

Reuse of Wikimedia Commons Cultural Heritage Images on the Wider Web

Elizabeth Kelly

Volume 14, numéro 3, 2019

URI : <https://id.erudit.org/iderudit/1088933ar>

DOI : <https://doi.org/10.18438/ebliip29575>

[Aller au sommaire du numéro](#)

Éditeur(s)

University of Alberta Library

ISSN

1715-720X (numérique)

[Découvrir la revue](#)

Citer cet article

Kelly, E. (2019). Reuse of Wikimedia Commons Cultural Heritage Images on the Wider Web. *Evidence Based Library and Information Practice*, 14(3), 28–51. <https://doi.org/10.18438/ebliip29575>

Résumé de l'article

Objective – Cultural heritage institutions with digital images on Wikimedia Commons want to know if and how those images are being reused. This study attempts to gauge the impact of digital cultural heritage images from Wikimedia Commons by using Reverse Image Lookup (RIL) to determine the quantity and content of different types of reuse, barriers to using RIL to assess reuse, and whether reused digital cultural heritage images from Wikimedia Commons include licensing information.

Methods – 171 digital cultural heritage Wikimedia Commons images from 51 cultural heritage institutions were searched using the Google images “Search by image” tool to find instances of reuse. Content analysis of the digital cultural heritage images and the context in which they were reused was conducted to apply broad content categories. Reuse within Wikimedia Foundation projects was also recorded.

Results – A total of 1,533 reuse instances found via Google images and Wikimedia Commons’ file usage reports were analyzed. Over half of reuse occurred within Wikimedia projects or wiki aggregator and mirror sites. Notable People, people, historic events, and buildings and locations were the most widely reused topics of digital cultural heritage both within Wikimedia projects and beyond, while social, media gallery, news, and education websites were the most likely places to find reuse outside of wiki projects. However, the content of reused images varied slightly depending on the website type on which they were found. Very few instances of reuse included licensing information, and those that did often were incorrect. Reuse of cultural heritage images from Wikimedia Commons was either done without added context or content, as in the case of media galleries, or was done in ways that did not distort or mischaracterize the images being reused.

Conclusion – Cultural heritage institutions can use this research to focus digitization and digital content marketing efforts in order to optimize reuse by the types of websites and users that best meet their institution’s mission. Institutions that fear reuse without attribution have reason for concern as the practice of reusing both Creative Commons and public domain media without rights statements is widespread. More research needs to be conducted to determine if notability of institution or collection affects likelihood of reuse, as preliminary results show a weak correlation between number of images searched and number of images reused per institution. RIL technology is a reliable method of finding image reuse but is a labour-intensive process that may best be conducted for selected images and specific assessment campaigns. Finally, the reused content and context categories developed here may contribute to a standardized set of codes for assessing digital cultural heritage reuse.

© Elizabeth Kelly, 2019



Ce document est protégé par la loi sur le droit d’auteur. L’utilisation des services d’Érudit (y compris la reproduction) est assujettie à sa politique d’utilisation que vous pouvez consulter en ligne.

<https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

érudit

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l’Université de Montréal, l’Université Laval et l’Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

<https://www.erudit.org/fr/>



Research Article

Reuse of Wikimedia Commons Cultural Heritage Images on the Wider Web

Elizabeth Kelly, C.A., D.A.S.
Digital Programs Coordinator
Loyola University
New Orleans, Louisiana, United States of America
Email: ejkelly@loyno.edu

Received: 23 Apr. 2019

Accepted: 25 May 2019

■ 2019 Kelly. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 4.0 International (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

DOI: 10.18438/ebliip29575

Abstract

Objective – Cultural heritage institutions with digital images on Wikimedia Commons want to know if and how those images are being reused. This study attempts to gauge the impact of digital cultural heritage images from Wikimedia Commons by using Reverse Image Lookup (RIL) to determine the quantity and content of different types of reuse, barriers to using RIL to assess reuse, and whether reused digital cultural heritage images from Wikimedia Commons include licensing information.

Methods – 171 digital cultural heritage Wikimedia Commons images from 51 cultural heritage institutions were searched using the Google images “Search by image” tool to find instances of reuse. Content analysis of the digital cultural heritage images and the context in which they were reused was conducted to apply broad content categories. Reuse within Wikimedia Foundation projects was also recorded.

Results – A total of 1,533 reuse instances found via Google images and Wikimedia Commons’ file usage reports were analyzed. Over half of reuse occurred within Wikimedia projects or wiki aggregator and mirror sites. Notable People, people, historic events, and buildings and locations were the most widely reused topics of digital cultural heritage both within Wikimedia projects

and beyond, while social, media gallery, news, and education websites were the most likely places to find reuse outside of wiki projects. However, the content of reused images varied slightly depending on the website type on which they were found. Very few instances of reuse included licensing information, and those that did often were incorrect. Reuse of cultural heritage images from Wikimedia Commons was either done without added context or content, as in the case of media galleries, or was done in ways that did not distort or mischaracterize the images being reused.

Conclusion – Cultural heritage institutions can use this research to focus digitization and digital content marketing efforts in order to optimize reuse by the types of websites and users that best meet their institution’s mission. Institutions that fear reuse without attribution have reason for concern as the practice of reusing both Creative Commons and public domain media without rights statements is widespread. More research needs to be conducted to determine if notability of institution or collection affects likelihood of reuse, as preliminary results show a weak correlation between number of images searched and number of images reused per institution. RIL technology is a reliable method of finding image reuse but is a labour-intensive process that may best be conducted for selected images and specific assessment campaigns. Finally, the reused content and context categories developed here may contribute to a standardized set of codes for assessing digital cultural heritage reuse.

Introduction

Cultural heritage institutions with digital images online want to know if and how those images are being reused. Whether the image was uploaded to a digital library by the institution or added to a website by an individual user, knowledge and understanding of digital image reuse helps cultural institutions determine the impact of their collections as well as whether they are meeting the needs of their users. One method of measuring reuse of digital images online is Reverse Image Lookup (RIL), in which the RIL service searches the internet for other versions of an image. Recent scholarship includes several RIL studies of digital cultural heritage media from specific collections or institutions. However, research by the Wikimedia Foundation has found that cultural heritage institutions with digital media in Wikimedia Commons, the media repository for Wikimedia Foundation projects, want better understanding of the impact of their uploaded media, in particular as it relates to institutional

goals (Research:Supporting Commons contribution, 2018). As increasing numbers of cultural heritage institutions upload their digital media to Wikimedia Commons, and as users add digital cultural media found during their own research, the opportunity and necessity of assessing the impact of these objects becomes more relevant.

This study attempts to gauge the impact that digital cultural heritage images from Wikimedia Commons have both in and beyond wiki projects by using RIL to determine quantity and quality of different types of reuse while also identifying barriers to assessing reuse in this way. Rooted in empirical evidence, this study will provide concrete examples of how digital cultural heritage from Wikimedia Commons is used outside of the Wikimedia landscape along with documented steps for finding and analyzing image reuse in order to facilitate greater reuse research among digital cultural heritage stakeholders, leading to improvements in efforts to make digital collections more widely available and reusable.

Media Reuse Studies

Media reuse research is still a relatively new field without standard or widely accepted definitions of use and reuse. The Digital Library Federation Assessment Interest Group (DLF-AIG) Content Reuse working group completed a 1-year Institute of Museum and Library Services (IMLS) grant in 2018 to evaluate the needs and functions of a digital library reuse toolkit, and in doing so also researched digital library stakeholder interpretations of use and reuse. While refined definitions of use and reuse by the group are forthcoming, at this time and for the purposes of this paper reuse will be defined as “how often and in what ways digital library materials are utilized and repurposed” and in what contexts (O’Gara et al., 2018). Collection curators, digital librarians, and archivists find value in assessing the reuse of their digital collections in order to show the collection’s reach and to determine who uses collections. This data can then be used to make decisions about collection development and digitization priorities as well as to negotiate increases in staffing and funding.

While digital library stakeholders find a great deal of value in assessing the reuse of their collections, they also find it very difficult to do. A survey administered by the DLF-AIG Content Reuse IMLS project team found that only 40% of respondents were gathering reuse data, usually from social media metrics or citation analysis (O’Gara et al., 2018).

There is also tension between cultural heritage organizations’ missions to provide access and a desire to maintain control over collections. Sometimes there are valid and commendable reasons for wishing to restrict access or mediate use and reuse of digital collections. Digital content misuse and cultural appropriation are concerns for digital library stakeholders (O’Gara et al., 2018). Ethnographic archives, especially those that document the history and cultures of marginalized populations, prove challenging to determine meaningful impact beyond simple

quantitative metrics such as clicks, likes, and downloads (Punzalan et al., 2018). Other times, however, archives unnecessarily attempt to control reuse of their online holdings via restrictive or unclear rights statements (Dryden, 2014).

While published literature about media reuse is still somewhat limited, the existing scholarship primarily focuses on use and reuse of specific archival and digital collections, reuse of generalized collections by scholars within specific areas of study, and reuse of specific types of media. These studies are often undertaken with the purpose of improving the services and technological infrastructure that make library and archival collections reusable by researchers. Studies involving focus groups, observational research, and citation analyses have evaluated the reuse of archival images by historians, archaeologists, architects, and artists (Beaudoin, 2014; Harris & Hepburn, 2013). Additional researchers, after creating or using digital media collections in their own work, have advocated for the creation of open-licensed digital collections of geology and film in order to enhance the research process for students and scholars alike (O’Sullivan, 2017; Rygel, 2013).

The reuse of digital cultural heritage media on social media platforms has received increasing attention in the scholarly literature over the course of the last decade. As noted in one study, “our data indicate that everyday users are repurposing digital content in ways that are meaningful to them, and they are acknowledging and fulfilling personal interests. These users are also sharing this content through a variety of environments on the Web, including popular social media platforms, blogs, and personal Web sites” (Reilly & Thompson, 2017). Social media platforms like Pinterest, which allow users to curate personal collections of images, blog posts, and other media from the web, have an “archival shape” due to their infrastructure that captures the provenance, or original source, of the item, making such platforms rich for analysis by media reuse

researchers (Summers, 2019). Examples of cultural heritage media reuse could include images downloaded from digital library collections and uploaded onto a Pinterest Pinboard, as well as those reproduced in commercial projects like artwork or included in official government reports (Thompson & Reilly, 2017). Reuse of digital cultural heritage media on Wikimedia Commons, Wikipedia, and other Wikimedia Foundation projects has also received scholarly attention in the last year (Kelly, 2018; Morley, 2018).

One of the most widely documented methods for evaluating digital image reuse involves RIL services such as Google images or TinEye, in which an image is either uploaded or an originating URL is input to the search platform and then duplicates and similar images are found online. RIL studies have been performed on images from NASA, academic digital libraries, the Library of Congress, and the British National Gallery (Kelly, 2015; Kirton & Terras, 2013; Kousha et al., 2010; Reilly & Thompson, 2014; Reilly & Thompson, 2017). In all of these studies, after duplicate images were found online, the context and purpose in which the images were reused was analyzed in order to determine who uses digital cultural heritage images and for what objective.

Cultural Heritage, Wikimedia, and Impact

A ready-made platform for sharing digital cultural heritage media and encouraging reuse can be found in Wikimedia Commons (commons.wikimedia.org), the Wikimedia Foundation's repository for photographs, artwork, video, sound, diagrams, and more. Many cultural heritage institutions have developed programs to upload their digital media to Wikimedia Commons and enhance Wikipedia articles with links to their collections and finding aids in order to increase traffic to their websites and repositories, typically with impressive results (Kelly, 2018). Digital cultural heritage media is added to Wikimedia Commons in a few ways:

- Cultural heritage institutions upload media from their own existing digital collections;
- Cultural heritage institutions upload media directly to the Commons, especially in the case of smaller institutions without existing digital repositories;
- Cultural heritage institutions and users upload media to other repositories or websites, such as Flickr and the Internet Archive, that are then crawled by bots and added to the Commons;
- Users upload media from cultural heritage institution digital collections;
- Users make their own digital reproductions of cultural heritage collections (for example, photographing a painting in a museum, or a document in an archive) and then upload them to the Commons.

Wikimedia Commons provides user guidelines on how to reuse media from the Commons on Wikimedia platforms as well as outside of the Wikimedia landscape (Commons:First steps/Reuse, 2019; Commons:Reusing content, 2018; Commons:Simple media reuse guide, 2018). But just as digital library stakeholders struggle to assess reuse of the media in their own repositories, Wikipedia editors and authors, or Wikipedians, struggle to assess reuse of projects, articles, and media from Wikimedia Foundation programs. Denny Vrandečić points out that readily available use metrics do not always show what is valuable or important, and instead "we should focus on measuring how much knowledge we allow every human to share in, instead of number of articles or active editors" (2014). Another Wikipedian argues that "The sum of human knowledge" is not the same concept as "the sum of what everyone is googling today" and that reach, importance, diversity and content gaps, uniqueness, and quality are all necessary primary measures of impact for the Wikimedia movement (User:The land, 2018). The Wikimedia Foundation "Supporting Commons contribution by GLAM

institutions” research project (GLAM standing for Galleries, Libraries, Archives, and Museums) noted that for cultural heritage organizations, “donating media to Commons is a means to an end. GLAM organizations and the volunteers who work with them want to know the media they upload is being used, and to be able to evaluate the impact of their donations against institutional goals” (Research:Supporting Commons contribution, 2018).

Aims and Methods

Research Questions

This study attempts to answer the following questions with the hopes of providing concrete strategies for assessing collection reuse to cultural heritage institutions:

1. What is the content of cultural heritage images found in Wikimedia Commons?
2. What content gets reused most often, and where?
3. Do reused cultural heritage images from Wikimedia Commons carry license or attribution information with them ?

Research Methods

A list of cultural heritage repositories, including museums, historical associations, and academic archives, among others, was generated from the archival discovery tool ArchiveGrid, and a random number generator was used to pull a sample of 66 institutions from the list for inclusion in this study. Searches were conducted over a two-week period for images from these institutions’ collections, determined primarily by examining the “Source” field in the Wikimedia Commons object metadata. While images documenting an institution’s buildings

or grounds were not included in the study, user-generated photos or videos of collections, such as pictures taken of an artwork or exhibit, were included. The number of results for each institution varied greatly, with some institutions not having any related images in Wikimedia Commons and others having hundreds of results. A list of all institutions and counts of their reuse results is available in Appendix A: 51 of the 66 institutions had digital images in Wikimedia Commons. As the purpose of this study was not to determine how many cultural heritage institutions have images in Wikimedia Commons, or how many images institutions have on average, not all results were analyzed; instead, at most 20 results from each institution were documented.¹

A total of 308 images from cultural heritage institutions were initially analyzed. A separate research project is underway to assess the validity of rights statements provided in Wikimedia Commons for all of these results. For the purposes of this study, a smaller subset was extracted for RIL analyses. All results from the initial 308 images with Creative Commons or other open licenses were selected for inclusion, as one research question pertinent to this study is how often evidence of open licensing is available when images are reused. These accounted for 44 images to be searched using RIL; an additional 126 public domain images, and two instances of images published with copyright permission from the Wikimedia Commons cultural heritage sample set, were selected for inclusion as well.

Wikimedia Commons includes wiki reuse information on the record page for uploaded media; the number of instances of reuse, both on Wikimedia Commons and on other wikis, was noted for each object (see Figure 1).

¹ The raw, cleaned dataset used for this research paper is available in the author’s Figshare repository (Kelly, 2019).



Figure 1
Screenshot of Wikimedia Commons file usage for “Hume Spring (c.1900) owned by Frank Hume (pictured far right).jpg”
([https://commons.wikimedia.org/wiki/File:Hume_Spring_\(c.1900\)_owned_by_Frank_Hume_\(pictured_far_right\).jpg](https://commons.wikimedia.org/wiki/File:Hume_Spring_(c.1900)_owned_by_Frank_Hume_(pictured_far_right).jpg)).



Figure 2
Screenshot of Google images result with multiple sizes.

Then each image was searched using the Google Chrome browser “Search by image” function. When available, the option to search Google for “all sizes” of the image, as opposed to just those

matching the original image, was selected to receive the greatest amount of results (see Figure 2).

For each image, a number of elements were recorded. These included:

- Repository Name
- Search Term
- Wikimedia Commons result URL
- Original Medium of Reused Media
- Content of Reused Media
- Wikimedia Licensing
- Reuse URL
- Reuse Context (Narrow)
- Reuse Context (Broad)
- Reuse License and Attribution
- Reuse License (Categorized)
- License Compatibility
- Notes

Most of the elements only required simple analysis of frequency counts. For elements with a greater level of subjectivity, such as “content of reused media” and “reuse context,” the content analysis method was used to examine each object, label it, and then categorize the labels into broader themes. Content analysis is a quantitative research method used to “examine large amounts of data in a systematic fashion, which helps to identify and clarify topics of interest” (Drisko & Maschi, 2015, pp. 25). Here, codes or categories were developed inductively, or without a prior scheme, rather than deductively, as reuse research is still in its infancy and existing codes and theory are diverse and not yet synthesized. However, it should be noted that content analysis of some type was conducted in all of the RIL studies previously mentioned, so the potential for integrating codes and developing a standard set for assessing cultural heritage via RIL may be a possibility in the future. In this study, the websites featuring Wikimedia Commons digital cultural heritage images were analyzed as to the site’s purpose. Many results were in languages other than English; for these, Google translate was used to infer the content of the site. Following the analysis and application of codes, tables and graphs were generated to assist in conveying the results of the study.

Results

From 171 digital cultural heritage Wikimedia Commons images searched in Google images, 34 did not have any results. Of the remaining 137 images, one had been deleted in Wikimedia Commons since initial data collection began and couldn’t be searched in Google, and two did not have any wiki results and only had results in Google images that were false positives. Over 25% of Google images results were also discarded as being unusable. These included dead links; false results in which the image was not found on the site; spam, porn websites, and sites blocked by the computer’s antivirus program; one instance of a website that was behind a paywall; and a site that Google translate could not decipher. To ensure that remaining analysis was based on true reuse, any result found by Google images that matched the “Source” field in Wikimedia Commons (for example, if the source of a painting was given as a museum, and Google images located the painting on the museum’s website) was removed from analysis. Finally, 21 results were for videos of a zoetrope at a museum. While instances of wiki reuse could be analyzed for these images, they were not suitable for Google images, so they were removed as well. After fully cleaning any unusable, false, non-reuse, or missing results, a total of 1,533 Google images and wiki search results from 51 cultural heritage institutions remained for analysis.

Approximately 5% of reuse cases from the total uncleaned data set, and 51% of the cleaned data set were associated with Wikimedia’s projects. This includes reuse on other Wikimedia Commons pages like galleries or featured images; reuse on other Wikimedia projects, like Wikipedia articles and Wikidata; reuse by wiki mirror sites, or exact replicas of wiki projects hosted at different URLs; and reuse by wiki aggregators, or sites that pull content straight from Wikimedia and repurpose it for readability, content curation, usability, or other reasons (such as Wikiwand and WikiVividly). While wiki aggregator and mirror results were

Table 1
Reuse Results for Wikimedia Commons Digital Cultural Heritage Images on Wikis and Related Sites

Wiki results	
<i>Result Type</i>	<i>Count</i>
wiki	611
wiki aggregator	158
wiki mirror site	9

Table 2
Content of Reused Wikimedia Commons Digital Cultural Heritage Images Found by Google Images^a

Reuse Content (not including wiki reuse)	<i>Count</i>	<i>Percent</i>
notable people	338	31%
people	251	23%
historic event	157	15%
buildings and locations	103	10%
historic object	34	3%
technology	33	3%
map	32	3%
animals	32	3%
landscape	25	2%
sports	24	2%
other	56	5%

^a "Other" includes fibre art, flowers and plants, outdoor photography, religious iconography, abstract art, diploma, currency, literature, and yearbook photos.

found through Google images, they weren't considered to be true examples of reuse as they simply copied entire Wikipedia articles or Wikimedia Commons galleries without providing any additional context or value to the original Wikimedia Commons object.

The subject matter of the digital images analyzed from Wikimedia Commons was coded, and then Google images results were analyzed to determine themes in what reusers of digital cultural heritage images are most likely to reuse. Note that these subjects are not one-to-one coordinates for each image; a single image could have multiple subjects. Instead, these numbers represent general areas that reusers of digital

cultural heritage tend to focus on when reusing images online. A full description of the codes used to label image content can be found in Appendix B. Notable People or people were included in more than half of the reuse results, while images documenting historical events and buildings and locations were also widely reused. Several categories identified in the initial image analysis were not reused at all outside of wiki products; these were book cover, book plate, data, diaries and personal letters, and library card.

Similar results can be found in analyzing just the reuse of these images on other wikis. The primary difference is that more of the image

Table 3

Content of Reused Wikimedia Commons Digital Cultural Heritage Images Found on Other Wiki Platforms (Wikipedia, Other Wikimedia Commons Page, Wiki Aggregators, and Wiki Mirror Sites)

Reuse Content (wiki only)	<i>Count</i>	<i>Percent</i>
notable people	386	36%
people	159	15%
buildings and locations	110	11%
historic event	96	9%
sports	52	5%
technology	42	4%
animals	33	3%
book cover	28	3%
fiber art	26	2%
currency	24	2%
landscape	23	2%
other ^b	83	8%

^b “Other” includes historic object, outdoor photography, map, advertisement, diaries and personal letters, literature, flowers and plants, religious iconography, abstract art, bookplate, diploma, data, and library card.

Table 4

Comparison of Wikimedia Commons Image Content Categories and Overall Reuse of Those Categories^c

Reuse Content Category	<i>Occurrences in Data Set (before reuse analysis)</i>	<i>Reuse Occurrences (wiki and Google images)</i>
people	38%	19%
notable people	24%	34%
buildings and locations	7%	10%
technology	4%	4%
sports	4%	4%
animals	4%	3%
historic event	3%	12%
landscape	2%	2%
other ^d	13%	13%

^c Table’s percentages do not sum to 100% due to rounding up small percentages,

^d “Other” includes outdoor photography, advertisement, book cover, historic object, map, diaries and personal letters, literature, religious iconography, data, fibre art, currency, flowers and plants, abstract art, bookplate, diploma, library card, and yearbook photos.

Table 5
Original Medium of Reused Images

Original Medium of Reused Media	Count	Percent
photograph	1104	72%
two-dimensional artwork	139	9%
illustration	56	4%
three-dimensional artwork	54	4%
map	44	3%
ephemera	42	3%
exhibit	34	2%
monograph	26	2%
Other ^e	34	1%

^e "Other" includes document, slide, drawing, newspaper, and three-dimensional object.

categories were reused in wiki products, with yearbook photos the only image content that was not reused at all. Also, while the content of reused images varies slightly depending on whether the image is reused on a wiki project or elsewhere, there is generally a strong correlation ($r=0.66$) between wiki reuse and non-wiki reuse.

Finally, for comparison's sake, the following table shows the percentage of instances for each reuse content category found within the initial cleaned data set. This shows a strong correlation between the number of images labeled with a content category and the number of times reused ($r=0.84$). However, people accounted for 38% of the data set but were only reused in 19% of reuse occurrences, while notable people accounted for 24% of the data set but were reused in 34% of instances. Historic events (3% original, 12% reuse) also had a higher level of reuse.

The original medium of the reused object was also documented and analyzed. Photographs accounted for nearly three quarters of all reuse.

When looking at reuse outside of wiki products, there are again clear trends in how and where digital cultural heritage images are being

reused. Social websites, defined here to include social media, blogs, discussion boards, online journals, and other websites whose primary purpose is user-generated content and interaction, account for just under half of reuse instances outside of wiki platforms. Media galleries, or user-curated collections of media (usually images), and news websites are also popular scenes for digital cultural heritage reuse. Only 11% of Google images results for Wikimedia Commons digital cultural images were on educational sites like research guides, encyclopedias, and historical timelines. Full definitions of the codes used to categorize reuse context are in Appendix C.

Slight variances in what subject matter is most viable for reuse on what type of websites can be found as well. While images representing notable people are the most popular reuse type across all websites, maps are almost exclusively found on social sites, whereas images representing historical objects are primarily reused by news sites. Delving further into what subjects are reused most by different types of websites may help cultural heritage institutions pinpoint where their digitization and marketing efforts should lie in order to meet institutional priorities.

Table 6

Context of Reuse of Wikimedia Commons Digital Cultural Heritage Images Found by Google Images^f

Google images reuse context	Count	Percent
social	371	49%
media galleries	137	18%
news	133	18%
education	80	11%
profiles of people and places	14	2%
commerce	9	1%
events	5	1%
web design and development	5	1%
tourism	1	0%

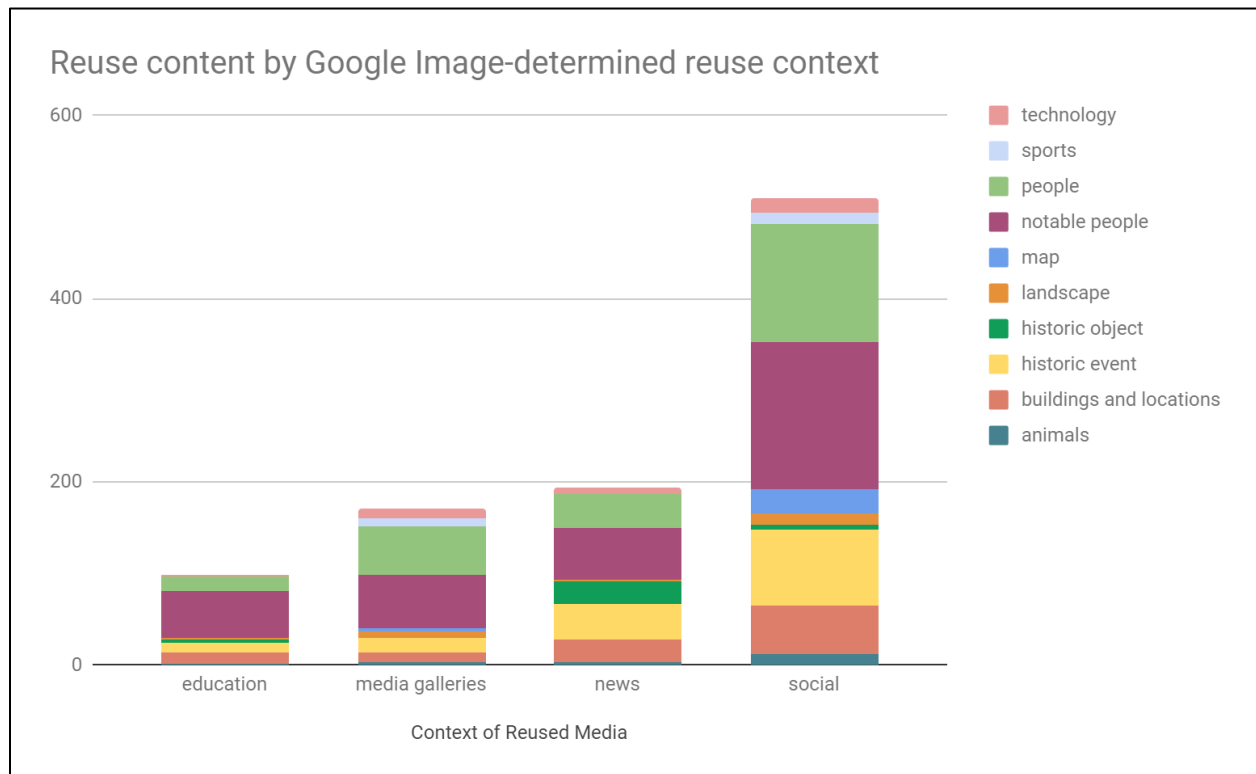
^fTable's percentages do not sum to 100% due to rounding up small percentages.

Figure 3

Reuse context of Wikimedia Commons digital cultural heritage content found by Google Images (excerpt).

Wikimedia provides ample guidelines on how wiki media should be shared from Wiki platforms, including providing appropriate attribution if required by the media's license. Of the sample set analyzed for this study, a mere 40 results out of a possible 755 non-wiki reuse instances had any type of license or copyright statement available. And in comparing the licenses provided in reuse instances, there were significant discrepancies between these and the licenses on Wikimedia Commons. "Compatible" refers to instances where the Wikimedia Commons object and the reused object had the exact same license. The "semi-compatible" designation was used when slight differences occurred, for example, the Wikimedia Commons license listed CC BY-SA 3.0, whereas the reused instance noted an updated CC BY-SA 4.0 license. The remaining "incompatible" results referred to wholly different licenses being applied, such as Wikimedia Commons marking an image as being in the public domain where another website included a Creative Commons or copyright statement alongside the object. The two images that were copyrighted but published to Wikimedia Commons with permission were reused four times outside of wiki products, but none of the reuse instances included a license or attribution.

Finally, a few other unexpected discoveries emerged in this analysis. While only 40 reuse instances provided some sort of license, 147

results, or 19% of non-wiki reuse results at least included some sort of credit, such as the name of the work and the cultural heritage institution that held it. Of these, 50 credited Wikimedia Commons or Wikipedia in some way, or linked back to the original image on Wikimedia Commons.

Also, in analyzing the reuse context of the digital cultural heritage images outside of Wikimedia, only three results appeared to be entirely "misused." These involved the following misidentifications or questionable reuse situations:

- A news article that uses an unlabeled photo of the 1966 UT Austin Tower shooter Charles Whitman's gun to illustrate new laws for gun amnesty in Canada;
- A blog post that mislabels an image of Gerald Ford as Richard Nixon;
- An image of railway workers laying the last rail of the Union Pacific Railroad in 1869, used to illustrate minimum wage.

Overall, reuse of cultural heritage images from Wikimedia Commons was either done without added context or content, as in the case of media galleries, or was done so in ways that did not distort or mischaracterize the image being reused.

Table 7

Compatibility of Reuse Licenses Found by Google Images with Original Wikimedia Commons License

Wikimedia and non-wiki reuse license compatibility	
<i>Compatibility evaluation</i>	<i>Count</i>
compatible	24
incompatible	8
semi-compatible	8

Conclusions

By identifying themes in what type of digital cultural heritage is reused online and where, we can begin to pinpoint possible strategies for cultural heritage institutions to maximize the impact of their digital images depending on institutional priorities. For example, institutions hoping to increase use of their collections by news organizations should focus Wikimedia Commons donation efforts on images related to notable people, historic photos of unidentified people, and historical events, but should also observe that photographs of historic objects ranked highly in reuse by news organizations. However, this study does not delve into great detail as to the content and context of images reused. In this sample set, all of the images labeled as “historic object” were photographs of University of Texas shooter Charles Whitman’s guns. Does this mean that images of weapons in general might be reused more by news organizations than other topics, or would images of other historic objects be reused as frequently? This question could be tested by conducting reuse analysis on Wikimedia Commons images of both historic weapons and generic images of weapons, or of historic weapons and other historic objects. Additional media reuse research should continue to narrow down what exactly makes one media object more reusable than another. Factors such as notability or fame, uniqueness, presentation, artistic merit, and others may be analyzed to further understand reuse priorities.

This study also does not attempt to measure the notability of specific cultural heritage institutions or collections. Previous scholarship documenting cultural heritage institutions voluntarily donating digital images to Wikimedia Commons focuses almost exclusively on large research universities, many of whom have internationally-recognized collections. It is unknown whether smaller institutions with lesser-known or niche collections would see similar increases in website traffic or similar reuse of their digital

images. While this study includes a variety of institution sizes and types, it does not attempt to qualify the notability of these institutions, nor of their collections or individual images. We can, however, see that there is a weak correlation ($r=0.27$) between how many images were searched from each repository and how many instances of reuse were found, so content and quality of the reused object may be larger factors in determining reuse than quantity of object per institution.

The research reported here shows that cultural heritage institutions have cause for concern about reuse of their collections without attribution. Only 9% of Creative Commons-licensed images that were reused outside of wiki projects were labeled as Creative Commons in their new context, only 19% of non-wiki reused images had any sort of credit at all, and most that did, did not include a reuse license or public domain statement. Still, at least for images that are in the public domain and don’t legally require a license or attribution, perhaps cultural heritage institutions should be less concerned with attribution and more concerned with increasing reuse. Unfortunately, a lack of proper attribution can make tracking reuse difficult, thus impeding the institution’s ability to measure the impact of their collections. Strategies such as using RIL to locate instances of reuse without text attribution included may be beneficial for image collections, but as of yet the RIL process is very labour-intensive and probably unfeasible for institutions to perform on all of their digital images on a regular basis. Instead, performing RIL reuse analysis on selected images may be undertaken for specific assessment campaigns, such as to assess reuse of a new collection after a year’s time, to show impact for annual reports and reviews, or to highlight the success of marketing campaigns the institution has undertaken related to a collection or object. The DLF-AIG IMLS grant project found that embedded metadata is one of the most-needed pieces of infrastructure for tracking reuse; the Wikimedia Foundation’s “Supporting Commons contribution by GLAM

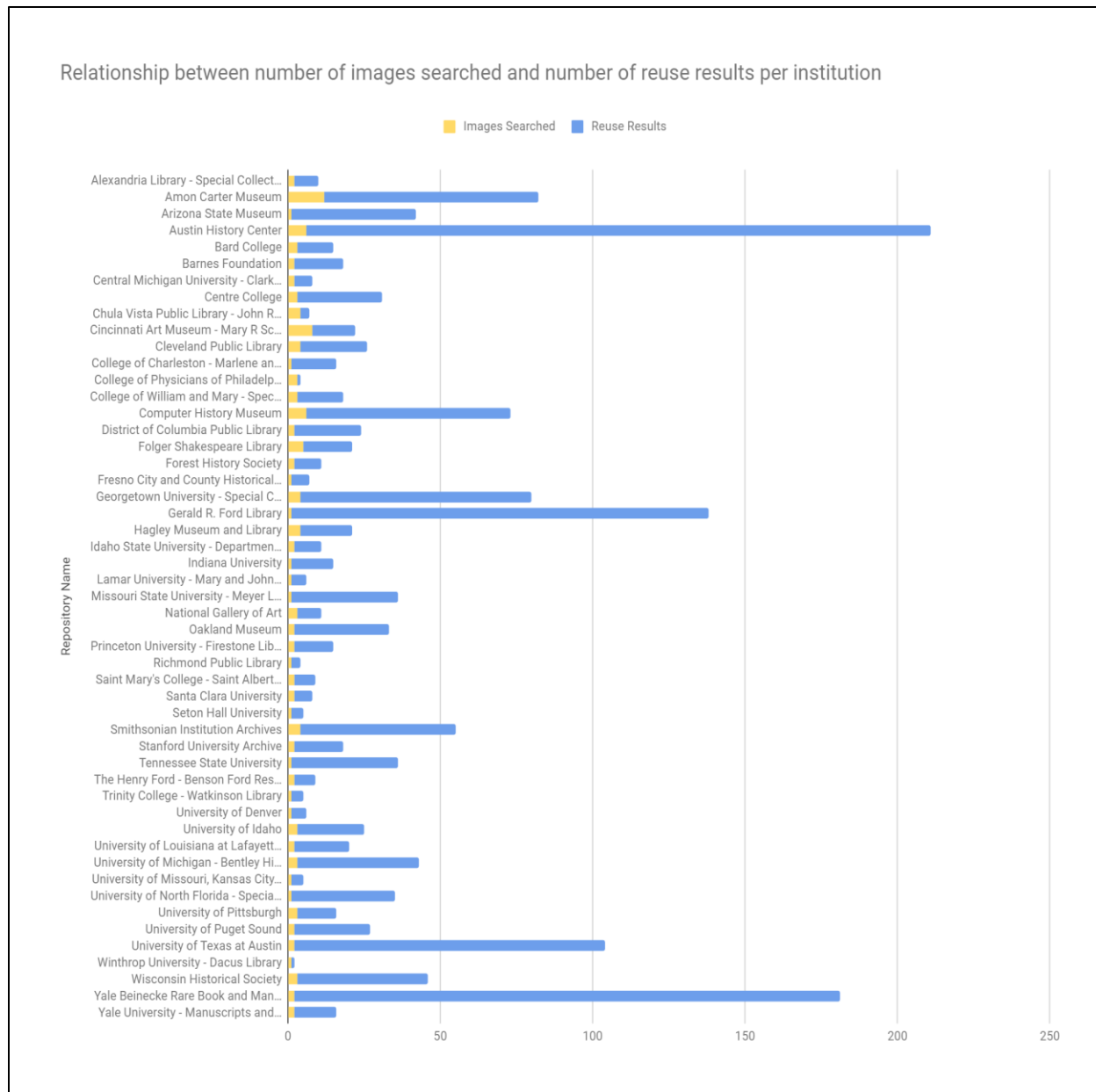


Figure 4

Relationship between number of images searched and number of reuse results per institution.

institutions" project similarly identified "demonstrating and preserving media provenance" as a priority (O'Gara et al., 2018; Research:Supporting Commons contribution, 2018). Improved infrastructure for embedded or sticky" metadata may allow reuse assessment without the need for formal attribution.

What cultural heritage institutions can begin to do with this research is to determine where their digitization efforts may have the most impact and alignment with institutional goals. The DLF-AIG IMLS grant project found that digital library practitioners had different priorities for where they hoped their digital resources would be reused; for example, some institutions might find more value in reuse by nationally-

recognized news organizations, others by students and scholars, still others by community groups (O’Gara et al., 2018). These goals will vary depending on the type, size, and mission of the institution the practitioner represents. By beginning to understand what types of Wikimedia Commons digital cultural heritage content are reused most often on what types of websites, practitioners can strategize which of their collections and objects they should focus on donating to Wikimedia Commons to reach the user communities they are most interested in connecting with.

While great care was taken in developing and analyzing the codes used for identifying content and context of reused images, it should be noted that content analysis as a method is highly subjective but often made less so by involving multiple researchers who “norm” their codes to come to agreement about classification. As this study was undertaken by a sole researcher, elements determined by content analysis may bear a higher level of subjectivity than is desired.

This paper contributes to media reuse literature, and to RIL research in particular by furthering understanding of what content categories are most likely to be reused and where, both within Wikimedia Foundation projects and on the wider web. Digital library practitioners should use the results of this study to develop digitization strategies that prioritize content attractive to the types of websites where reuse would most align with their institutional missions. This research also emphasizes the need for better education and infrastructure related to licensing and rights for digital content reuse, as reused digital cultural heritage images from Wikimedia Commons rarely includes attribution or licensing information. The content categories developed here may be combined with content categories found in other RIL studies to begin synthesizing a common code of subjects for assessing image reuse. By continuing to deepen understanding of digital cultural heritage reuse, we can better assess the

impact of our collections online and strive to meet the needs of current and potential users in line with institutional priorities and missions.

References

- Beaudoin, J. E. (2014). A framework of image use among archaeologists, architects, art historians and artists. *Journal of Documentation*, 70(1), 119–147. <https://doi.org/10.1108/JD-12-2012-0157>
- Commons:First steps/Reuse. (2019). In *Wikimedia Commons*. Retrieved 11 Apr. 2019 from https://commons.wikimedia.org/wiki/Commons:First_steps/Reuse
- Commons:Reusing content outside Wikimedia/technical. (2018). In *Wikimedia Commons*. Retrieved 11 Apr. 2019 from https://commons.wikimedia.org/wiki/Commons:Reusing_content_outside_Wikimedia/technical
- Commons:Simple media reuse guide. (2018). In *Wikimedia Commons*. Retrieved 11 Apr. 2019 from https://commons.wikimedia.org/wiki/Commons:Simple_media_reuse_guide
- Drisko, J., & Maschi, T. (2015). *Content analysis*. New York, NY: Oxford University Press.
- Dryden, J. (2014). Just let it go? Controlling reuse of online holdings. *Archivaria*, (77), 43–71. Retrieved from <https://archivaria.ca/index.php/archivaria/article/view/13486>
- Harris, V., & Hepburn, P. (2013). Trends in image use by historians and the implications for librarians and archivists. *College & Research Libraries*, 74(3). <https://doi.org/10.5860/crl-345>

- Kelly, E. J. (2015). Reverse image lookup of a small academic library digital collection. *Codex: The Journal of the Louisiana Chapter of the ACRL*, 3(2), 80–92. Retrieved from <http://journal.acrlla.org/index.php/code/article/view/101>
- Kelly, E. J. (2018). 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>
- Kelly, E. J. (2019). *2019 Wikimedia Commons digital cultural media analysis (Version 1)* [figshare].
- Kirton, I., & Terras, M. (2013). Where do images of art go once they go online? A reverse image lookup study to assess the dissemination of digitized cultural heritage. MW2013: Museums and the Web 2013, Portland, OR. Retrieved 22 Apr. 2019 from <https://mw2013.museumsandtheweb.com/paper/where-do-images-of-art-go-once-they-go-online-a-reverse-image-lookup-study-to-assess-the-dissemination-of-digitized-cultural-heritage/>
- Kousha, K., Thelwall, M., & Rezaie, S. (2010). Can the impact of scholarly images be assessed online? An exploratory study using image identification technology. *Journal of the American Society for Information Science & Technology*, 61(9), 1734–1744.
- Morley, James. (2018). Use and impact of cultural heritage images on Wikimedia Commons and Wikipedia. In *Catching the Rain*. Retrieved 18 Mar. 2019 from <http://www.catchingtherain.com/portfolio/use-and-impact-of-cultural-heritage-images-on-wikimedia-commons-and-wikipedia/>
- O'Gara, G. M., Woolcott, L., Joan Kelly, E., Muglia, C., Stein, A., & Thompson, S. (2018). Barriers and solutions to assessing digital library reuse: Preliminary findings. *Performance Measurement & Metrics*, 19(3), 130–141. <https://doi.org/10.1108/PMM-03-2018-0012>
- O'Sullivan, S. (2017). Archives for education: The creative reuse of moving images in the United Kingdom. *The Moving Image*, 17(2), xvi–19. <https://doi.org/10.5749/movingimage.17.2.0001>
- Punzalan, R. L., Marsh, D. E., & Cools, K. (2018). Beyond clicks, likes, and downloads: Identifying meaningful impacts for digitized ethnographic archives. *Archivaria*, 84. Retrieved from <https://archivaria.ca/index.php/archivaria/article/view/13614>
- Reilly, M., & Thompson, S. (2014). Understanding ultimate use data and its implication for digital library management: A case study. *Journal of Web Librarianship*, 8(2), 196–213. <https://doi.org/10.1080/19322909.2014.901211>
- Reilly, M., & Thompson, S. (2017). Reverse image lookup: Assessing digital library users and reuses. *Journal of Web Librarianship*, 11(1), 56–68. <https://doi.org/10.1080/19322909.2016.1223573>
- Research:Supporting Commons contribution by GLAM institutions. (2018). In *Meta-Wiki*. Retrieved 15 Apr. 2019 from [https://meta.wikimedia.org/wiki/Research:Supporting Commons contribution by GLAM institutions](https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions)

[ch:Supporting Commons contribution by GLAM institutions](#)

- Rygel, M. C. (2013). Share and share alike: Using Wikimedia Commons to disseminate geophotography. *Abstracts with Programs - Geological Society of America*, 45(7), 381–381.
- Summers, E. (2019). Archival Shapes. In *Inkdroid*. Retrieved 11 Apr. 2019 from <https://inkdroid.org/2019/01/03/archival-shapes/>
- Thompson, S., & Reilly, M. (2017). “A picture is worth a thousand words”: Reverse image lookup and digital library assessment. *Journal of the Association for Information Science & Technology*, 68(9), 2264–2266.
<https://doi.org/10.1002/asi.23847>

- User:The land/thinking about the impact of the Wikimedia movement. (2018). In *Meta-Wiki*. Retrieved 15 Apr. 2019 from https://meta.wikimedia.org/wiki/User:The_Land/Thinking_about_the_impact_of_the_Wikimedia_movement
- Vrandečić, D. (2014). A new metric for Wikimedia. In *Wikipedia Signpost*. Retrieved 22 Apr. 22 2019 from https://en.wikipedia.org/w/index.php?title=Wikipedia:Wikipedia_Signpost/2014-08-20/Op-ed&oldid=671617000

Appendix A
List of Cultural Heritage Institutions with Reuse Results

Cultural Heritage Institution	<i>Total Reuse</i>	<i>Wikimedia Commons</i>	<i>Google Image</i>
Alexandria Library	8	7	1
Amon Carter Museum	49	28	21
Arizona State Museum	41	26	15
Austin Public Library - Austin History Center	205	56	149
Bard College	12	9	3
Barnes Foundation	16	9	7
Central Michigan University - Clarke Historic Library	6	5	1
Centre College - Grace Doherty Library	28	27	1
Chula Vista Public Library	3	3	0
Cincinnati Art Museum	14	6	8
Cleveland Public Library	22	20	2
College of Charleston	15	11	4
College of Physicians of Philadelphia	1	1	0
College of William and Mary	15	12	3
Computer History Museum	67	41	26
District of Columbia Public Library	22	10	12
Folger Shakespeare Library	16	11	5
Forest History Society	9	6	3

Fresno City and County Historic Society Archives	6	5	1
Georgetown University	76	36	40
Gerald R. Ford Library	137	40	97
Hagley Museum and Library	17	8	9
Idaho State University	9	9	0
Indiana University	14	5	9
Lamar University	5	4	1
Missouri State University	35	7	28
National Gallery of Art	8	1	7
Oakland Museum	31	12	19
Princeton University - Firestone Library	13	6	7
Richmond Public Library	3	3	0
Saint Mary's College	7	4	3
Santa Clara University	6	4	2
Seton Hall University	4	4	0
Smithsonian Institution Archives	51	44	7
Stanford University Archive	16	14	2
Tennessee State University	35	11	24
The Henry Ford - Benson Ford Research Center	7	0	7
Trinity College	4	4	0
University of Denver	5	2	3

University of Idaho	22	15	7
University of Louisiana at Lafayette	18	3	15
University of Michigan - Bentley Historic Library	40	29	11
University of Missouri, Kansas City	4	0	4
University of North Florida	34	5	29
University of Pittsburgh	13	11	2
University of Puget Sound	25	22	3
University of Texas at Austin	102	52	50
Winthrop University	1	0	1
Wisconsin Historic Society	43	27	16
Yale Beinecke Rare Book and Manuscript Library	179	93	86
Yale University - Manuscripts and Archives	14	10	4

Appendix B

Image Content Codes

abstract art: fine art lacking recognizable visual references

advertisement: images used for the purpose of promoting a product or service, usually for monetary gain

animals: non-human biological organisms from the kingdom Animalia

book cover: the front of a published monograph

bookplate: identification labels used by monograph owners

buildings and locations: architectural structures, cityscapes, towns, and non-landscape locales

currency: representations of paper or coin money

data: tables and figures used for illustrative purposes to convey information

diaries and personal letters: manuscript materials such as personal writings and correspondence

diploma: paper documenting graduation from some level of education

fibre art: fine art composed of natural or synthetic components like yarn, thread, and string; examples include tapestries, rugs, and embroidery

flowers and plants: multicellular organisms from the kingdom Plantae

historic event: documentation of occurrences with remarkable significance

historic object: documentation of objects with remarkable significance

landscape: natural scenery

library card: identification used to access items at a library

literature: written works, usually published monographs

map: visual depiction of geographic spaces

notable people: individuals identified by name due to their cultural or historical recognizability on Wikimedia Commons

outdoor photography: camera images of the outdoors

people: primarily unidentified individuals primarily or, in a few cases, identified because their images came from yearbook scans but were otherwise not to be found identified elsewhere online

religious iconography: fine art created for the specific purpose of use in or by religious organizations and individuals

sports: athletic events, spaces, or people associated with specific athletic activities

technology: machines and systems used for carrying out technical processes

yearbook photos: images captured for school publications documenting an academic year

Appendix C
Reuse Context Codes

Broad Code	Narrow Codes	Definition
commerce	art store DVD reproduction for purchase trade catalogue	websites whose primary purpose is the sale of commercial products
education	academic website dictionary digital exhibit digital library eBook encyclopedia Google Arts & Culture page infographic institution website on this day presentation quiz quote website report research guide slide deck timeline tutorial video	reference resources such as dictionaries, encyclopedias, research guides, digital libraries and exhibits, timelines, presentation slides, infographics, "on this day" websites, and academic websites
events	event post movie listing	news or other websites with calendar or public relations-related announcements about specific events like workshops, classes, performances, and exhibits
media galleries	clip art gallery Flickr media gallery stock image gallery	websites made up of manually or automatically-generated collections of images
news	article magazine news article newsletter press release	online publishing by television, online, radio, and print news organizations, as well as magazines and other websites for current events
profiles of people and places	city or company profile person profile	generalized biographies or profiles of cities and towns found on specialty topic, non-educational websites

social	blog discussion board Facebook Google Plus journal message board pin board Pinterest reddit song lyrics annotation site Tweet Twitter aggregator	social networks (Facebook, Pinterest, Twitter), blogs, discussion boards, online journals, and other web 2.0 websites whose primary purpose is user-generated content and interaction
tourism	travel site	travel websites
web design and development	keyword trends	tools for website development such as identifying keyword trends for Search Engine Optimization