Commons. Of the 43 objects licensed with Creative Commons licenses, 27 included the “ShareAlike” designation that requires derivative creations use the same license. Figure 2 provides a visual breakdown of the licenses used in WMC.

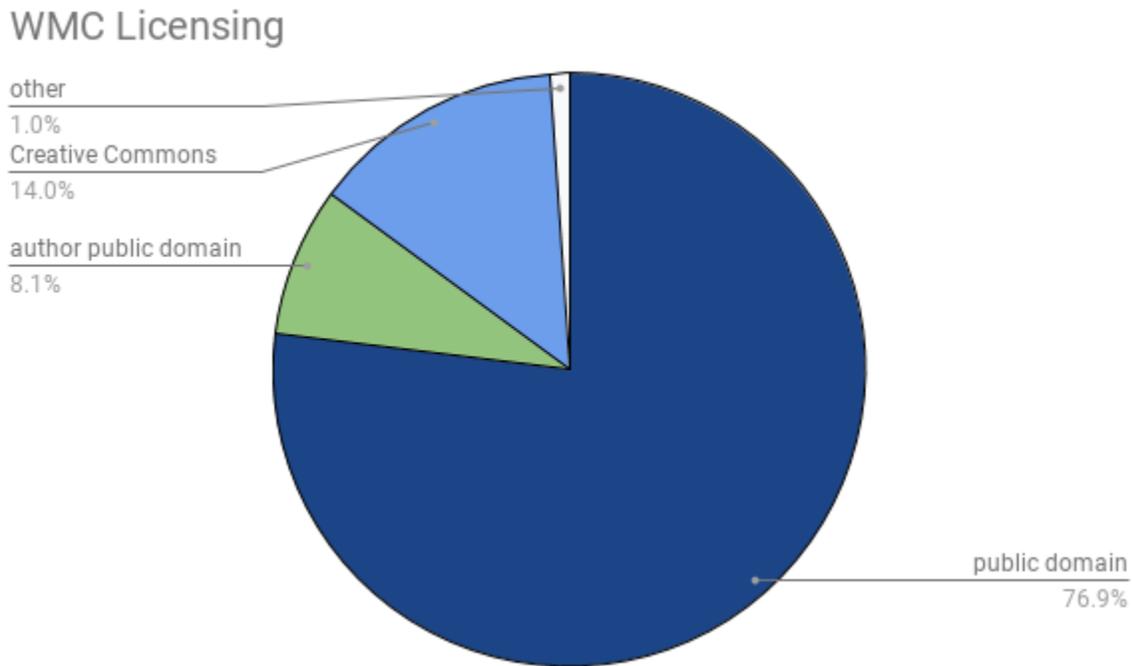


Figure 2. WMC licensing. Licenses used for analyzed cultural heritage objects uploaded to Wikimedia Commons.

Based on an evaluation of provided creation date and creator information, an impressive 91% of these licenses and rights statements were determined to accurately depict the copyright, public domain, or open-license status of each object. Only 3% were interpreted to be incorrect, while 6% could not be determined using the information provided, as seen in Table 2.

Accuracy of WMC Right Statement (n = 308)	Number	Percent
Correct	280	91
Incorrect	10	3
Unsure	18	6

Table 2. Evaluation of WMC rights statements. Validity of rights statements for analyzed digital cultural heritage objects on Wikimedia Commons, as compared to the creation date and original source.

Of the 308 WMC objects evaluated, 114 had working links to external sources with which to compare the WMC rights statement or license. Seventy-six percent of the licenses were compatible with each other, 16% were incompatible, and 4% were

inconclusive—usually because the rights statement on the original source was vague or misleading. Another 4% were compatible but incorrect (see Table 3).

Compatibility of WMC and Original Source Rights Statements (n = 114)	Number	Percent
Compatible	87	76
Incompatible	18	16
Inconclusive	5	4
Compatible but incorrect	4	4

*Table 3.* Compatibility of WMC and original source rights statements

### Discussion

This study finds that rights statements and licenses for digital cultural heritage objects on WMC are largely trustworthy, perhaps even more so than their counterparts in institutionally created digital libraries. This may not come as a surprise, as the professional literature shows that many institutions have struggled with determining the copyright status of their digital collections and applying accurate rights statements. It is possible that the collaborative and peer-reviewed nature of WMC creates greater opportunities for evaluation and correction of rights statements compared to a cultural heritage institution in which library practitioners may have limited training, support, and feedback in their rights and licensing assessment of digital objects. Further research should continue to identify the effect of clear and accurate rights statements on digital object reuse and the barriers faced by library practitioners in aptly applying such statements.

In response to the question of what are the most common inaccuracies found in WMC cultural heritage object rights statements, a deeper dive into the 10 WMC objects with incorrect rights statements is necessary. Nine of the 10 objects were labeled as either Creative Commons or CC0 1.0 Universal (CC0 1.0), referred to in Figure 2 as “author public domain.” The purpose of the CC0 designation is for a rightsholder to declare that they are releasing the rights to something for which they own the rights. In the case of the WMC objects mentioned here, the objects were already in the public domain and could not be further licensed (and subsequently released) by a Wikipedian. While this may seem like a minor distinction, it points to confusion shared by both cultural heritage institutions and individual users as to who can designate a license. Two of these results were credited to individual authors who seem to misunderstand that digital surrogates of two-dimensional objects are still in the public domain. Three results were cited in the “source” field as being from cultural heritage collections, but there were no corresponding digital library links with which to compare license statements. Four

results came from cultural heritage digital libraries that had incorrect rights statements on their websites (or, in one case, numerous Creative Commons licenses were listed for the same object). Again, this demonstrates the need for clear and correct rights statements to ensure correct licensing when objects are reused.

While over 90% of WMC digital cultural heritage objects seem to have accurately applied licensing, only 76% of those licenses are compatible with, or match, corresponding digital libraries and websites where the object was presumably found before being uploaded to WMC. Again, the cause of inconsistency stems more from inaccuracies in the original source than from mistakes on WMC's part, perhaps due to WMC's constant cycle of peer review. For the 18 results where the original source license was incompatible with WMC, WMC was correct and the mistake was found in the original source. For the four sources where the original and WMC licenses were compatible but incorrect, presumably the license was copied from the original source to WMC.

However, a characteristic of WMC's markup and requirements for licensing information does highlight room for improved education and/or documentation for some of the intricacies of copyright on WMC. Of the analyzed WMC objects, 115 were flagged as missing either a copyright or public domain tag. The copyright and public domain tags are further divisions of the legal justifications for an object being in the public domain, and these tags include designations such as ("Commons:Copyright tags," n.d.):

- Copyright expired due to age
- Objects released into the public domain by the creator
- Objects created by animals and therefore not copyrightable

There are hundreds of available tags, which demonstrates the complexity of trying to provide clear rights statements to a diverse and potentially international audience. From this preliminary analysis, it is unclear if users do not always understand why their objects are public domain and therefore leave out the tag, if the necessity of embedding the tag using wiki markup is a barrier, or if some other reason prevents users from fully articulating the reason for the designated licensing. But the lack of precise copyright and public domain tags means that some WMC object licensing is vague; this also prevents accurate sorting of WMC objects by licensing categories.

This analysis points to a relatively reliable level of accuracy in general for digital cultural heritage objects on WMC, but there is still room for improvement, particularly if the credibility or reliability of WMC is at stake. The structured data project page on WMC details some possible developments in how licensing, rights, and permissions may be implemented for media files in the future ("Commons:Structured data/Get involved/Feedback requests/First licensing consultation," n.d.). Structured data—or data coded in a specific, fixed field—could

allow data regarding rights for WMC media to be more useful in the long term. A discussion of two different approaches to implementing structured data rights statements was conducted through April 2018 on a related “Commons talk” page (“Commons talk:Structured data/Get involved/Feedback requests/First licensing consultation,” n.d.). The approaches detailed included adapting existing WMC practices for the structured license and copyright model and adapting existing (and potentially interoperable) structured rights statements, like [rightsstatements.org](http://rightsstatements.org). As users on the “Commons talk” page discuss, the audience for WMC and for digital libraries that use [rightsstatements.org](http://rightsstatements.org) differs enough that full-on implementation of [rightsstatements.org](http://rightsstatements.org) may be inadvisable. The purpose of [rightsstatements.org](http://rightsstatements.org) is to ensure clear and concise rights information so that users quickly know what they can and cannot do with uploaded media. In many cases the cultural heritage institution uploading media to a digital library has more documented information about the media’s rights than anyone else and should (ideally) be relied upon as having correct rights statements. The community infrastructure of WMC, and the reuse of digital media from multitudinous resources on WMC, few or none of which are owned by the Wikimedia Foundation, requires both that rights statements be clear and concise and also that the cultural heritage institution’s underlying reasoning for selection of the license also be made available. This could point to the need for multiple required metadata fields on WMC. In addition to an open-text “source” field, the following may all be needed: a controlled vocabulary field akin to [rightsstatements.org](http://rightsstatements.org), copyright/public domain tags, and another field with evidence proving the accuracy of the license. The more that is required in a WMC record, particularly if the record necessitates that the Wikipedian be familiar with the way wiki tags and markup are formatted, the more training and documentation might be needed to safeguard the reliability of rights statements on WMC.

Adding these same metadata fields to cultural heritage digital libraries, explaining the reasoning behind designated rights statements, and providing evidence of accuracy could be even more effective since institutions have struggled with this work in the past. Typically the cultural heritage institution is only required to provide proof of the rights statement if requested by a publisher or in the event of a legal challenge to the license; otherwise no additional proof of the license is usually included in a digital library, thus obscuring the decisions behind the application of different rights statements and making it difficult to determine their accuracy. In addition, a system for feedback or a working group with copyright expertise who could provide analysis of library practitioners’ stated licensing and rights information in digital library metadata when requested could lead to better use of rights statements overall. While professional and educational organizations like the Society of American Archivists, Creative Commons, the Institute for Museum and Library Services–funded Library Copyright Institute, and Harvard University’s

CopyrightX offer training and certification courses in copyright and open licensing, there is obviously still a need for better training of library practitioners, especially those working with unique cultural heritage materials.

Finally, it should be noted that the author is not advocating for all cultural heritage to be made available digitally and licensed openly for use on WMC and other projects. There are certainly cases where access restrictions are necessary for reasons both legal and ethical, especially in cases of collections from marginalized communities. As detailed in their 2010 article, Eschenfelder and Caswell identified some beneficial uses of access restriction such as “regulating access and use of culturally sensitive materials in order to protect the source group that generated the material” and protecting the rights of individuals. While outside of the scope of this article, the Documenting the Now project advocates for the ethical collection of social media archives, and the Architecting Sustainable Futures project looks to enable community-based archives to document marginalized communities while emphasizing “the importance of controlled access and authority,” an integral part of Mukurtu, an open-source content management system built with and for indigenous communities (Architecting Sustainable Futures, “About,” n.d.; DocNow, “Documenting the Now,” n.d.; Historypin, 2018). Considerations relevant to protecting the rights of source communities to their cultural heritage materials often fall outside of copyright law and require additional nuanced permissions and terms of use developed by or in consultation with source communities and creators.

### **Conclusion**

Undoubtedly the prevalence of digital library objects on WMC speaks to the importance of including clear and accurate rights statements in digital libraries to ensure that reuse of digital library objects falls within the parameters of copyright law and institutional terms of use. The results of this study show that, though rare, inaccuracies in rights statements for digital cultural heritage objects on WMC are usually the fault of the cultural heritage institutions using incorrect or misleading rights statements in their digital libraries and websites. Digital cultural heritage objects that are in the public domain should be clearly labeled as such to encourage their reuse, which is in line with their cultural heritage institutions’ missions. Greater education and documentation of some of the intricacies of copyright law that affect cultural heritage institutions, such as the differences in copyright for digital reproductions of two- and three-dimensional objects, could also lead to greater accuracy in rights statements and licensing on WMC. Further research should be conducted to determine the effect of inaccurate rights statements that appear on WMC on the wider reuse of digital cultural heritage objects.

## References

- Anonymous. (2009). Collections are for use, but is Wikipedia the outlet? *Library Journal*, 134(19), 15–16.
- Architecting Sustainable Futures. (n.d.). About the project [Webpage]. Retrieved April 8, 2019, from <https://www.architectingsustainablefutures.net/about>
- Ballinger, L., Karl, B., & Chiu, A. (2017). Providing quality rights metadata for digital collections through rightsstatements.org. *Pennsylvania Libraries: Research & Practice*, 5(2), 144–158. <https://doi.org/10.5195/palrap.2017.157>
- Bauder, J., & Emanuel, J. (2012). Being where our faculty are: Emerging technology use and faculty information-seeking workflows. *Internet Reference Services Quarterly*, 17(2), 65–82. <https://doi.org/10.1080/10875301.2012.718316>
- Brinkerink, M. (2015, June 23). Dutch cultural heritage reaches millions every month. OpenGLAM. Retrieved from <https://openglam.org/2015/06/23/dutch-cultural-heritage-reaches-millions-every-month/>
- Capell, L., & Williams, E. (2016). *Assigning rights statements to legacy digital collections*. Poster session presented at the Rare Books & Manuscript Section conference, Coral Gables, FL. Retrieved from <https://scholarlyrepository.miami.edu/librarypapers/2>
- Chesney, T. (2006, November). An empirical examination of Wikipedia's credibility. *First Monday*, 11(1). Retrieved from <https://firstmonday.org/article/view/1413/1331>
- Commons:Categories. (n.d.). Wikimedia Commons. Retrieved April 4, 2019, from <https://commons.wikimedia.org/wiki/Commons:Categories>
- Commons:Copyright tags. (n.d.). Wikimedia Commons. Retrieved April 4, 2019, from [https://commons.wikimedia.org/wiki/Commons:Copyright\\_tags](https://commons.wikimedia.org/wiki/Commons:Copyright_tags)
- Commons:Email templates. (n.d.). Wikimedia Commons. Retrieved April 4, 2019, from [https://commons.wikimedia.org/wiki/Commons:Email\\_templates](https://commons.wikimedia.org/wiki/Commons:Email_templates)
- Commons:First steps/License selection. (n.d.). Wikimedia Commons. Retrieved April 4, 2019, from [https://commons.wikimedia.org/wiki/Commons:First\\_steps/License\\_selection](https://commons.wikimedia.org/wiki/Commons:First_steps/License_selection)
- Commons:Hirtle chart. (n.d.). Wikimedia Commons. Retrieved April 4, 2019, from [https://commons.wikimedia.org/wiki/Commons:Hirtle\\_chart](https://commons.wikimedia.org/wiki/Commons:Hirtle_chart)
- Commons:Licensing. (n.d.). Wikimedia Commons. Retrieved April 4, 2019, from <https://commons.wikimedia.org/wiki/Commons:Licensing>
- Commons:Non-copyright restrictions. (n.d.). Wikimedia Commons. Retrieved April 4, 2019, from [https://commons.wikimedia.org/wiki/Commons:Non-copyright\\_restrictions](https://commons.wikimedia.org/wiki/Commons:Non-copyright_restrictions)

- Commons:Structured data/Get involved/Feedback requests/First licensing consultation. (n.d.). Wikimedia Commons. Retrieved April 3, 2019, from [https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Get\\_involved/Feedback\\_requests/First\\_licensing\\_consultation](https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Feedback_requests/First_licensing_consultation)
- Commons talk:Structured data/Get involved/Feedback requests/First licensing consultation. (n.d.). Wikimedia Commons. Retrieved April 8, 2019, from [https://commons.wikimedia.org/wiki/Commons\\_talk:Structured\\_data/Get\\_involved/Feedback\\_requests/First\\_licensing\\_consultation](https://commons.wikimedia.org/wiki/Commons_talk:Structured_data/Get_involved/Feedback_requests/First_licensing_consultation)
- Cooban, G. (2017). Should archivists edit Wikipedia, and if so how? *Archives & Records*, 38(2), 257–272. <https://doi.org/10.1080/23257962.2017.1338561>
- Creative Commons. Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0). (n.d.). Retrieved April 4, 2019, from <https://creativecommons.org/licenses/by-sa/3.0/>
- Documenting the Now. (n.d.). Retrieved April 8, 2019, from <https://www.docnow.io/>
- Dowell, M. L., & Bridges, L. M. (2019). A perspective on Wikipedia: Your students are here, why aren't you? *Journal of Academic Librarianship*, 45(2), 81–83. <https://doi.org/10.1016/j.acalib.2019.01.003>
- Dryden, J. (2011). Copyfraud or legitimate concerns? Controlling further uses of online archival holdings. *American Archivist*, 74(2), 522–543. <https://doi.org/10.17723/aarc.74.2.d5g2700q5612l4w7>
- Dryden, J. (2012). Cavalier or careful? How users approach the rights management practices of archival repositories. *Journal of Archival Organization*, 10(3/4), 191–206. <https://doi.org/10.1080/15332748.2012.787847>
- Elder, D., Westbrook, R. N., & Reilly, M. (2012). Wikipedia lover, not a hater: Harnessing Wikipedia to increase the discoverability of library resources. *Journal of Web Librarianship*, 6(1), 32–44. <https://doi.org/10.1080/19322909.2012.641808>
- Erickson, K., Heald, P. J., Homberg, F., Kretschmer, M., & Mendis, D. (2015). *Copyright and the value of the public domain: An empirical assessment* (SSRN Scholarly Paper No. ID 2571220). Retrieved from World Intellectual Property Office website: [https://www.wipo.int/edocs/mdocs/mdocs/en/wipo\\_ip\\_econ\\_ge\\_1\\_15/wipo\\_ip\\_econ\\_ge\\_1\\_15\\_ref\\_erickson.pdf](https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_econ_ge_1_15/wipo_ip_econ_ge_1_15_ref_erickson.pdf)
- Eschenfelder, K. R., & Caswell, M. (2010). Digital cultural collections in an age of reuse and remixes. *First Monday*, 15(11). <https://doi.org/10.5210/fm.v15i11.3060>
- Ferriero, D. S. (2011). On the growing relationship between the National Archives and Wikipedia. *Vital Speeches of the Day*, 77(10), 367–369.

- Galloway, E., & DellaCorte, C. (2014). Increasing the discoverability of digital collections using Wikipedia. *Pennsylvania Libraries: Research & Practice*, 2(1), 84–96. <https://doi.org/10.5195/palrap.2014.60>
- Hirtle, P. B. (2001, Fall). Unpublished materials, new technologies, and copyright: Facilitating scholarly use. *Journal of the Copyright Society of the USA*, 49(1). Retrieved from <https://ecommons.cornell.edu/handle/1813/58>
- Historypin. (2018). *Moving beyond colonial models of digital memory* (based on planning grant LG-72-16-0113-16). Institute of Museum and Library Services. Retrieved from [https://drive.google.com/file/d/1xjIN\\_oyOKm8iQc\\_yuz6drEVNqGNGaDoD/view](https://drive.google.com/file/d/1xjIN_oyOKm8iQc_yuz6drEVNqGNGaDoD/view)
- Kelly, E. J. (2018a). Content analysis of Google alerts for cultural heritage institutions. *Journal of Web Librarianship*, 12(1), 28–45. <https://doi.org/10.1080/19322909.2017.1369374>
- Kelly, E. J. (2018b). Use of Louisiana's digital cultural heritage by Wikipedians. *Journal of Web Librarianship*, 12(2), 85–106. <https://doi.org/10.1080/19322909.2017.1391733>
- Lally, A. M., & Dunford, C. E. (2007). Using Wikipedia to extend digital collections. *D-Lib Magazine*, 13(5/6). <https://doi.org/10.1045/may2007-lally>
- Miller, R. (2014, August 29). Millions of historic images posted to Flickr [Blog post]. Retrieved from <https://blog.archive.org/2014/08/29/millions-of-historic-images-posted-to-flickr/>
- Nagaraj, A. (2018). Does copyright affect reuse? Evidence from Google Books and Wikipedia. *Management Science*, 64(7), 3091–3107. <https://doi.org/10.1287/mnsc.2017.2767>
- Needham, C. (2017). Understanding copyfraud: Public domain images and false claims of copyright. *Art Documentation: Bulletin of the Art Libraries Society of North America*, 36(2), 219–230. <https://doi.org/10.1086/694241>
- Perrin, J. M., Winkler, H., Daniel, K., Barba, S., & Yang, L. (2017). Know your crowd: A case study in digital collection marketing. *Reference Librarian*, 58(3), 190–201. <https://doi.org/10.1080/02763877.2016.1271758>
- Resolution: Licensing policy. (n.d.). Wikimedia Foundation governance wiki. Retrieved April 4, 2019, from [https://foundation.wikimedia.org/wiki/Resolution:Licensing\\_policy](https://foundation.wikimedia.org/wiki/Resolution:Licensing_policy)
- Schlosser, M. (2009). Unless otherwise indicated: A survey of copyright statements on digital library collections. *College & Research Libraries*, 70(4), 371–385. <https://doi.org/10.5860/0700371>
- Sims, N. (2017). Rights, ethics, accuracy, and open licenses in online collections. *College & Research Libraries News*, 78(2), 79–82. <https://doi.org/10.5860/crln.78.2.9620>

- Snyder, J. (2013). Wikipedia: Librarians' perspectives on its use as a reference source. *Reference & User Services Quarterly*, 53(2), 155–163. <https://doi.org/10.5860/rusq.53n2.155>
- Stitzlein, H., Han, M.-J. K., & Benson, S. R. (2018). Unraveling challenges: Rights statements in digital cultural heritage collections. *Journal of Library Metadata*, 18(3/4), 135–150. <https://doi.org/10.1080/19386389.2018.1540328>
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques & procedures for developing grounded theory* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Szajewski, M. (2013). Using Wikipedia to enhance the visibility of digitized archival assets. *D-Lib Magazine*, 19(3/4). <https://doi.org/10.1045/march2013-szajewski>
- Terras, M. (2015a). Opening access to collections: The making and using of open digitised cultural content. *Online Information Review*, 39(5), 733–752. <https://doi.org/10.1108/OIR-06-2015-0193>
- Terras, M. (2015b). So you want to reuse digital heritage content in a creative context? Good luck with that. *Art Libraries Journal*, 40(4), 33–37. <https://doi.org/10.1017/S0307472200020502>
- Vetter, M., & Harrington, S. (2013). Integrating special collections into the composition classroom: A case study of collaborative digital curriculum. *Research Library Issues*, 283, 16–20. <https://doi.org/10.29242/rli.283.4>

Appendix A: List of Repositories

Agua Caliente Cultural Museum  
Alexandria Library, Special Collections  
Amon Carter Museum  
Arizona State Museum Library and Archives  
Arkansas History Commission  
Austin Public Library, Austin History Center  
Bard College, Center for Curatorial Studies Library  
Barnes Foundation, Archives, Libraries, and Special Collections  
Bryn Mawr College  
Central Michigan University, Clarke Historical Library  
Centre College, Grace Doherty Library  
Chula Vista Public Library, John Rojas Local History Room  
Cincinnati Art Museum, Mary R Schiff Library and Archives  
Cleveland Public Library  
College of Charleston, Marlene and Nathan Addlestone Library  
College of Physicians of Philadelphia  
College of William and Mary, Special Collections Research Center  
Computer History Museum  
District of Columbia Public Library  
Eastern Kentucky University, Crabbe Library  
Folger Shakespeare Library  
Forest History Society  
Fresno City and County Historical Society Archives  
Fresno Pacific University, Hiebert Library  
Georgetown University, Special Collections  
Gerald R. Ford Library  
Hagley Museum and Library, Manuscripts and Archives Department  
Harvard Film Archive  
Harvard University, Loeb Music Library  
Holocaust Center of Northern California Archives  
Idaho State University, Department of Special Collections and University Archives  
Indiana University, Folklore Collection  
Lamar University, Mary and John Gray Library  
Lancaster County Historical Society, Archives  
Los Angeles Maritime Museum  
Missouri State University, Meyer Library  
National Gallery of Art  
Oakland Museum  
Princeton University, Firestone Library, Latin American Ephemera Collections

Richmond Public Library  
Saint Mary's College, Saint Albert Hall  
Santa Clara University  
Seton Hall University  
Smithsonian Institution Archives  
Stanford University, Archive of Recorded Sound  
Tennessee State University  
Tennessee Technical University, Angelo and Jeanette Volpe Library  
The Henry Ford, Benson Ford Research Center  
The New School, Archives and Special Collections and Kellen Design Archives  
Trinity College, Watkinson Library  
University of Denver, Penrose Library  
University of Idaho  
University of Louisiana at Lafayette, Edith Garland Dupre Library  
University of Michigan, Bentley Historical Library  
University of Missouri, Kansas City, Miller Nichols Library  
University of North Florida, Special Collections  
University of Pittsburgh  
University of Puget Sound, Collins Memorial Library  
University of St. Thomas, O'Shaughnessy-Frey Library  
University of Tennessee at Chattanooga  
University of Texas at Austin, Alexander Architectural Archive  
University of Utah, J. Willard Marriott Library  
Visual Communications Archives and Media Resource Library  
Winthrop University, Dacus Library  
Wisconsin Historical Society Library and Archives  
Wood Memorial Library and Museum  
Yale University, Beinecke Rare Book and Manuscript Library  
Yale University, Irving S. Gilmore Music Library  
Yale University, Manuscripts and Archives

## Appendix B: Results by Repository

Repository	WMC Results Analyzed
Alexandria Library, Special Collections	2
Amon Carter Museum	13
Arizona State Museum	1
Austin Public Library, Austin History Center	14
Bard College, Center for Curatorial Studies Library	3
Barnes Foundation	3
Bryn Mawr College	12
Central Michigan University, Clarke Historical Library	3
Centre College, Grace Doherty Library	13
Chula Vista Public Library, John Rojas Local History Room	1
Cincinnati Art Museum, Mary R Schiff Library and Archives	13
Cleveland Public Library	18
College of Charleston, Marlene and Nathan Addlestone Library	7
College of Physicians of Philadelphia	11
College of William and Mary, Special Collections Research Center	8
Computer History Museum	9
District of Columbia Public Library	10
Eastern Kentucky University, Crabbe Library	1
Folger Shakespeare Library	9
Forest History Society	6
Fresno City and County Historical Society Archives	1
Georgetown University, Special Collections	13
Gerald R. Ford Library	6
Hagley Museum and Library	10
Harvard University Loeb Music Library	1
Idaho State University, Department of Special Collections and University Archives	2
Indiana University	4
Lamar University, Mary and John Gray Library	1
Lancaster County Historical Society, Archives	2
Missouri State University, Meyer Library	1
National Gallery of Art	7
Oakland Museum	9
Princeton University, Firestone Library, Latin American Ephemera Collections	2

Richmond Public Library	1
Saint Mary's College, Saint Albert Hall	2
Santa Clara University	5
Seton Hall University	1
Smithsonian Institution Archives	10
Stanford University Archive	2
Tennessee State University	1
The Henry Ford, Benson Ford Research Center	5
Trinity College, Watkinson Library	1
University of Denver	4
University of Idaho	7
University of Louisiana at Lafayette, Edith Garland Dupre Library	3
University of Michigan, Bentley Historical Library	6
University of Missouri, Kansas City, Miller Nichols Library	3
University of North Florida, Special Collections	1
University of Pittsburgh	8
University of Puget Sound	6
University of Tennessee at Chattanooga	1
University of Texas at Austin	3
University of Utah	1
Winthrop University, Dacus Library	2
Wisconsin Historical Society	9
Yale University, Beinecke Rare Book and Manuscript Library	5
Yale University, Manuscripts and Archives	5